

# ZIN

Studia informacyjne  
Information studies

VOL. 55 2017 NO. 2(110)

p-ISSN 0324-8194

e-ISSN 2392-2648



---

**WDiB UW**

---



STOWARZYSZENIE  
BIBLIOTEKARZY  
POLSKICH

**ZAGADNIENIA  
INFORMACJI  
NAUKOWEJ**  
Studia informacyjne

**ISSUES IN  
INFORMATION  
SCIENCE**  
Information Studies

VOL. 55 2017 NO. 2(110)  
p-ISSN 0324-8194  
e-ISSN 2392-2648



**WDIiB UW**



STOWARZYSZENIE  
BIBLIOTEKARZY  
POLSKICH



Warszawa 2017

## ISSUES IN INFORMATION SCIENCE – INFORMATION STUDIES

The core purpose of this journal is to provide a forum for the dissemination of scientific papers and research results in the field of information science and other disciplines which analyze social and technological aspects of various information-related activities performed by contemporary communities. Moreover, the journal is to disseminate critical reviews and summaries of new publications in the field of information science and reports from important conferences discussing contemporary information problems.

In addition to that, after fifty years *Issues in Information Science* (ZIN) becomes bilingual to reach foreign readers – papers are available either in Polish or English.

The reconstructed journal is published with slightly extended title *Issues in Information Science – Information Studies* (ZIN – *Information Studies*) – the subtitle emphasizes the interdisciplinary nature of its subject profile covering a broad spectrum of issues studied by various academic disciplines and professional activity domains related to access to resources of recorded information and knowledge and the use of these resources by contemporary man and society. Other subjects to be covered by ZIN involve: 1) theoretical ponderings on the practice of information-related activities performed by various communities, 2) the results of research on the conditions influencing those activities and ways of improving methods and tools employed for the activities in question, 3) the methodology of information science research, information science history and education concerning the information science. The subject profile of semiannual ZIN – *Information Studies* covers, among else, the issues of:

- information science in relation to library science, archival science, museology and other disciplines researching preservation and access to scientific and cultural heritage
- information and knowledge management
- traditional and online scholarly communication
- information and knowledge organization
- metadata theory and practice
- Web 2.0
- Semantic Web
- information architecture
- information websites usability
- digital humanities
- human-computer interaction
- natural language processing
- information retrieval
- use of information and behavior of the information users
- social response to modern information technologies
- digital humanities
- information and digital skills
- information policy
- information ethics.

ZIN – *Information Studies* is addressed to: 1) information science teachers and lecturers, researchers and students, 2) practitioners of information-related activities who analyze methods and tools used to implement those activities in various domains and organizational environments, 3) politicians and donors related to information activities in various domains. The journal content may also be of some interest to teachers, students and researchers in other disciplines of science which deal with various aspects of information existence and use in the contemporary world.

ZIN – *Information Studies* is included in 'B' list of journals scored by Polish Ministry of Science and Higher Education and indexed by: Central European Journal in Social Sciences and Humanities (CEJSH), Cambridge Scientific Abstracts (CSA), Library and Information Science and Technology Abstracts (LISTA), Polish Bibliography of Book Studies (PBB), Knowledge Organization Literature and Polish Scholarly Bibliography (PBN).

## ZAGADNIENIA INFORMACJI NAUKOWEJ – STUDIA INFORMACYJNE

Głównym celem niniejszego czasopisma jest zapewnienie forum dla rozpowszechniania artykułów naukowych i wyników badań z zakresu nauki o informacji (informatologii) oraz innych dyscyplin, w których podejmowane są analizy społecznych i technologicznych aspektów działalności informacyjnej prowadzonej w różnych sferach współczesnego życia społecznego. Czasopismo służyć ma również rozpowszechnianiu krytycznych recenzji i omówień publikacji z tego zakresu oraz problemowych sprawozdań z ważnych konferencji poświęconych współczesnym problemom informacyjnym.

W minionym pięćdziesięcioleciu *Zagadnienia Informacji Naukowej* (ZIN) były czasopismem publikującym teksty wyłącznie po polsku, a zatem adresowanym tylko do czytelnika polskiego.

W nowej formie czasopismo adresowane jest zarówno do czytelnika polskiego jak i zagranicznego, publikujemy artykuły zarówno w języku polskim jak i angielskim. Obecnie czasopismo ukazuje się pod rozszerzonym tytułem: ZIN – *Studia Informacyjne*. Dodany podtytuł podkreśla interdyscyplinarny charakter jego profilu tematycznego, który obejmuje szeroki zakres problemów podejmowanych przez dyscypliny akademickie i dziedziny działalności zawodowej związane z zapewnianiem dostępu do utrwalonych zasobów informacji i wiedzy oraz ich wykorzystywaniem przez współczesnego człowieka i współczesne społeczeństwo. Czasopismo publikuje też artykuły prezentujące teoretyczną refleksję o praktycznej działalności informacyjnej prowadzonej w różnych dziedzinach i obszarach życia społecznego, a także wyniki badań służących poznaniu różnych uwarunkowań tej działalności oraz doskonaleniu jej metod i narzędzi. Na łamach ZIN publikowane są także artykuły poświęcone metodologii badań informatologicznych, historii nauki o informacji oraz edukacji w zakresie nauki o informacji. Profil tematyczny półrocznika ZIN – *Studia Informacyjne* obejmuje m.in. problematykę:

- nauki o informacji w powiązaniu z bibliotekoznawstwem, archiwistyką, muzeologią innymi dyscyplinami zajmującymi się problematyką zachowania i zapewnienia dostępu do dziedzictwa nauki i kultury
- zarządzania informacją i wiedzą
- komunikacji naukowej i cyfrowej komunikacji naukowej
- organizacji informacji i wiedzy
- teorii i praktyki metadanych
- zagadnień Web 2.0
- zagadnień Sieci Semantycznej
- architektury informacji
- projektowania użytecznych serwisów informacyjnych
- humanistyki cyfrowej
- interakcji człowiek – komputer
- przetwarzania języka naturalnego
- wyszukiwania informacji
- wykorzystywania informacji i zachowań informacyjnych użytkowników
- społecznej recepcji nowoczesnych technologii informacyjnych
- kompetencji informacyjnych i cyfrowych
- polityki informacyjnej
- etyki informacyjnej.

*Zagadnienia Informacji Naukowej – Studia Informacyjne* adresowane są do wykładowców, badaczy i studentów nauki o informacji, a także praktyków działalności informacyjnej, krytycznie analizujących metody i narzędzia jej realizacji w różnych środowiskach dziedzinowych i organizacyjnych oraz polityków i donatorów działalności informacyjnej w różnych dziedzinach. Lektura czasopisma może też zainteresować wykładowców, studentów i badaczy innych dyscyplin, które zajmują się równymi aspektami funkcjonowania informacji we współczesnym świecie.

*Zagadnienia Informacji Naukowej* znajdują się na liście B czasopism punktowanych Ministerstwa Nauki i Szkolnictwa Wyższego. Czasopismo jest indeksowane w bazach: Central European Journal in Social Sciences and Humanities (CEJSH), Cambridge Scientific Abstracts (CSA), Library and Information Science and Technology Abstracts (LISTA), Polska Bibliografia Bibliologiczna (PBB), Knowledge Organization Literature, Polska Bibliografia Naukowa (PBN).



# Spis treści | Contents

ROZPRAWY. BADANIA. MATERIAŁY | THESES. RESEARCH. MATERIALS

|   |     |
|---|-----|
| Mieczysław Muraszekiewicz<br><i>Innovativeness and Information Processes in the High-Tech Environment</i><br>[Innowacyjność i procesy informacyjne w przedsiębiorstwach wysokich technologii] .....   | 7   |
| Quoc-Tan Tran<br><i>Ethically Responsible Knowledge Organization Systems: Towards an Intercultural User Interface</i> [Etycznie odpowiedzialne systemy organizacji wiedzy: w kierunku międzykulturowego interfejsu użytkownika] .....   | 16  |
| Sebastian Grabowski, Mieczysław Muraszekiewicz<br><i>Modelowanie ekosystemów informacyjnych dla innowacyjnych społeczności programistycznych</i> [Modeling of Information Ecosystems for Innovative Programming Communities] .....  | 30  |
| Marcin Roszkowski<br><i>Formaty danych strukturalnych w zasobach World Wide Web</i><br>[Structured Data Formats for World Wide Web] .....   | 46  |
| Łukasz Opaliński<br><i>Bibliometryczna metodologia prognozowania i oceny rozwoju dyscyplin naukowych. Analiza piśmiennictwa. Część 2. Badania porównawcze, hybrydowe, statystyczne, analizy dokumentów patentowych, ścieżek rozwoju dyscyplin oraz pozostałe oryginalne podejścia metodologiczne</i> [Methods to Foresee and Assess the Development of Scientific Disciplines. Literature Analysis. Part 2. Comparisons, Hybrid and Statistical Methods, Analysis of Patents and Main Paths of Literature Development and Other Original Approaches in Terms of Predictive Methodology] ..... | 73  |
| Łukasz Opaliński, Marcin Jaromin<br><i>Zastosowanie statystycznej analizy szeregów czasowych do krótkoterminowego prognozowania rozwoju dyscyplin naukowych</i> [Application of Statistical Time Series Analysis in Short-Term Forecasting of the Development of Scientific Disciplines] .....  | 106 |
| Łukasz Iwasiński<br><i>Quantified Self. Self-tracking a problem tożsamości</i> [Quantified Self. Self-Tracking and Identity] .....  | 126 |

Ewa Głowacka, Mirosław Górny, Małgorzata Kisilowska, Zbigniew Osiński  
*Wpływ cyfryzacji infrastruktury informacyjnej na procesy badawcze w humanistyce. Wybrane aspekty zjawiska [Impact of Information Infrastructure Digitization on Research Processes in the Humanities. Selected Aspects of the Phenomenon]* ..... 137

Marzena Świgoń, David Nicolas  
*Postawy i zachowania młodych naukowców – doniesienie z pierwszego etapu badań międzynarodowych, ze szczególnym uwzględnieniem polskich respondentów [Attitudes and Behavior of Junior Researchers – Early Findings of an International Study with a Particular Focus on Polish Researchers]* ..... 156

Anna Kamińska  
*Co mogą zaoferować cyfrowej humanistyce biblioteki i ośrodki informacji? [What Kind of Services Could Libraries and Information Centers Offer to Digital Humanities?]* ..... 171

#### RECENZJE I OMÓWIENIA | REVIEWS

*Wśród zagranicznych książek [Foreign Publications]* (Jacek Wojciechowski) ..... 182

*Przegląd polskich nowości wydawniczych [New Polish Publications]* (Anna Stanis) .. 192

#### KRONIKA | CHRONICLE

*Europejska, Śródziemnomorska i Bliskowschodnia Konferencja Systemów Informacyjnych – EMCIS (Coimbra, Portugalia, 7–8 września 2017) [European, Mediterranean and Middle Eastern Conference on Information Systems – EMCIS (Coimbra, Portugal, September 7–8, 2017)]* (Mariusz Luterek) ..... 196

*Europejska Konferencja Edukacji Informacyjnej (Saint-Malo, Francja, 18–21 września 2017) [European Conference on Information Literacy (Saint-Malo, France, September 18–21, 2017)]* (Zuza Wiorogórska) ..... 198

*XIV Krajowe Forum Informacji Naukowej i Technicznej (Zakopane, 20–22 września 2017) [Fourteenth National Forum for Scientific and Technical Information (Zakopane, September 20–22, 2017)]* (Renata Frączek) ..... 201

WSKAZÓWKI DLA AUTORÓW | GUIDELINES FOR AUTHORS ..... 205

# Innovativeness and Information Processes in the High-Tech Environment

Mieczysław Muraszkiewicz

*Faculty of Economics and Informatics in Vilnius, Lithuania  
Branch of the University of Białystok*

---

## Abstract

**Purpose/Thesis:** The purpose of the study is to examine the relationship among modernization, innovativeness, and information processes in high-technology environments, and provide recommendations on boosting innovativeness, taking the informational dimension of innovation into account. Currently, knowledge and innovations have become the driving force of the world development. It is claimed that innovativeness, as opposed to inventiveness, can be subject to planning and management and as such is usually developed mostly in high-tech organizations and enterprises.

**Approach/Methods:** The method enabling high-tech organizations to create favorable conditions in order to boost innovativeness and develop innovations includes, inter alia: (i) encouraging such organizations and companies to build points of bifurcations within their own structures and business processes where free creative thoughts and experiments could happen; (ii) integrating main information systems that support innovativeness (e.g. systems dealing with company products and know-how, human resources, knowledge management, customer relationships management, and crowdsourcing). The corporate portal seems to be the most appropriate place for such integration, (iii) tapping into the model of open innovation and (iv) offering mobile access to corporate resources.

**Results and conclusions:** The outcome of the research is the outline of a methodology for boosting innovativeness in high-tech environments that may help company managers and decision makers to organize innovation projects.

**Originality/Value:** The value of the research consists in the recognition of the fact that innovation within the high-tech environment needs to be organized, managed and provided with a reliable information support. The latter can be achieved with the integration of various information processes related to the implementation of innovation projects within a corporate portal.

## Keywords

High-tech. ICT. Innovation. Knowledge society. Mobile technology. Modernization. Process integration.

*Received: 10 December 2017. Reviewed: 19 December 2017. Accepted: 21 December 2017.*

---

## 1. About Modernisation and Innovativeness

Given a large complexity and speed of the present high-tech world and business, a broader contextual outlook including social, political and even philosophical aspects is necessary while thinking of its evolution and prospects, especially when it comes to innovation that is one of the mainsprings of change.

Today, when vivid debates and hot discussions sometimes leading to serious political conflicts, on the choice whether we need more liberalism and free market or we should



strengthen and advance the achievements of the welfare state take place in Europe, it seems that those protagonists are right who argue that this dilemma is misleading (Blair, 2005). The real game about the present and future patterns, people's betterment and social cohesion, the role and rules of business making, and about maintaining our cultural and political identity is played on a different ground. The point is not to confront liberalism with social solidarity and welfarism. This game is about something different; it is about modernisation. The subject of the game is to understand, accept and absorb the changes caused by faster and faster developments of technology, especially the information and communications technology, mobile technology, new management techniques, and new financial mechanisms, which transform economic and political spheres and change the very fabric of social architecture and social capital. It is also about setting up new schemes, patterns and procedures to replace the existing dysfunctional ones and to discontinue all the features that are obsolete and counterproductive. We claim that substantial part of social attention and energy should be drawn and channelled to boost modernisation, in particular its high technology facets and information processes of different kind, scope and coverage, rather than to be engaged in fruitless disputes on the advantages of liberalism over welfarism, or vice versa. Undoubtedly, should the modernisation endeavour be carried on, this dilemma will turn out irrelevant.

At the outset, let us elaborate a bit on the meaning of the term "modernisation". Is it a straight continuation of the Enlightenment project initiated by French and British thinkers, philosophers, scientists of the seventeenth and eighteenth centuries, which by means of the driving forces of capitalism has led the so-called Western countries to economic, military and political power? Or, perhaps, as Friedrich Nietzsche already argued, it is the reason of overwhelming nihilism, labour alienation, human's reification, and as his followers such as Theodor Adorno and Max Horkheimer of the Frankfurt School added, a reason of modern evil such as fascism, Bolshevik's tyranny and cruel wars of the twentieth century (Horkheimer & Adorno, 2002). Undoubtedly, modernisation includes both threats, however, in different proportions over time. It is the force that has two faces, a sword of two edges. After the twentieth century lessons of sheer barbarism, after the ecological catastrophes caused by the abuse and immaturity of technology, for instance in the nuclear power plant of Chernobyl or Fukushima or chemical factory in Indian Bhopal, it seems today that by means of these experiences, political reason, and social awareness we are able to minimise negative consequences of modernisation, and to exploit and enjoy its indubitable potential to improve and enrich the quality of our life.

Here, we understand modernisation broadly, i.e. from simple technological improvements facilitating day-to-day life, to new technologies and their various applications systematically integrated with the social and economic fabric of the society, to new financial, organisational, educational, political and social models. Modernisation is thus not limited to the area of economics; it is a process of a much larger reach, encompassing practically all aspects of our life. The main mechanisms of modernisation are various policies, methodologies and tools that are used to devise, shape and organise social processes and economy based on knowledge and relying on the achievements of science and technology. Such terms as "information society" or "knowledge society" used in the context of a modernisation discourse reflect a common conviction that information and knowledge are key factors of modernisation processes.

It is now worthwhile to ask: What is the link that joins information and knowledge with the practice of production and organisational processes occurring in the knowledge society, or in other words, to ask about an operational factor of modernisation? We believe that one of such factors which is particularly important and directly influencing our present and future life is innovativeness. The innovations are this force that transforms discoveries, research and experiments outcomes in a variety of applications and new solutions, which have a direct impact on the form, appearance and structure of the world. If we assume that research and discoveries is the engine of modernisation processes, then we can legitimately consider innovations a fuel for this engine that affects its efficiency and effectiveness. This is not a new opinion and appreciation of the innovation's role. We can find a similar belief expressed in a slightly different manner in the classic book by the Austrian-American economist Joseph A. Schumpeter entitled *The Theory of Economic Development* already published in 1911 (Schumpeter, 1934), which almost a hundred years later provided an intellectual background for the promotion of innovativeness as an important building block and a mainspring of societal development strategies, lately so eagerly employed by the European Commission in its Lisbon Strategy (The Lisbon European Council, 2000).

Innovativeness has been subject to a great deal of papers, reports books, seminars, workshops, and conferences. It seems however that in this ocean of proposals, models, and methodologies still there is an epistemological gap. Namely, one has not managed to work out an approach to innovativeness in which social life or professional life is a starting point of innovation endeavour rather than business reasons such as profit or technological reason like the improvement or enhancement of device functionality. In this approach life should be considered as a multithread well-ordered and organised process including many actors, in which, however, suddenly come out particular points where as a result of original, fresh, non-routine thinking and acting new elements emerge. This is the moment of disruption when a new product or service disrupts an existing setting (e.g. a segment of a market) and value network. The phenomenon of "disruptive innovation", as well as this very term, was identified, analysed and introduced to the innovation discourse by the American scholar Clayton M. Christensen in the mid of the last decade of the previous century, and now is recognised as one of the most influential business idea of the early 21st century (Dyer et al., 2011).

We have already arrived at the point where an explanatory remark has to be made. One has to distinguish innovation from discovery, innovativeness from inventiveness. Innovativeness consists in a new use of the existing techniques, technologies and/or devices to set up new solutions, goods, services or processes. Innovativeness is a creative exploitation of the known things, yet, in a new way, in a new situation or for reaching a new objective. Inventiveness, however, is different for it is aimed at creating entirely new things that have not existed so far. When Sir Winston Churchill was the First Lord of Admiralty in 1911 he initiated a series of deep reforms. Perhaps the most important one was the transition of the Royal Navy from coal to oil, which according to many historians significantly contributed to the British success on the sea and to defeating the Kaiserliche Marine. A more recent example is smartphone, a device engineered for listening to the music, watching photos and videos, and playing games on the move. The smartphone is a creative amalgamation of previously existing ideas, technologies, and brilliant marketing such as Sony's walkman, digital standards MP3(4) and QuickTime.

We consider innovativeness and innovations a major driving force for both the economy and society that helps introduce the society into the era of liquid modernity understood as the world of reason, chances, opportunities and development (Bauman, 2000). Because of a quite common mistake that mixes up innovativeness with inventiveness, it is often said that innovations cannot be planned, that all attempts to set up durable pro-innovation organisational structures in enterprises cannot bring tangible and steady outcomes because the innovations are the results of accidents or fortunate circumstances and, therefore, cannot be efficiently managed. We are of a different opinion, namely, innovation processes can be subject to management, hence the subject to planning, implementation and further applications. For high-tech organisations and companies to maintain a comparative/competitive advantage the issue of innovation development has become of paramount strategic importance, especially where at stake is the survivability on a highly competitive marketplace.

## 2. Boosting Innovations in High-Tech Enterprises

A classic albeit a slightly cliché-type characteristic of high-tech enterprises is that, *inter alia*: (i) they use cutting-edge technologies; (ii) they feature high innovativeness and creativity of staff; (iii) they develop and apply high performance work models; (iv) the lifetime of the applied solutions and technologies by these companies is short; (v) they are exposed to highly competitive market; (vi) they suffer from heterogeneous and floating customer segments; and (vii) their staffers enjoy high salaries (Jacobfeuerborn, 2005). This is really a challenging milieu, where external and internal challenges are key development factors that determine and drive daily operations and employees' behaviour. Undoubtedly, business culture, rules of operation and relationships with the market as practiced by the high-tech companies are dramatically different from those of conventional enterprises. For high-tech companies the major challenge consists in handling business initiatives in such a way that they must balance relevance and speed to the market with the need for flexibility and robustness of the underlying technologies, while taking into account that business process may extend beyond organisational boundaries to address an entire value chain/network. The high-tech enterprises have to set up new standards in terms of business processes integration, management, human resources development, work performance and pro-customer orientation. Towards this end, they tremendously need information, knowledge, customers – and employees-centered managerial solutions and social awareness about the implications of new technologies and services proliferation. All this requires broadly understood innovation, and innovativeness management.

Innovativeness management is not an easy task. Peter F. Drucker argues that the development through innovations requires what he calls a “creative destruction”, and introducing a certain level of disequilibrium into the existing order (Drucker, 1985). We are of the opinion that a pro-innovation environment requires a balanced and regular organisational structure, especially in its decision, manufacturing and/or service rendering centres, but it also includes some disordered sub-systems (sometimes located at the “peripheries” of the company) that may *prima facie* even look slightly chaotic. These sub-systems of disequilibrium are actually the sites where innovations are born. They are the fulcrums of

innovation emergence and crystallisation. In these places of bifurcation semi-chaos boosts innovativeness. Now, we can say that innovativeness planning and management have to identify such sub-systems or to establish frames for them to emerge (towards this end, a few helpful methods are given in (Kelley, 2001; Napier et al., 1998)), and then to evaluate and make use of what these sub-systems produce, and if appropriate, to integrate the results with the rest of the organisation.

Innovativeness is the process that combines technological, organisational and psychological factors, which is getting more and more democratised, mainly owing to information technology and mobile communications technology. It should be noted that the democratisation of innovativeness is a new and promising phenomenon in the history of innovativeness whose consequences are hardly predictable. Information technologies allow more employees, users and customers to be involved in pro-innovation activities (Hippel, 2005), what gives rise to the emergence of innovation networks (Hippel, 2002). In what follows we shall elaborate on this matter a little bit more.

The key-word *creativity economy*, which assumes innovativeness and engagement of employees and customers in a lively dialog, becomes the most relevant characteristic of the present advanced micro-economy. The relationship between manufacturers and service providers and customers become the key to understand, devise and perform business. In this approach the way a company understands and handles its clients shapes the way it is managed. In the world of today, when traditional sources and means of growth are not sufficient to boost growth, only the companies having the leaders and managers who constantly think of customers and innovation, and new types of employees equipped with cutting edge competences can successfully face the incoming challenges. It seems that what is needed now, while preserving the advantages of the Six Sigma paradigm focused mainly on productivity and quality, is a leap from it towards an innovation culture and creative thinking driven by customers' actual and prospective needs.

This leap requires a re-engineering of the information infrastructure of an enterprise towards the integration of main information systems related to business processes management and human resources management that exist in the enterprise with a customer relationship management system. Schematically, this idea is depicted in the Figure 1, where the customer relationship management system encompasses and somehow determines and "inspires and nourishes" other major information systems of the company. As a digression let us note that the concept and often the practice of the staff competence building in high-tech organisations and enterprises still leaves a lot to desire. This may sound really surprising given the number of business schools populating all corners of the world. Part of the answer might be that the business schools generally follow the American model with a traditional emphasis on finance and strategy rather than the "softer" side of management that these days become more and more important and productive. The paper (Jacobfeuerborn, 2005) provides detailed characteristics of and requirements addressed to human resources development systems, human capacities building systems, including e-learning, knowledge management systems, and customer relationship management systems in the context of high-tech organisations and enterprises.

In our approach the term "integration" features several dimensions. First of all, it assumes a presence of a well-defined, consistent bottom line of the company and an overall corporate development policy which are the basis for efficient implementation of business processes,

whether they are decision making, resource management and planning, costs control, coordination, or any other activities within the company. Moreover, “integrated” means that the system has to address needs of different categories of users within the company as well as outside the enterprise, including some participants of the distribution chain, mainly company customers, and protagonists of the supply chain. The third point is that the system has to be integrated with other computerised systems of the company so that a unified platform for providing interoperability and a seamless flow and exchange of data is in place. Next, it is of paramount importance that the interface, access mechanism and navigation philosophy are unified for all the components of the integrated system. Last but not least, the integration includes mobile access to corporate information resources and applications. Employees and managers, and in some cases also customers, have to be given a possibility to interact with the enterprise while on the move. This feature tremendously enhances flexibility of business processes, what is particularly important for fostering innovativeness. Mobility frees creativeness from place and time, which perfectly addresses the motto “innovation needs to be free”.

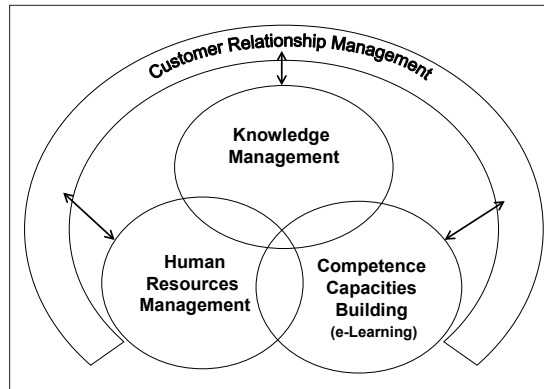


Fig. 1: Integration of information resources and processes

We argue that a portal methodology, including remote and mobile access and mechanisms, is the most appropriate to meet the above integration requirements. The justification is as follows. An enterprise portal is a point of entry that connects employees directly to structured business content and processes through a consistent interface. Company portals are multi-functional and make a multiplicity of various information and services available at a single location. A corporate portal gives its users the impression that all the company information resources and services are available in one place by means of a unified interface that can be personalised to users’ aesthetic and behavioural preferences. With a single log on, portal technology provides the capability to aggregate content from multiple sources, integrates company services and business processes, integrate workflow from multiple sources, offers access to analytical information and, if necessary, facilitates commercial transactions. Portals are the right tools for creating specific collaborative communities for company’s own human resources, salespeople, and customers. Noteworthy, many organisations are already using or researching a portal solution owing to the quick return on investment that comes from a quantum improvement in usability and information organisation.

### 3. Summary

Let us recapitulate, in a nutshell, our line of reasoning adopted in this essay. We have argued that the developed world should not leave or compromise the track of modernisation that in spite of its drawbacks has thus far proven efficient and successful to address our needs and expectations regarding the standard of living and prospect. The dilemma between liberalism and welfarism is false and misleading for the world needs both the economic and social development. Presently, the driving forces of this development are information, knowledge and above all innovations. The places where innovations like to be born are mostly high-tech organisations and enterprises. To help high-tech entities give birth to innovations one has to establish favourable conditions. We have considered three aspects related to this goal, which are as follows. Firstly, one has to allow and even encourage high-tech companies for having and/or establishing in their organisational structures some points of bifurcation where free creative thoughts of the employees and experimentations could happen. Secondly, it is necessary to integrate under the umbrella of a corporate portal equipped with a unified user-friendly interface the main information system helping to handle such business processes as human resources development, human capacities building, knowledge management, and customer relationships management. These are the first processes to be integrated in order to boost innovativeness in the high-tech companies. Obviously, the integration should not be stopped at this stage and has to be continued to encompass other corporate information systems. Thirdly, access to corporate information resources and applications has to be made mobile for creativity and innovativeness have to be freed from place and time.

### References

- Bauman, Z. (2000). *Liquid Modernity*. Cambridge: Polity Press.
- Blair, T. (2005). Prime Minister Tony Blair's address to the EU Parliament. *The Guardian* [online] [12.12.2017], <https://www.theguardian.com/politics/2005/jun/23/speeches.eu>
- Drucker, P. F. (1985). *Innovation and Entrepreneurship*. New York: Harper and Row.
- Dyer, J., Gregersen, H., Christensen, C., M. (2011). *The Innovator's DNA: Mastering the Five Skills of Disruptive Innovators*. Boston, Mass.: Harvard Business Review Press.
- EC (2000). *The Lisbon European Council – An agenda of economic and social renewal for Europe* [online]. Rapid. European Commission Press Release Database [12.12.2017], [http://europa.eu/rapid/press-release\\_IP-00-191\\_en.htm](http://europa.eu/rapid/press-release_IP-00-191_en.htm)
- Hippel, E., von (2002). *Horizontal Innovation Networks – By and For Users* [online]. MIT Sloan School of Management Working Paper No. 4366-02. CiteSeer [12.12.2017], <http://citeseerx.ist.psu.edu/viewdoc/download?sessionid=FF1FFB9752ED41A560A91C0B4CBB383A?doi=10.1.1.12.8692&rep=rep1&type=pdf>
- Hippel, E., von (2005). *Democratizing Innovation* [online]. The MIT Press. Massachusetts Institute of Technology [12.12.2017], <http://web.mit.edu/evhippel/www/sources.htm>
- Horkheimer, M., Adorno, Th.W. (2002). *Dialectic of Enlightenment: Philosophical Fragments (Cultural Memory in the Present)*. Redwood City, Ca.: Stanford University Press.
- Jacobfeuerborn, B. (2005). *Information Systems for Boosting Competence and Management in High-Tech Organizations*. Warszawa: MOST Press.
- Kelley, T. (2001). *The Art of Innovation*. New York, NY: Doubleday.



Napier, R., Sidle, C., Sanaghan, P. (1998). *High Impact Tools and Activities for Strategic Planning*. New York, NY: McGraw-Hill.

Schumpeter, J. A. (1934). *The Theory of Economic Development*. Cambridge, Mass.: Harvard University Press.

---

## Innowacyjność i procesy informacyjne w przedsiębiorstwach wysokich technologii

### Abstrakt

**Cel/Teza:** Celem pracy jest przesledzenie związków występujących pomiędzy ogólnymi procesami modernizacyjnymi, innowacyjnością i procesami informacyjnymi w organizacjach wysokich technologii oraz sformułowanie rekomendacji dotyczących wspierania innowacyjności, biorąc pod uwagę jej informacyjny wymiar. Obecnie istotną siłą sprawczą rozwoju świata jest wiedza i innowacje. Artykuł wyraźnie odróżnia innowacyjność od wynalazczości – ta pierwsza bowiem może być przedmiotem planowania i zarządzania i, jako taka, ma miejsce przede wszystkim w organizacjach i przedsiębiorstwach wysokich technologii.

**Koncepcja/Metody badań:** Metoda pozwalająca organizacjom wysokich technologii tworzyć warunki korzystne dla innowacyjności i opracowywać innowacyjne rozwiązania polega m.in. na: (i) zachęcaniu takich organizacji do tworzenia w ramach swych struktur organizacyjnych i procesów biznesowych punktów bifurkacji, w których może ujawniać się kreatywność i są warunki do eksperymentowania z pomysłami, (ii) integrowaniu głównych korporacyjnych systemów informacyjnych, np. systemów dotyczących produktów, know-how, zasobów ludzkich, zarządzania wiedzą, zarządzania relacjami z klientami, czy mechanizmów crowdsourcingu; szczególnie dogodnym miejscem dla takiej integracji jest portal korporacyjny (iii) wykorzystaniu potencjału modelu otwartych innowacji, (iv) zapewnieniu mobilnego dostępu do korporacyjnych zasobów informacyjnych.

**Wyniki i wnioski:** Przedstawiono zarys metody wspierania procesów innowacyjnych w organizacjach i przedsiębiorstwach wysokich technologii. Metoda ta może pomóc decydom i menedżerom w projektowaniu i realizacji projektów innowacyjnych.

**Oryginalność/Wartość poznawcza:** Opracowana metoda jednoznacznie pokazuje, że innowacyjność w organizacjach wysokich technologii wymaga systematycznego podejścia opartego na planowaniu i regułach zarządzania oraz na stworzeniu solidnego wsparcia informacyjnego. Wsparcie to można zapewnić przez całościową integrację zasobów i procesów informacyjnych w ramach portalu korporacyjnego.

### Słowa kluczowe

Innowacja. Integracja procesów informacyjnych. Modernizacja. Społeczeństwo wiedzy. Technologie mobilne. Wysokie technologie.

---

*Professor MIECZYŚLAW MURASZKIEWICZ is a full professor at the Institute of Computer Science at Warsaw University of Technology. His research interests cover intelligent information systems, methods of knowledge representation, information services for research and development sector, and the relationships between technology and culture as well as issues of innovativeness. His recent publications are: M. Muraszkievicz (2014). An Essay on Information Overload. Zagadnienia Informatyki Naukowej, 52(1), 7–18; B. Jacobfeuerborn, M. Muraszkievicz (2014). Some Challenges and Trends in Information Science. In: R. Bembenik et. al. (eds) Intelligent Tools for Building a Scientific Information Platform: From Research to Implementation (3–14). Springer [Studies in Computational Intelligence No. 541]; B. Jacobfeuerborn, M. Muraszkievicz (2013). Media, Information Overload, and Information Science. In: R. Bembenik et. al.*

*(eds) Intelligent Tools for Building Scientific Information Platform. Advanced Architectures and Solutions (3–13). Springer [Studies in Computational Intelligence No. 467].*

*Contact to the Author:*

*m.muraszkiewicz@ii.pw.edu.pl*

*Uniwersytet w Białymstoku*

*Wydział Ekonomiczno-Informatyczny w Wilnie*

*Kalwarijū g. 135*

*Lt-08221 Vilnius, Lithuania*

*and*

*Instytut Informatyki*

*Wydział Elektroniki i Technik Informacyjnych*

*Politechnika Warszawska*

*Nowowiejska 15/19*

*00–665 Warszawa*



# Ethically Responsible Knowledge Organization Systems: Towards an Intercultural User Interface

Quoc-Tan Tran

*GERiiCO, University of Lille, France*

---

## Abstract

**Purpose/Thesis:** This paper discusses the challenges of creating a theoretic framework within the context of an intercultural and ethically responsible knowledge organization system (KOS).

**Approach/Methods:** First, the paper explores ethical and societal concerns linked to the development of KOS. Second, it illustrates a way to tackle this ethical factor by proposing an applicable architecture for intercultural interfaces which respects cultural diversity on a global scale.

**Results and conclusions:** The author emphasizes the importance of opening up the notion of cultural inclusiveness, to weigh not only linguistic diversity but also other cultural and social aspects, such as geography, religious affiliations, tradition, historical elements, ethnic. This kind of approach should be attentive to intersectionality and cultural interoperability.

**Originality/Value:** By addressing local circumstances, the author offers insights into essential approaches that take into account cultural diversity when designing KOS and access interfaces to knowledge.

## Keywords

Cultural interoperability. Cultural warrant. Ethical warrant. Knowledge organization. Knowledge organization system.

*Received: 3 November 2017. Reviewed: 29 November 2017. Accepted: 15 December 2017.*

---

## 1. Introduction

The development of digital libraries, digital collections, and multilingual resource portals has been pushing forward efforts on the pooling of scattered resources in order to offer them to the public. At the international level, cultural institutions have worked on the conditions of heritage works exchange for which they are responsible and on the metadata used for identifying and describing the works. The challenges of technical and semantic interoperability have been and continue to be at the heart of all online-content development projects (Favier & Mustafa El Hadi, 2013). Moreover, extending the concept of interoperability is what Papy (2015) suggested to understand both what makes access to shared resources possible and what makes them usable in a timely manner.

Since knowledge organization systems (KOS) are mechanisms for organizing information, they are at the core of libraries, archives, and museums (LAMs). The effective integration of KOS in digital environments would facilitate the integration of the vast corpora of recorded knowledge in heritage institutions and born-digital resources on the Web. Most efforts until recently are being put on the development of multilingual/multicultural systems and local

adaptations to feed those systems, but language is only one of the many aspects. There are other social aspects regarding some of the levels and sources of multiculturalism (Barát, 2008), or epistemological aspects such as the recognition of diversity as something inherent to KOS, and the recognition of the impact of human and social factors into knowledge organization (KO) activities (Guimarães, 2015).

This paper stems from results and the reflection during the designing phase of EKOS-Interface project, where the author works as a design engineer. The project aims to deal with two convergent themes: cultural interoperability and ethics in KO. The ultimate goal of EKOS-Interface is to provide an intercultural interface designed for knowledge discovery in multilingual and multicultural contexts by delivering full integration of KOS. The paper begins with examining the theoretical construction of ethical dimension in KOS. The primary purpose of this part is to tackle the issues of cultural interoperability in KO which can be briefly defined as “the degree to which knowledge and information is anchored to a unified model of meaning across cultures” (Vossen et al., 2010, 292).

The next part focuses on the creation and monitoring of ethically and globally accessible and culturally acceptable KOS. It is necessary to highlight the relevance of the works dedicated to conceptual foundation and theoretical construction of the cultural and ethical aspects that have been impacting the field (Olson, 2002; Beghtol, 2002a; 2005; García Gutiérrez, 2002; Hjørland, 2008; Campbell, 2009; Milani & Guimarães, 2011). In particular, the line of this research is linked to scholars who tried to integrate multiple knowledge representation systems or knowledge organization schemes and take into consideration the social and ethical dimensions of KO (Hudon, 1997; Green et al., 2002; Kublik et al., 2003; Mustafa El Hadi, 2015; López-Huertas, 2016).

## 2. Theoretical construction of ethically responsible KOS

### 2.1. *System heterogeneity and universality across KO schemes*

KOS covers all types of schemes and structures for organizing information and promoting knowledge management. They include not only classifications that organize materials at a general level and highly structured vocabularies such as thesauri but also “less traditional schemes, such as semantic networks and ontologies” (Hodge, 2000, 3). Due to that scope, one of the challenges we face in knowledge organization and representation is “heterogeneity of systems both at the level of expressions and structure of conceptual content” (Mustafa El Hadi, 2015, 577). Some theoretical conceptions of ethics in knowledge organization and representation can be highlighted to confront the challenges of promoting processes, tools, and products that are not tied to a given dominant ideology, and which respect the varied forms of knowledge (Green et al., 2002; García Gutiérrez, 2002; López-Huertas, 2016).

Olson (2002) suggested the ethical use of “the power to name”, as the representation of information itself presupposes a power which, making use of processes and tools that are not neutral but rather constructed and agreed upon, leads to constructing products that will act as a representation of the document or as a documentary surrogate. In this sense, although KOS can reflect the mainstream culture of society, they need to be

constantly opened to multiple cultural approaches to avoid exploitation, marginalization, and cultural imperialism.

Regarding the universality of human conceptual systems, one needs to deal first and foremost with human languages to tackle the issues of system heterogeneity. The multicultural dimension of multilingual KOS could be perceived in such a way to recognize the diverse underlying idiosyncratic views that are present in different semantic structures. Hudon (1997) noticed that languages are not only a set of words and rules put together; their conceptual and lexical structures also reflect the way their speakers see and interact with the world. She examined whether multilingual thesauri can act as tools that connect cultures and facilitate interlinguistic communication. Green et al. (2002) looked at a concept's hierarchical level and determined whether it can affect the likelihood of the universality across KO schemes.

## *2.2. Cultural hospitality as an ethical warrant*

Culture is a collective and evolving construct. This concept describes the various phenomena that make up the collective beliefs and activities of a particular group of people. Discussions of culture refer to shared values, history, language, collective memory, social attitudes, preferences, and practices (Beghtol, 2002b). Today, the use of information and communications technology allows for more exchanges among people from a variety of linguistic and sociocultural contexts. Thanks to this cultural diversity, a new reality has developed in the form of cultural exchanges between people belonging to different collectives.

The benefits of cultural interoperability in KO are multiple. First of all, in an intercultural universe, cultural interoperability allows a specific system to integrate with the cultural environment of the other. Arboit and Guimarães (2015) emphasized, from a Bakhtinian perspective, the importance of intersubjective dialogue and responsive understanding in the way an individual interacts with others to “identify himself/herself with others and see the world through his/her own system of values” (Arboit & Guimarães, 2015, 326).

This dual concept permeated by dialogism is also beneficial for the elimination of harmful stereotypes caused by difference. Secondly, in the digital age, cultural interoperability allows one community to broaden its cultural knowledge as well as its skills developed in the interaction with others, regarding whether general knowledge (concepts, disciplinary knowledge, etc.), knowledge specific to the professional environment (management rules, organizational culture, etc.), procedural knowledge (methods, operating rules, procedures, etc.) or know-how which is knowledge updated by learned lessons. Thirdly, by respecting commitment, cultural interoperability reflects an experience of change, which is a process of transformation that involves all senses, knowledge, and behaviors (Al Sahyouni Bou Fadel, 2016).

Beghtol (2002b; 2005) introduced the concept of “cultural warrant” which allows the integration of information and knowledge across cultural, social, national, spatial, temporal, linguistic and domain boundaries, and promotes a “cultural hospitality” by means of KOS opened to incorporate new concepts and to establish appropriate semantic and syntactic relationships among the old and the new concepts. Furthermore, the problems of globalization for KOS can be approached by broadening the concept of hospitality, according to Beghtol (2002a). She argued that “new technologies have made the increased globalization

of information resources and services possible. In this situation, it is ethically and intellectually beneficial to protect cultural and information diversity” (Beghtol, 2002a, 507).

Relying on foundation principles for the ethical treatment of different cultures following the basis of the Universal Declaration of Human Rights, Beghtol (2002a) emphasized the need of building a global and local access to information in any language, available at any location at any time and for any purpose for any individual, culture, ethnic groups, or domain. For that reason, Beghtol concluded that the concept of “cultural hospitality” can act as a theoretical framework for the ethical warrant of KOS.

### 3. Approaches to cultural interoperability

#### 3.1. *Community of practice*

Cultural diversity is a significant issue for many emerging communities of practice, and KO is not an exception. Etienne Wenger, a social learning theorist, hypothesized that the actors of the eco-construction met in the form of “community of practice” (CoP) (Wenger, 1998). The notion of CoP, according to Wenger, refers to all the social, information and communication practices put in place by groups or communities. Focusing on the processes of creating and sharing knowledge, Wenger distinguished three dimensions of CoP: the symbolic one that generates the feeling of belonging to a group and provokes mutual commitment; the cognitive one based on the sharing of available tools and resources by the whole community; and the social one that is responsible for the common work around a shared vision. These three dimensions relate to the concept of interoperability which, particularly in the field of information sciences, relies on the openness, sharing, adaptation, reconciliation of components, policies, and practices.

From this perspective, the tools and resources built in a CoP, which could be remobilized in various situations, allow the continuity and sustainability of the work activities for the actors within the domain. The use of adapted and stable means of communication contributes to the collective action and participates in mutual recognition among the actors. The characterization of their social practices thus depends on the relationship they establish with the new technologies (Soumagnac-Colin, 2016).

#### 3.2. *Domain analysis*

Hjørland (1995; 2002) proposed hypotheses for a domain-analytic paradigm by dissecting its object of inquiry – the domain – which is something that has “a boundary, a specific terminology, basic unit concepts, terms, semantic relations, classification schemes and a shared ontology” (Mustafa El Hadi, 2015, 601).

Tennis (2003) developed Hjørland’s hypotheses by defining the two axes of domain analysis which are needed to consider when analyzing a domain:

Areas of Modulation must state 1) the totality of what is covered in the domain analysis – the extension and 2) what it is called – its name. The Degrees of Specialization must 1) qualify the domain – state its focus and 2) state where the domain is positioned against other domains – its intersection (Tennis, 2003, 194).

Domain analysis in its broad scope can be adapted to a general framework addressing cultural interpretability in KOS. Smiraglia (2012) extended domain analytical approach in KO. By using epistemological analysis, he observed that

the domain is best understood as a unit of analysis for the construction of a KOS. That is, domain is a group with an ontological base that reveals an underlying teleology, a set of common hypotheses, epistemological consensus on methodological approaches, and social semantics (Smiraglia 2012, 114).

Guimarães et al. (2015) proposed a domain analytical approach to identify theoretical referents and epistemic communities within a particular domain of the researchers that constitute a KO scientific community. On the practical level, Mustafa El Hadi (2015) proposed the use of reference tools such as terminologies and thesauri to retrieve the construction and sharing of meaning which refers to a “community of experts” or “discourse community”, while Guimarães et al. (2015) advocate the bibliometric methods.

### ***3.3. Global/local knowledge organization***

Smiraglia (2014) defined a synergistic era as one where “information is seen as a cultural action, information objects are cultural artifacts and cross-institutional information-sharing is a form of cultural synergy” (as cited in Smiraglia, 2015, 297).

Therefore, it requires resource description that is “synergistic, rather than bibliocentric” (Smiraglia, 2015, 297). Guimarães (2015) summarizes the main challenges of KOS in a world permeated by a tension between the global and the local approaches, including the “recognition of diversity as something inherent to KOS” and the recognition of KOS as “tools to promote a global dialogue” (8). These above-mentioned authors, when addressing KO’s cultural dimension, shared Olson’s (2002) vision that the challenge of facing marginalization can be solved by making limits permeable rather than redefining it or constructing a new limit, by making spaces, rather than filling them, and by addressing the relevant discourse in a given context.

Based on the assumption that context matters in any access to knowledge, Adler et al. (2016) proposed a global/local KO view to conceptualize the relationship between the universal and the particular. The tension between the “global” (Paul Otlet’s universalist vision) and the “local” (culturally anchored) remains a fundamental theme in classification, authority control, and links for expanding resource discovery from a local perspective to the global environment.

## **4. Towards an intercultural user interface**

### ***4.1. Applicable architecture***

Language is one of the many aspects, such as age, geography, religious affiliations, tradition, historical elements, racial mix, gender, ideology, and other intangibles that reflect the levels and sources of multiculturalism. The overall vision of improving cultural interoperability is to enhance the accessibility of cultural heritage for the public, to pave the way for new digital humanities approaches, and to contribute to the various efforts to tame the information

flood. The EKOS-Interface project is expected to specify a high-level applicative architecture for intercultural interface design, which could serve as a model for local KOS, and their interoperability with global KOS (Fig. 1).

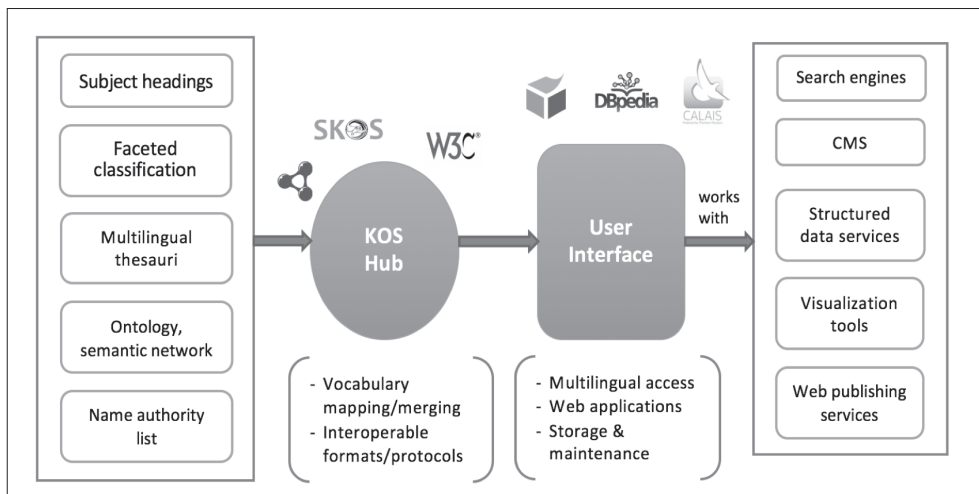


Fig. 1. Applicable architecture for intercultural interface

Barát (2008) listed four possible solutions to linguistic and cultural barriers in KO: the usage of multilingual thesauri, multilingual subject headings, the adaptation and usage of classification systems which are not based on language as the UDC, and machine translation or machine-aided translation. While the use of automatic translation or machine-aided translation in interfaces does not seem to be a feasible alternative yet, most efforts are being put on the development of multilingual/multicultural systems and local adaptations to feed those systems. This strategy is an increasingly vital element to ensure that access to information proceeds through any number of different portals, gateways, and search engines, many geared to particular audiences and subject areas.

EKOS-Interface is perceived and based on the observations of Aitchison & Dextre Clarke (2004), which indicate that, to achieve interoperability of systems, we must design our KO tools (vocabularies, thesaurus, etc.) for easy integration into downstream applications such as content management systems, indexing/meta-tagging interfaces, search engines, and portals. Also, the interface should be multilingual to support access to information resources in multiple languages and to facilitate cross-cultural communication in an increasingly global information society.

#### 4.2. Building and testing the conceptual model

There are challenges involved in creating a user-friendly interface that will meet the needs of a global multilingual access to knowledge. Our work focuses on the creation and monitoring of ethically and globally accessible and culturally acceptable KOS. The core issue of such information architecture is how to consider cultural diversity when designing KOS and access interfaces to knowledge. To provide an intercultural interface intended

for knowledge discovery in multilingual/multicultural contexts and to deliver full integration of KOS, we consider three types of activities (see Fig. 2), which form the iterative research-experimentation-evaluation cycle:

- Data collection and domain analysis (Phase 1);
- Method development (Phase 2);
- Experimentation and evaluation (Phase 3).

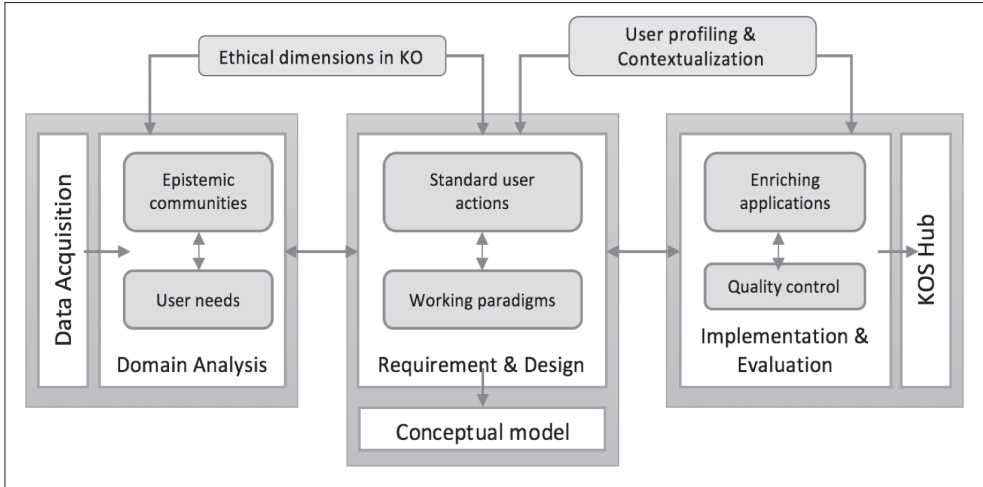


Fig. 2. Constitution of three principal phases

The project's output will be an intercultural interface which potential is to display not only equivalences with the pivot language but also with other communities/languages using the central structure/bone as a node (KOS hub). The main idea is to have concepts of more than two cultures on the screen to display (the switching language together with the local adaptations). This interface will have some common features with VIAF in the way that it is supposed to display all the cultures and languages. The difference is that we would like to have clear-cut displays of one culture which will connect to all the mappings standing behind the first circle. The focus is not to compare but to accommodate all the linguistic and cultural communities.

Regarding KOS functioning, there are alternative interfaces for novice and expert users based on HCI evaluation of user interface techniques in Phase 2. This feature allows expert users to issue the mappings between their language-cultures and the switching language. In this case, the end-user interface should include various functions related to data annotation, validation, classification, and so on. We would give novice users access to our system that already handles mappings via the web, and they would have the only display option. These suggested features in the web application (i.e., interface) should not be too complicated to develop once we decide the set of user profiles, based on the results and observations in Phase 1 (theoretical framework) and Phase 2 (methodology).

One major concern lies in the testing and the evaluation tasks in Phase 3, as the theory of knowledge organization operates with a number of assumptions about user needs and user behavior. For instance, classifications schemes – which have significant value in



a multilingual environment as they are language independent – are often criticized for their culturally biased and rigid structure. Faceted classifications, however, are considered more flexible and less rigid (Broughton 2006; Hudon & Mustafa El Hadi 2017). The variety of categories identified in various classification systems and indexing languages suggests that the faceted classification is just as much a useful analytical tool for category identification as a modeling tool for domain structuring and KOS construction. KOS construction based on facet analysis has been used in graphical user interfaces resulting in many application of networked knowledge organization systems (NKOS) in web information systems and services. Nonetheless, the application of the intercultural approach on facet analytical theory in a real-life scenario was not sufficiently tested. Few research studies have examined how knowledge organization can influence or alter the user behavior and culture in the digital environment. The above example of facet classifications confirms Hjørland's opinion that we do not know how many of the assumptions about knowledge classifications and their power in knowledge browsing in cross-collection, cross-language and cross-cultural would be true today if we give thought to the challenges caused by digital technologies “at both the practical and at the theoretical levels” (Hjørland, 2012, 299).

### 4.3. *Integration of multiple authorities*

The intercultural interface is designed in such a way that users from a given culture would adapt a universal language to their context and use it to interoperate with other cultures while taking advantage of a distributed network of adaptations. This section provides some examples of how the KOS hub integrates multiple controlled vocabularies. This central institution, which is in charge of the switching language, would manage all mappings and make them available to the different cultures or communities.

- (1) Exonym and endonym: many geographical places have different spellings and pronunciation (exonyms) in different languages due to historical reasons or a geographical feature that extends over more than one country. For its initial development, our demonstration uses English as the main language, and for each subject, the exonyms are included if they are available in local KOSs.
  - Brixen (Italian: Bressanone)
  - Lviv (German: Lemberg; Polish: Lwów)
  - Lille (Dutch: Rijsel; German: Ryssel)
  - Mosul (Arabic: الموصل / al-Mawsil; French: Mossoul; Kurdish (Kurmanji): Mûsil; Kurdish (Sorani): مووسڵ)
  - Wrocław (Czech: Vratislav; German: Breslau; Hungarian; Boroszló; Latin: Vratislavia)
  - English Channel [the] (Breton: Mor Breizh, “Sea of Brittany”; Cornish: Mor Bre-tannek, “British Sea”; French: Manche [la], “Sleeve [the]”; German: Ärmelkanal, “Sleeve Channel”)

Another issue is that a geographical name may have almost the same pronunciation but slightly different writing (endonym) in two languages, sometimes due to transliteration, transcription, or the orthographies which do not match (Kerfoot & Närhi, 2006). Some examples are Sao Paulo (for São Paulo), Malaga (for Málaga) or Lubeck (for Lübeck). In all these cases mentioned above, we preserve the original graphic form with the same diacritics given by Wikidata and provide interlanguage links to the variations.



- (2) Cognate name: Many given names are derived from the Bible, names of early Christian saints, or varieties of ancient languages. Though they can be spelled different ways, all cognates of the same name have a common etymological origin. For each subject, we present its equivalents (cognates) in different languages or various linguistic areas:
  - Peter (given name), in Peter the Great (French: Pierre Ier le Grand; Polish: Piotr I Wielki). Notes: Derived from Πέτρος (Petros) which means “stone” in Greek, the name Peter became popular, particularly in the community of Christians, due to the same name of a renown apostle. Other cognates of Peter include Petrus (Biblical Latin), Pere (Catalan), Pierre (French), Piotr (Polish), Pedro (Spanish; Portuguese), Petro (Ukrainian), or Phêrô (Vietnamese)
  - Catherine (given name) (German: Katharina; Polish: Katarzyna)
  - Louis (given name) (Dutch: Lodewijk; Italian: Luigi; German: Ludwig; Romanian: Ludovic; Portuguese: Luís).
- (3) Abortion: For a concept that copes with complicated realities, the subject portal provides other variants or matching concepts from other schemes to accommodate different social and cultural contexts. Even though the French word for the term is *avortement*, the French press seems to endorse medical terminology as it utilizes instead *IVG* (the short for *interruption volontaire de grossesse*). On the contrary, V.T.O.P. (voluntary termination of pregnancy) exists in English, but its use is less prevalent.
- (4) Euthanasia: Similar to the case of “abortion,” we present related terms (Death with Dignity) and closely matching concepts (Assisted-Dying; Physician-Assisted Death/Dying; Physician-Assisted Suicide) to the subject “euthanasia”. It should be noted that this term has a negative meaning in German-speaking countries because it associates with the Nazi eugenics policies, and the term *sterbehilfe* (assisted dying) is used instead.

These two examples mentioned above shed light on how a concise concept which contains the essence of complicated realities can be displayed in a subject portal together with its variations, related terms, exact matching or closely matching concepts from other schemes. Some attention is paid to use of multilingual authorities and multiple authority identifiers as a means of structuring information in the intercultural interface. A display that shows all “cultures” and links them in a single graph can be helpful to visualize and access to variations that might have been more established in other cultures. A possible solution is to navigate from various authority files such as VIAF, WorldCat, ISNI, SUDOC to an integrated record for each subject in the hub. This type of managing authority records based on linked authority identifiers is being utilized by large controlled vocabularies, such as those of Library of Congress, National Library of France, German National Library, and by library resource management systems like Ex Libris Alma.

## 5. Future works

Regardless of the type of architecture, the application should enable access to all mappings by navigating linked authority records, allowing the interface to display as many different “cultures” as desired. This application relies on the solution that endorses the equal

treatment among communities and cultures. On the one hand, it enables the creation of local classifications for different communities or cultures that would interconnect using one classification system as a switching language. On another hand, users from a given culture would adapt a universal language to their context and use it to interoperate with other cultures while taking advantage of a distributed network of adaptations. Each community would focus only on their own mapping, while the central institution in charge of the switching language would manage all mappings and make them available to the different communities. Local interfaces would have the potential to show not only an equivalence with the universal language but also with other communities' languages using the central institution as a node.

Discussions and research on terminology mappings and architectures are still going on in the knowledge organization field. We can cite, for example, the "noodle architecture" (Voß, 2016) that is supposedly more appropriate for the web. In this architecture, there would not be any central terminology, but all terminologies would be treated equally. It would be one of the alternative approaches to consider in a discussion on diversity and multiple cultures.

## 6. Conclusion

This paper attempts to provide elements to support the theoretical construction of ethically responsible KOS. It examines the practices and challenges of cultural interoperability and the place it occupies today in the information society marked by activities of intercultural exchange. It also illustrates a way to tackle the ethical factor of multilingualism, cultural hospitality, and the power to name by proposing an applicable intercultural interface. The goal of building and testing the conceptual model is to promote a semantic technology oriented and standards-based approach to managing controlled vocabularies and reuse of existing classification data to improve resource discovery in digital libraries which will help in achieving KOS sustainability. In this way, the ethical factor is integrated into the processes of development and revision of conceptual structures not only for information retrieval but also classification and indexing.

The purpose of the intercultural interface design is to address the problem of developing a mode of access and organization of knowledge which takes into consideration cultural diversity on a global scale. By far in its developing phase, the EKOS-Interface project is firstly a theoretical reflection that nourishes a proof-of-concept that is the culturally-inclusive and ethically-oriented KOS, which represents the primary deliverable. Through some demonstrations, the author tries to address global cultural biases by designing an inclusive KOS interface and acknowledging the importance of opening up the notion of cultural inclusiveness, i.e., focusing not only linguistic diversity (as a large part of the existing literature does), but also other social aspects, such as geography, religious affiliations, tradition, historical elements, ethnic, gender, which has been the primary object of research for the culture-oriented branch in KO. After all, one step towards achieving cultural interoperability is to promote processes, tools, and services that are not tied to any given dominant ideology, and which respect the varied forms of knowledge.

## 7. Acknowledgements

Part of this work has been carried out in the context of the designing phase of EKOS-Interface, a joint project between GERiiCO Laboratory, University of Lille (France) and the Information Science Department, São Paulo State University (Brazil). The author would like to thank Prof. Widad Mustafa El Hadi (University of Lille) and Dr. Daniel Martínez-Ávila (São Paulo State University) for their precious advice and suggestions.

## References

- Adler, M.A., Tennis, J.T., Martínez-Ávila, D., Guimarães, J.A.C., Mai, J.-E., Olesen-Bagneux, O., Skouvig, L. (2016). Global/Local Knowledge Organization: Contexts and questions. *Proceedings of the Association for Information Science and Technology*, 53(1), 1–4.
- Aitchison, J., Dextre Clarke, S. (2004). The Thesaurus: A Historical Viewpoint, with a Look to the Future. *Cataloging and Classification Quarterly*, 37(3/4), 177–191.
- Al Sahyouni Bou Fadel, R. (2016). L'interopérabilité culturelle et l'interculturalité. *Communication* [online], 34(1), [11.12.2017], <http://journals.openedition.org/communication/6636>; DOI : 10.4000/communication.6636
- Arboit, A.E., Guimarães, J.A.C. (2015). The Ethics of Knowledge Organization and Representation from a Bakhtinian Perspective. *Knowledge Organization*, 42(5), 324–331.
- Barát, Á.H. (2008). Knowledge Organization in the Cross-cultural and Multicultural Society. In: *Culture and Identity in Knowledge Organization*. Proceedings of the Tenth International ISKO Conference 5–8 August 2008, Montréal, Canada. *Advances in Knowledge Organization*, vol. 11 (91–97). Würzburg: Ergon Verlag.
- Beghtol, C. (2002a). A Proposed Ethical Warrant for Global Knowledge Representation and Organization Systems. *Journal of Documentation*, 58(5), 507–532.
- Beghtol, C. (2002b). Universal Concepts, Cultural Warrant and Cultural Hospitality. In: *Challenges in Knowledge Representation and Organization for the 21st Century Integration of Knowledge Across Boundaries*. Proceedings of the Seventh International ISKO Conference, Granada, 10–13 July 2002, Granada, Spain. *Advances in Knowledge Organization* vol. 8, (45–49). Würzburg: Ergon Verlag.
- Beghtol, C. (2005). Ethical Decision-Making for Knowledge Representation and Organization Systems for Global Use. *Journal of the American Society for Information Science and Technology*, 56(9), 903–912.
- Broughton, V. (2006). The Need for a Faceted Classification as the Basis of All Methods of Information Retrieval. *Aslib Proceedings: New Information Perspectives*, 58(1/2), 49–72.
- Campbell, D.G. (2009). Tensions Between Language and Discourse in North American Knowledge Organization. NASKO: North American Symposium on Knowledge Organization [online]. *Proceedings from North American Symposium on Knowledge Organization*, vol. 2 (10–16) [11.12.2017], <https://journals.lib.washington.edu/index.php/nasko/issue/view/863>
- Favier, L., Mustafa El Hadi, W. (2013). L'interopérabilité des systèmes d'organisation des connaissances : une nouvelle conception de l'universalité du savoir? In: *Recherches ouvertes sur le numérique* (22–38). Paris: Lavoisier.
- Green, R., Bean, C., Hudon, M. (2002). Universality and Basic Level Concepts. In: *Challenges in Knowledge Representation and Organization for the 21st Century: Integration of Knowledge Across Boundaries*. Proceedings of the Seventh International ISKO Conference, Granada, 10–13 July 2002, Granada, Spain. *Advances in Knowledge Organization*, vol. 8 (311–317). Würzburg: Ergon Verlag.
- Guimarães, J. A. C. (2015). *Time and Space: Two Axes for Slanted Knowledge Organization* (Global/Local Knowledge Organization Conference and Workshop, University of Copenhagen, August

- 2015) [online]. Jens-Eric Mai. Professor of Information Studies [11.12.2017], <http://jenserikmai.info/glocalko/Guimaraes.pdf>
- Guimarães, J.A.C., Martínez-Ávila, D., Alves, B.H. (2015). Epistemic Communities in Knowledge Organization: An Analysis of Research Trends in the Knowledge Organization Journal [online]. *Knowledge Organization – Making a Difference: The Impact of Knowledge Organization on Society, Scholarship and Progress. ISKO UK Biennial Conference 13th – 14th July 2015, London*. ISKO UK. The UK Chapter of ISKO [11.12.2017], <http://www.iskouk.org/content/epistemic-communities-knowledge-organization-analysis-research-trends-knowledge-organization>
- Gutiérrez, A.L.G. (2002). Knowledge Organization from a “Culture of the Border”: Towards a Trans-cultural Ethics of Mediation. In: *Challenges in Knowledge Representation and Organization for the 21st Century: Integration of Knowledge Across Boundaries*. Proceedings of the Seventh International ISKO Conference, Granada, 10–13 July 2002, Granada, Spain. *Advances in Knowledge Organization* vol. 8 (516–522). Würzburg: Ergon Verlag.
- Hjørland, B., Albrechtsen, H. (1995). Toward a New Horizon in Information Science: *Journal of the American Society for Information Science and Technology*, 46(6), 400–425.
- Hjørland, B. (2002). Domain Analysis in Information Science: Eleven Approaches – Traditional as Well as Innovative. *Journal of Documentation*, 58(4), 422–462.
- Hjørland, B. (2008). What is Knowledge Organization? *Knowledge Organization*, 35(2), 86–101.
- Hjørland, B. (2012). Is Classification Necessary After Google? *Journal of Documentation*, 68(3), 299–317.
- Hodge, G. (2000). *Systems of Knowledge Organization for Digital Libraries: Beyond Traditional Authority Files* [online]. Digital Library Federation, Council on Library and Information Resources [11.12.2017], <https://www.clir.org/wp-content/uploads/sites/6/pub91.pdf>
- Hudon, M., Mustafa El Hadi, W. (2017). Introduction. La classification à facettes revisitée. De la théorie à la pratique. *Les Cahiers du numérique*, 13(1), 9–24.
- Hudon, M. (1997). Multilingual thesaurus construction-integrating the views of different cultures in one gateway to knowledge and concepts. *Information Services & Use*, 17(2/3), 111–123.
- Kerfoot, H., Närhi, E. M. (2006). *Manual for the national standardization of geographical names*. New York: United Nations Publication.
- Kublik, A., Clevette, V., Ward, D., Olson, H. (2003). Adapting Dominant Classifications to Particular Contexts. *Cataloging & Classification Quarterly*, 37(1–2), 13–31.
- López-Huertas, M.J. (2016). The Integration of Culture in Knowledge Organization Systems. In: *Knowledge Organization for a Sustainable World: Challenges and Perspectives for Cultural, Scientific, and Technological Sharing in a Connected Society*. Proceedings of the Fourteenth International ISKO Conference, 27–29 September 2016, Rio de Janeiro, Brazil. *Advances in Knowledge Organization*, vol. 15 (13–28). Würzburg: Ergon Verlag.
- Milani, S.O., Guimarães, J.A.C. (2011). Biases in Knowledge Representation: An Analysis of the Feminine Domain in Brazilian Indexing Languages. *NASKO: North American Symposium on Knowledge Organization* [online], 3(1), 94–104 [11.12.2017], <http://journals.lib.washington.edu/index.php/nasko/article/view/12794>, <http://dx.doi.org/10.7152/nasko.v3i1.12794>
- Mustafa El Hadi, W., Favier, L. (2014). Bridging the Gaps Between Knowledge Organization and Digital Humanities. In: *Knowledge Organization in the 21st Century: Between Historical Patterns and Future Prospects*. Proceedings of the Thirteenth International ISKO Conference, 19–22 May 2014, Kraków, Poland. *Advances in Knowledge Organization*, vol. 14 (477–485). Würzburg: Ergon Verlag.
- Mustafa El Hadi, W. (2015). Cultural Interoperability and Knowledge Organization Systems. In: J. A. C. Guimarães & V. Dodebei (eds.), *Organização do Conhecimento e Diversidade Cultural*. Proceedings of the 3rd Brazilian ISKO-Conference (575–606). Marília, São Paulo: Fundação para o Desenvolvimento do Ensino, Pesquisa e Extensão (FUNDEPE).
- Olson, H.A. (2002). *The Power to Name: Locating the Limits of Subject Representation in Libraries*. Dordrecht: Kluwer Academic Publishers.

- Papy, F. (2015). *Bibliothèques numériques : interopérabilité et usages*. London: ISTE editions.
- Smiraglia, R. (2012). Epistemology of Domain Analysis. In: R.P. Smiraglia, H. – L. Lee (eds.), *Cultural Frames of Knowledge* (111–124). Würzburg: Ergon Verlag.
- Smiraglia, R. (2014). *Cultural Synergies in Information Institutions*. Heidelberg – New York – Dordrecht – London : Springer.
- Smiraglia, R. (2015). Bibliocentrism Revisited: RDA and FRBR00. *Knowledge Organization*, 42(5), 296–301.
- Soumagnac-Colin, K. (2016). Construire l'interopérabilité culturelle autour de la documentation et de l'information dans une communauté de pratique. *Communication* [online], vol. 34(1), [11.12.2017], <http://journals.openedition.org/communication/6670> ; DOI : 10.4000/communication.6670
- Tennis, J. (2003). Two Axes of Domains for Domain Analysis. *Knowledge Organization*, 30(3), 191–195.
- Vossen, P., Agirre, E., Bond, F., Bosma, W., Herold, A., Hicks, A., ... Tesconi, M. (2010). KYOTO: A Wiki for Establishing Semantic Interoperability for Knowledge Sharing Across Languages and Cultures. In: *Handbook of Research on Culturally-Aware Information Technology: Perspectives and Models* (265–293). Hershey, New York: Information Science Reference.
- Voß, J. (2016). *coli-conc Technical Report 3: A Novel Approach to Terminology Mappings*. [online] Zenodo [11.12.2017], <https://doi.org/10.5281/zenodo.48740>
- Wenger, E. (1999). *Communities of Practice: Learning, Meaning, and Identity*. Cambridge University Press.
- 

## Etycznie odpowiedzialne systemy organizacji wiedzy: w kierunku międzykulturowego interfejsu użytkownika

### Abstrakt

**Cel/Teza:** W artykule omówiono wyzwania związane z wyznaczaniem teoretycznych założeń dla tworzenia systemu organizacji wiedzy (SOW), spełniającego postulaty międzykulturowości i odpowiedzialności etycznej.

**Koncepcja/Metody badań:** Najpierw poddano analizie problemy etyczne i społeczne związane z tworzeniem SOW. Następnie zilustrowano sposób rozwiązania problemu etycznego podejścia do projektowania SOW, proponując architekturę dla międzykulturowych interfejsów, która uwzględnia szeroko rozumianą różnorodność kulturową w skali globalnej.

**Wyniki i wnioski:** Autor podkreśla znaczenie otwarcia pojęcia kulturowej inkluzywności i objęcia nim nie tylko kwestii różnorodności językowej, ale także innych aspektów kulturowych i społecznych, które mogą stanowić źródło rozwiązań godzących w poczucie równości i etyczności, takich jak zmiany geopolityczne, przynależności religijne, tradycja, elementy historyczne, etniczność. Takie podejście powinno uwzględniać interseksjonalność i kulturową interoperacyjność.

**Oryginalność/Wartość poznawcza:** Analizując lokalne uwarunkowania, autor oferuje wgląd w podstawowe podejścia uwzględniające różnorodność kulturową w projektowaniu SOW i interfejsów zapewniających dostęp do wiedzy.

### Słowa kluczowe

Interoperacyjność kulturowa. Nakaz kulturowy. Nakaz etyczny. Organizacja wiedzy. System organizacji wiedzy.

---

QUOC-TAN TRAN is a doctoral student at GERiiCO Laboratory, University of Lille (France). He holds a double Bachelor's degree in Information Systems and Business from Swinburne University of Technology

*(Australia). He served as the interface designer for Aboriginal Smart Art, the information system for Aboriginal art market in Victoria, Australia. He developed a thesaurus on the theme of "Jewish heritage" to support automatic indexing in Galicia Jewish Museum in Poland. He holds a Research Master's degree in Library and Information Science from University of Lille in 2016. His area of investigation is cultural interoperability in knowledge organization systems.*

*Contact to the Author:*

*quoc-tan.tran@etu.univ-lille3.fr*

*12 rue de Philadelphie*

*59800 Lille, France*



# Modelowanie ekosystemów informacyjnych dla innowacyjnych społeczności programistycznych

Sebastian Grabowski

*Orange Polska S.A.*

Mieczysław Muraszkievicz

*Wydział Elektroniki i Technik Informacyjnych  
Politechnika Warszawska*

---

## Abstrakt

**Cel/Teza:** Celem artykułu jest przedstawienie generycznego modelu ekosystemu informacyjnego wspierającego innowacyjne społeczności programistyczne, zwłaszcza takie, które tworzą się i działają w środowiskach miejskich.

**Koncepcja/Metody badań:** Metodę badań oparto na analizie wybranych, zagranicznych sieciowych ekosystemów wspierających społeczności programistyczne (m.in. Deutsche Telekom Development Garden i AT&T Developers Portal) oraz współtworzonych przez Sebastiana Grabowskiego społeczności Open Middleware 2.0 Community i BIHAPI. Zanalizowano strukturę tych ekosystemów i ich *modus operandi*. Wnioski wyciągnięte z analizy, z naciskiem na czynniki wspomagające procesy innowacyjne, posłużyły do zbudowania generycznego modelu ekosystemu informacyjnego, wspierającego innowacyjne społeczności programistyczne

**Wyniki i wnioski:** Przeprowadzone badania ekosystemów oraz doświadczenia własne zebrane przez autorów w wyniku współpracy ze społecznościami programistycznymi w ramach różnych projektów i programów, w tym programów Open Middleware 2.0 Community i BIHAPI oraz licznych hackathonów, potwierdziły istotną rolę, jaką już odgrywają i mają do odegrania w przyszłości innowacyjne społeczności programistyczne. Okazało się, że szczególnie efektywnym i produktywnym modelem organizowania się i działania takich społeczności jest korzystanie z otwartych platform, a w szczególności z otwartego oprogramowania, otwartych danych oraz otwartych innowacji.

**Zastosowanie praktyczne:** Przedstawiony model generyczny ekosystemu informacyjnego wspierającego innowacyjne społeczności programistyczne może posłużyć jako punkt wyjścia do dalszych prac nad konstruowaniem, architekturą, funkcjami i zarządzaniem społecznościami programistycznymi, których znaczenie w tworzeniu innowacji technicznych i wspierania nowatorskich rozwiązań gospodarczych i społecznych, nabiera szczególnego znaczenia w kontekście wysiłków wyrwania się Polski z pułapki średniego rozwoju.

**Oryginalność/Wartość poznawcza:** Zaproponowany model ma charakter generyczny, co oznacza, że może być adaptowany do środowisk programistycznych o różnym profilu i celach. Jest on oparty na otwartych platformach dotyczących danych i oprogramowania oraz kładzie nacisk na funkcjonowanie w trybie charakterystycznym dla otwartych innowacji, sam wykazując nowatorskie podejście do funkcjonowania programistów. Model ma charakter otwarty, co tworzy solidną podstawę do dalszych prac nad modelowaniem i realizacją ekosystemów informacyjnych wspierających środowiska programistyczne. Wartość modelu została potwierdzona w praktyce, m.in. w czasie trwania programów Open Middleware 2.0 Community oraz BIHAPI.

## Słowa kluczowe

API. Ekosystem informacyjny. Interfejs programowania aplikacji. Otwarte dane. Otwarte innowacje. Otwarte oprogramowanie. Społeczność programistyczna.

Otrzymany: 23 października 2017. Zrecenzowany: 24 listopada 2017. Zatwierdzony: 27 grudnia 2017.

## 1. Wstęp

Informatyka i jej zastosowania to dziedzina, która horyzontalnie przenika niemal wszystkie obszary współczesnego życia, niosąc ze sobą jakościowe zmiany w gospodarce, administracji, obronności, edukacji, służbie zdrowia, kulturze, rozrywce i generalnie w stylu życia. Rosnące moce obliczeniowe i transmisyjne sprzętu teleinformatycznego na niewiele jednak by się zdały, gdyby nie nieprzerwany i znaczący rozwój oprogramowania narzędziowego i aplikacyjnego oraz wręcz lawinowy wzrost liczby różnorodnych aplikacji. Tylko w Stanach Zjednoczonych Ameryki Północnej w 2016 r. pracowało 3.97 mln twórców oprogramowania (Global Developer, 2017); szacuje się, że w Polsce jest ich obecnie około 250 tys. Programiści, podobnie jak inne społeczności, wytworzyli swą własną kulturę, z własnym stylem pracy, sposobem bycia i komunikowania się oraz z własnym slangiem. Świat programistów, ze względu na stałe kontakty z wysokimi technologiami i z uwagi na szybkie zmiany tych technologii oraz różnorodność tematyki, wymaga kreatywności, gotowości do ustawicznego uczenia się, zachęcając jednocześnie do innowacyjnego myślenia i działania.

Cechami tego świata, które są szczególnie istotne dla programistów i ich rozwoju, są otwartość i sieciowość. Programiści, nawet jeśli pracują w różnych zespołach czy organizacjach współpracują ze sobą, wymieniając się informacjami, doświadczeniami i elementami swojej pracy, zarówno jeśli chodzi o metody, jak i konkretne produkty (programy komputerowe, zbiory danych), tworząc w ten sposób sieci, które w wielu przypadkach stają się częściami większych ekosystemów. Jednym z powodów tej współpracy jest stały „głód” informacji, który towarzyszy indywidualnym programistom i zespołom programistycznym. Każda zatem inicjatywa, każdy system informacyjny, które mogą im pomóc w uzyskaniu relewantnych informacji i wiedzy wywołuje w środowiskach programistycznych zainteresowanie i chęć wykorzystania w bieżącej pracy. Niniejszy artykuł pokazuje, jak zaspokoić ten „głód” informacji i jednocześnie wesprzeć społeczności programistyczne w tworzeniu innowacyjnych rozwiązań. Można to osiągnąć tworząc ekosystemy informacyjne skierowane do programistów. Generyczny model takiego ekosystemu został przedstawiony w tym artykule. Jego prezentację poprzedzono charakterystyką procesów innowacyjnych oraz krótkim przeglądem istniejących ekosystemów, które zawierają komponenty przeznaczone dla programistów.

## 2. Procesy innowacyjne

Innowacje i innowacyjność są przedmiotem licznych debat, seminariów, publikacji, konferencji i badań naukowych. Obecność tematów poświęconych innowacyjności w dyskursie publicznym i branżowym wynika z tego, że innowacje w czasach postępującej globalizacji, rozwoju środków komunikacji, opartych na Internecie stacjonarnym i mobilnym,



narastającej konkurencyjności, przemożnego znaczenia rynków finansowych i ogromnych postępów nauki i techniki stały się jednym z głównych czynników pozwalających na uzyskanie przewagi porównawczej i/lub konkurencyjnej, a niekiedy wręcz czynnikiem decydującym o utrzymaniu się na rynku. W szerszym planie uważa się, że innowacyjność jest motorem nowoczesnej gospodarki, co prowadzi do konkluzji, że państwo powinno nie tylko tworzyć ramy prawne, finansowe, organizacyjne, edukacyjne i promocyjne do wspierania procesów innowacyjnych, ale powinno też uczynić zagadnienia innowacyjności stałym elementem swojej strategii i polityki gospodarczej i społecznej. Ten drugi wymiar, czyli wymiar społeczny, nie pojawia się tu przypadkowo – innowacyjność bowiem nie dotyczy wyłącznie spraw, urzędzeń i systemów technicznych; jest dla niej także miejsce w opracowywaniu, realizacji i ulepszaniu procesów i instytucji społecznych i politycznych w skali makro, oraz w organizowaniu i wielowymiarowym wspieraniu społeczności i inicjatyw lokalnych.

Istnieje szeroka gama definicji terminu „innowacje” i wielość podejść do zagadnień innowacyjności. Za szczególnie trafne przyjmuje się w tej pracy krótkie określenie, które powiada, że innowacją jest nowy produkt, usługa, lub zastosowanie, które powstały w wyniku nowatorskiego połączenia elementów już istniejących, lub też nowe zastosowanie dostępnych urządzeń czy rozwiązań. Przykładem innowacji jest smartfon, który skonstruowano z wcześniej znanych i istniejących podzespołów, ale nie jest innowacją na przykład tranzystor, który słusznie ma status wynalazku, a więc czegoś zupełnie nowego w momencie swego pojawienia się w domenie elektroniki. Warto zatem odróżniać innowacje od wynalazków i odkryć naukowych. Te ostatnie są w swej naturze różne i od innowacji, i od wynalazków – odkrycia naukowe bowiem dotyczą zjawisk i procesów oraz ich opisu w postaci praw i zasad. Mogą one, rzecz prosta, być pomocne w pracy nad innowacjami, czy też stanowić inspiracje dla wynalazców, ale ich etiologia jest zupełnie inna – są one w gruncie rzeczy motywowane ciekawością poznawczą, a nie powodami użytecznymi. Jeśli zaś chodzi o granicę pomiędzy innowacjami i wynalazkami, to bywa ona rozmyta i może być przedmiotem sporu. O ile jednak wynalazczość i odkrycia naukowe są w zasadzie procesami spontanicznymi, trudno poddającymi się kontroli i rygorystycznemu zarządzaniu (często są wręcz dziełem przypadku), to procesy innowacyjne można traktować jako działania „wstecznego” projektowania, które rozpoczyna się „od końca”, to znaczy od określenia potrzeby na nowy przedmiot (produkt, usługę lub zastosowanie), czyli w gruncie rzeczy od specyfikacji finalnego przedmiotu, i następnie krok po kroku, drogą dekonstrukcji rozłożenie go na składniki elementarne, którymi są dostępne komponenty, procesy i technologie uwzględniając przy tym czynniki ergonomiczne, finansowe, kulturowe i inne. W świecie korporacyjnym procesy innowacyjne są zazwyczaj przedmiotem starannego planowania, zarządzania, kontroli, oceny i marketingu, czego znakomite przykłady dostarczają m.in. firmy Apple i Facebook.

W literaturze dotyczącej innowacyjności istnieje mnogość modeli procesów innowacyjnych. Ta mnogość *notabene* świadczy o tym, że nie istnieje jeden, uniwersalny model i jedno podejście do innowacyjności przydatne we wszystkich okolicznościach i dla wszystkich organizacji, co zresztą nie powinno dziwić, gdyż innowacyjność, to proces o dużej różnorodności i złożoności. Jedną z lekcji badań nad innowacjami jest to, że każdy projekt innowacyjny jest z natury jednostkowy, specyficzny i wymaga przystosowania ogólnych zasad i dobrych praktyk dotyczących innowacyjności do lokalnych uwarunkowań oraz do kultury organizacji, w której ma miejsce, uwzględniając przy tym trendy i ograniczenia narzucone przez otoczenie, którym w przypadku innowacji przeznaczonych do komercjalizacji jest

rynek. Jest jednak jedna rzecz wspólna dla udanych projektów innowacyjnych prowadzonych w świecie korporacyjnym i w organizacjach publicznych, a mianowicie dokładny plan działania i rygorystyczne kontrolowanie jego wykonania. Praktyka pokazuje, że wprowadzanie innowacyjnych rozwiązań często wywołuje opór środowiska, głównie z powodu konieczności zmiany przyzwyczajeń, nawyków, nauczenia się nowych procedur, czy wręcz na skutek naruszenia lokalnych relacji i interesów. Z tego powodu istotne jest, aby zarówno prace nad innowacjami, jak i wdrażanie innowacji miały wsparcie i mocną autoryzację grupy kierowniczej i zarządczej organizacji.

B. Jacobfeuerborn i M. Muraszkiewicz (2012) zaproponowali liniowy model procesu informacyjnego ze sprzężeniem zwrotnym. Jest to dobry punkt wyjścia do rozszerzenia procesu modelowania – oto jego przebieg: pracę nad innowacją (wszystko jedno czy będzie to nowy produkt, usługa, czy zastosowanie) rozpoczyna się od zdefiniowania „zadania innowacyjnego”, które określone jest przez „potrzebę innowacyjną” zamawiającego. W zależności od natury i złożoności zadania jego język opisu jest bardziej lub mniej sformalizowany. Do wykonania zadania zostaje powołany zespół (w szczególności jednoosobowy), wyposażony w kompetencje dotyczące samej innowacji, jak i środowiska, do którego ma zostać wprowadzona. Zespół dokonuje morfologii innowacji, rozkładając jej przedmiot na czynniki prostsze oraz, stosując techniki generowania pomysłów i ich oceny (np. drogą burzy mózgów i filtrowania pomysłów), proponuje zbiór alternatywnych rozwiązań zadania innowacyjnego. Następnie przeprowadzana jest kwerenda badawcza w dostępnych zasobach informacyjnych, naturalnie z uwzględnieniem zasobów Internetu, w celu możliwie wszechstronnego pozyskania informacji o składnikach proponowanych rozwiązań (cena, patenty, licencje itp.) oraz – w przypadku innowacji przeznaczonych do komercjalizacji – informacji o sytuacji na rynku, do którego jest ona adresowana. Kolejnym etapem jest zbudowanie prototypu i jego ewaluacja. Po pozytywnym zakończeniu tego etapu innowacja gotowa jest do wdrożenia, w szczególności do komercjalizacji, jeśli ma charakter rynkowy. Trzeba zauważyć, że we wszystkich fazach procesu (z wyjątkiem fazy definiowania i komercjalizacji) możliwe jest cofnięcie się do fazy wcześniejszej, jeśli stwierdzono przeszkody lub błędy w pracy nad innowacją. W praktyce zaproponowany wyżej model może być przedmiotem różnego rodzaju rozszerzeń, uszczegółowień i mutacji w zależności od przeznaczenia. Jednak już w obecnej postaci dobrze modeluje na poziomie ogólnym trzy rodzaje procesów innowacyjnych, a mianowicie:

- (1) Ambitne, znaczące innowacje, które powstają jako wynik dużych projektów naukowo-badawczych, prowadzonych w profesjonalnych laboratoriach i w ośrodkach badawczych finansowanych przez państwo lub korporacje. Są to przedsięwzięcia, w których mają miejsce badania naukowe, często angażujące ośrodki uniwersyteckie. Ponieważ badania naukowe są kosztowne i z definicji obciążone relatywnie dużym ryzykiem niepowodzenia, proces innowacyjny oparty na tych badaniach wymaga znacznych nakładów finansowych i jest przedsięwzięciem ryzykownym, narażonym na porażkę.
- (2) Innowacje usprawniające istniejące produkty, usługi lub zastosowania. W środowisku korporacyjnym są to rutynowe działania, których celem są lokalne ulepszenia pozwalające na utrzymanie przewagi porównawczej na rynku. Poziom ryzyka i nakłady finansowe na takie działania są zazwyczaj niewielkie.
- (3) Innowacje z przysłowiowego „garażu”, realizowane przez pojedyncze osoby, startupy i małe/średnie przedsiębiorstwa. Są to niewielkie inicjatywy (choć, jak pokazuje

historia, mogące przynieść przełomowe produkty, czego sztandarowymi przykładami są komputer osobisty i wyszukiwarka Google), niewymagające dużych nakładów finansowych i złożonych form i procedur organizacyjnych.

Produkcja innowacji w społecznościach programistycznych opiera się na wariancie (c) wyżej przedstawionego modelu. Pojedynczy programiści lub niewielkie zespoły programistów, często tworzących startup, działają właśnie według modelu „innowacje z garażu”. Jednakże istotne i wychodzące daleko poza klasyczny model innowacji z garażu jest to, iż uczestnicy społeczności programistycznej należą do ekosystemu, w którym szczególną rolę odgrywają przepływy informacji. Podmioty ekosystemu, w zależności od okoliczności, współpracują lub konkurują ze sobą, tworząc w ten sposób własną kulturę, której cechy przypominają opisy znane z amerykańskiej Doliny Krzemowej. Model ekosystemu informacyjnego społeczności programistycznych, z uwzględnieniem czynnika innowacyjnego, zostanie przedstawiony w dalszej części tego artykułu.

W odniesieniu do społeczności programistycznych szczególne znaczenie ma koncepcja *otwartych innowacji*. Otwarte innowacje, o czym przekonują autorzy artykułu (Hartmann & Trott, 2009), wprowadzono do dyskusji już w latach 70. XX w. (Griffiths & Pearson, 1973; Pearson et al., 1979), a potem do teorii i praktyki innowacyjności wprowadził je H. Chesbrough w pierwszej dekadzie XXI w. (Chesbrough, 2003; Chesbrough et al., 2006). W odróżnieniu od podejścia „ogrodu za murem” (ang. *walled garden*), praktykowanego powszechnie w ubiegłym stuleciu, według którego prace nad innowacjami prowadzone są w zamkniętym obszarze firmy czy w organizacji i mają status poufnych, głównie z powodu obawy przed konkurencją, podejście otwartych innowacji zakłada, że w procesie prac nad nowymi rozwiązaniami biorą udział podmioty zaproszone spoza środowiska inicjatora zadania innowacyjnego. Mogą to być laboratoria i ośrodki badawcze, organizacje pozarządowe, agencje regulujące rynek powołane przez państwo czy reprezentanci przyszłych użytkowników innowacji. Na tak rozumianej otwartej innowacyjności B. Jacobfeuerborn oparł swój model ekosystemu otwartej innowacyjności (Jacobfeuerborn, 2012). W modelu tym, oprócz wymienionych już uczestników, są jeszcze dwa elementy, a mianowicie „wczesni użytkownicy” oraz „fabryka innowacji”. Ten pierwszy termin wprowadził do literatury E. M. Rogers (1962). Są to osoby i/lub organizacje, które jako pierwsze uzyskują – kupując na rynku lub otrzymując od innowatora na podstawie specjalnych relacji (na przykład do testowania) – produkt lub usługę innowacyjną. Rola wczesnych użytkowników jest szczególnie istotna, oni bowiem ukształtują opinie o produkcie, co wpłynie na sukces lub jego brak na rynku. Wczesnymi użytkownikami są niekiedy prosumenci, którzy współpracowali z innowatorem w czasie definiowania problemu i generowania pomysłów. Tu warto dodać, że choć znalezienie się w grupie wczesnych użytkowników może środowiskowo być przedmiotem swego rodzaju prestiżu (czego wymownym przykładem jest wczesne pozyskanie produktów firmy Apple), czy uzyskaniem przewagi porównawczej, to jednak wiąże się to z możliwością pewnego ryzyka finansowego i zawodu „moralnego” związanego z tym, że produkt nie spełnia oczekiwań, nierzadko bowiem bywa, że firmy pod naciskiem konkurencji wprowadzają na rynek produkty w wersji beta, a więc nie do końca przetestowanej, a nawet niekompletnej, co do funkcjonalności.

Drugi element, czyli „fabryka innowacji” odnosi się do specjalistów, środków technicznych, finansowych i organizacyjnych, które organizacja podejmująca zadanie innowacyjne stawia do dyspozycji zespołu innowacyjnego. B. Jacobfeuerborn wymienia tu m.in.: własne

zasoby informacyjne organizacji w postaci różnorodnych baz danych oraz wiedzy i doświadczenia zawodowego pracowników organizacji, środki i techniki wywiadu gospodarczego, w tym zaawansowane heurystyki przeznaczone do wyszukiwania informacji w Internecie, metody analizy patentowej, a także narzędzia do prototypowania oraz systemy komunikacji z klientami: CRM (Customer Relationship Management System) oraz CEM (Customer Experience Management System), jeśli organizacja takowe posiada (Jacobfeuerborn, 2012).

Powyższa uwaga, dotycząca ryzyka związanego z innowacjami, zachęca do poczynienia szerszego komentarza. Otóż za sprawą swoistej „propagandy” na rzecz innowacyjności uprawianej przez środki masowego przekazu, polityków i działaczy gospodarczych i samorządowych terminy „innowacje” i „innowacyjność” zostały w powszechnym odczuciu nacechowane pozytywnie, co zresztą należy uznać za efekt ogólnie korzystny. Umknęło jednak uwadze to, że innowacyjność oprócz korzyści może być źródłem kłopotów, a niekiedy nawet poważnych zagrożeń. Więcej na ten temat, a w szczególności na temat zagrożeń można znaleźć w artykule Jacobfeuerborna i Muraszkiewicza (2013). Tutaj wśród zagrożeń warto wymienić to, że innowacje, zwłaszcza w zakresie robotyki i sztucznej inteligencji mogą doprowadzić do zlikwidowania wielu miejsc pracy, a nawet całych zawodów, czego współczesnym przykładem są pojazdy autonomiczne (bez kierowcy), które prawdopodobnie spowodują utratę pracy milionów kierowców samochodów dostawczych i autobusów komunikacji miejskiej na całym świecie. Innym niekorzystnym efektem jest to, że nawet niewielkie innowacje, polegające na przykład na zmianie kształtu czy elementów interfejsu użytkownika (dobrym przykładem są tu smartfony) powodują, że wciąż użyteczne dotychczasowe produkty są zastępowane przez ich nowe wersje. Efekt ten już dawno dostrzegł Joseph Schumpeter, jeden z pierwszych badaczy innowacyjności i jej roli w kapitalizmie (Schumpeter, 1989). Odkładając na bok dyskusję na temat tego, że nieustająca i rosnąca konsumpcja jest motorem współczesnego kapitalizmu, nie ma żadnych wątpliwości, że ten w gruncie rzeczy pozorny rodzaj innowacji, prowadzi do nieuzasadnionego wyczerpywania zasobów naturalnych planety.

### 3. Ekosystemy społeczności programistycznych

W jednym z wcześniejszych artykułów S. Grabowskiego (2016) znajduje się opis i analiza wybranych sieciowych ekosystemów działających w obszarze technik i usług teleinformatycznych oraz opis społeczności programistycznych, które wytworzyły się wokół tych ekosystemów. Tutaj dokonamy obszerniejszego przeglądu wybranych ekosystemów.

#### 3.1. *DT Developer Garden*

DT Developer Garden działał w latach 2009–2015 i w swym zamierzeniu był projektem niemieckiego operatora telekomunikacyjnego Deutsche Telekom, który udostępnił firmom i programistom otwarte API<sup>1</sup> i inne narzędzia programistyczne. Pierwotnie DT Developer

<sup>1</sup> API (ang. *Application Programming Interface*), czyli interfejs programowania aplikacji. Jest to sposób (zrealizowany zwyczaj w formie oprogramowania) zapewniający komunikację pomiędzy aplikacją korzystającą i/lub współdziałającą z systemem, którego funkcjonalnym i/lub ergonomicznym rozszerzeniem jest ta aplikacja.

Garden działał jako punkt kontaktowy dla kreatywnych programistów, udostępniający interfejsy API oferowane przez Deutsche Telekom w celu realizacji usług dla stron internetowych i aplikacji, takich jak mashupy, widżety i wtyczki. Potem przekształcił się w program, którego celem było przyspieszenie cyklu tworzenia oprogramowania ułatwiającego tworzenie nowych aplikacji. Oprócz tradycyjnych API „telekomunikacyjnych” obsługujących wiadomości SMS i MMS oraz połączenia głosowe, DT Developer Garden udostępniał narzędzia do transkrypcji mowy na tekst i usługi geolokalizacyjne. Uzupełnieniem zestawu usług oddanego do dyspozycji społecznościom programistów były narzędzia do zarządzania i rozwoju aplikacji oraz aplikacje do testowania wydajności i analizy kodu. Partnerami DT Developer Garden były takie społeczności i grupy zorganizowane lub wspierane m.in. przez duże firmy, jak Microsoft Developer Network (MSDN), Intel Developers Services, IBM Mobile Foundation, czy M2M Developer Community.

### ***3.2. Ribbit British Telecom i Twilio***

Założona w 2006 r. firma Ribbit podeszła do problemu dostarczania usług komunikacyjnych inaczej niż jej ówczesni konkurenci. Celem firmy było udostępnienie programistom tworzącym aplikacje platformy pozwalającej na interakcję z sieciami komórkowymi operatorów telekomunikacyjnych. Z jej pomocą programiści mogli włączać do swoich stron internetowych na przykład usługi komunikacji głosowej i wysyłania wiadomości tekstowych. Użytkownicy mogli również realizować połączenia telefoniczne przez Internet do wybranego numeru za pośrednictwem prostego widżetu. Ribbit umożliwił implementację za pomocą otwartych API następujących usług: obsługę wiadomości SMS i MMS, obsługę poczty głosowej oraz usługi głosowe. Programiści mogli korzystać z następujących technik narzędziowych: Adobe Flash, Java, JavaScript, PHP, .NET, Silverlight, RESTful API. Rabbit zamknął swoją platformę w 2011 r., rekomendując jej użytkownikom przeniesienie aktywności do platformy firmy Twilio, która jest firmą specjalizującą się w komunikacji opartej na chmurze obliczeniowej. Udostępniane przez Twilio API pozwala programistom tworzyć aplikacje umożliwiające wysyłanie i odbieranie wiadomości tekstowych, a także wykonywanie i odbieranie połączeń głosowych. Platforma Twilio działa w modelu PaaS (Platform-as-a-Service), a opłata za korzystanie z jej usług jest naliczana w zależności od czasu użycia i wielkości wykorzystanych zasobów. Z rozwiązań udostępnianych przez Twilio może korzystać każdy z zastrzeżeniem, że nie wykorzysta ich do rozsyłania spamu.

### ***3.3. AT&T Developers portal***

AT&T jest jednym z największych przedsiębiorstw telekomunikacyjnych na świecie. Z jego laboratoriów wyszły takie urządzenia i rozwiązania, jak: tranzystory, system operacyjny Unix, czy języki C, C++. AT&T było jedną z pierwszych firm, które już od 1996 r. udostępniały przez Internet narzędzia programistyczne, pakiety SDK (Software Development Kit) oraz podzbiory kodu źródłowego. Od 2012 r. program współpracy z programistami został rozszerzony przez dodanie języka HTML5 oraz oprogramowania ułatwiającego tworzenie systemów i aplikacji mobilnych. AT&T Developers Program, równoległe z oferowaniem technicznego wsparcia programistom, co polega m.in. na udostępnianiu API oraz narzędzi do sprawniejszego, wygodniejszego czy bardziej zaawansowanego tworzenia aplikacji oraz



na udostępnianiu różnego rodzaju opracowań i dokumentów, organizuje grupy zrzeszające osoby zainteresowane i amatorsko praktykujące programowanie i tworzenie aplikacji.

### 3.4. *Open Middleware 2.0 Community*

Spółeczność Open Middleware 2.0 Community powstała w 2011 r. w ówczesnym Centrum Badawczo-Rozwojowym Orange Labs jako inicjatywa S. Grabowskiego i jego współpracowników. Celem tej inicjatywy było stworzenie mechanizmu budowania społeczności programistów i umożliwienie im bezpłatnego korzystania z zaawansowanych funkcji sieci Orange, dostępnych w formie otwartego API. W ramach Open Middleware 2.0 Community młodzi programiści, głównie studenci, tworzą własne aplikacje z obszaru telekomunikacji i informatyki. W ciągu kilku lat działalności opracowano kilkadziesiąt innowacyjnych rozwiązań, m.in. aplikację dla osób wymagających opieki, społecznościową grę miejską, usługę dla pasażerów komunikacji publicznej, system parkingowy oraz system do obsługi imprez publicznych. We współpracy ze szkołami wyższymi powstało wiele prac inżynierskich i magisterskich oraz kilkadziesiąt publikacji naukowych. Obecnie Open Middleware 2.0 Community współpracuje z następującymi uczelniami: Politechniką Warszawską, Politechniką Łódzką, Uniwersytetem Warmińsko-Mazurskim, Uniwersytetem Ekonomicznym w Poznaniu, Uniwersytetem Mikołaja Kopernika, Politechniką Gdańską, Politechniką Krakowską Politechnika Śląską, Uniwersytetem Jagiellońskim oraz Polsko-Japońską Akademią Technik Komputerowych.

### 3.5. *Ekosystem informacyjny BIHAPI (Business Intelligence Hackathon Application Programming Interfaces)*

BIHAPI jest rozszerzeniem programu Open Middleware 2.0 Community. Jest to ekosystem, którego uczestnikami oprócz programistów, inwestorów, przedstawiciele środowiska akademickiego oraz firm z obszaru technik teleinformatycznych, udział biorą przedstawiciele władz miejskich (m.in. Warszawy i Krakowa) oraz aktywiści miejscy. Ekosystem BIHAPI jest pomyślany i działa jako cykliczny konkurs skierowany głównie na tworzenie aplikacji wspierających transformację miast w stronę tzw. *smart cities*. Pierwsza edycja konkursu została zorganizowana w 2013 r. we współpracy z: firmą Oracle Communications jako partnerem technologicznym, IQPartners jako partnerem inwestycyjnym oraz z Urzędem Miasta Warszawy i Politechniką Warszawską. Głównymi celami konkursu były: (i) popularyzacja otwartych API jako narzędzia do tworzenia aplikacji mobilnych i serwerowych (Grabowski, 2014), (ii) promocja usług telekomunikacyjnych dostępnych przez API, (iii) wykorzystanie oraz udostępnianie otwartych danych znajdujących się w posiadaniu polskich miast, (iv) budowa środowiska społecznego skupionego wokół otwartych API oraz otwartych danych.

## 4. *Generyczny model ekosystemu informacyjnego społeczności programistycznych*

Spółeczności programistyczne różnią się od wielu innych tworzących się oddolnie społeczności. Różnica nie polega jednak na tym, że ich etiologia, struktura czy sposób

funkcjonowania są diametralnie różne od odpowiadających im komponentów w innych społecznościach. Polega ona na specyficznej kompozycji, konfiguracji, wewnętrznych relacjach i *modus operandi* społeczności programistycznych. Współcześnie powstają spontanicznie lub inspirowane przez instytucje państwowe, publiczne i organizacje gospodarcze, w tym *par excellence* biznesowe, heterogeniczne społeczności programistyczne oparte na sieci, głównie na Internecie stacjonarnym i mobilnym. W skład tych społeczności wchodzi programiści indywidualni, nieformalne zespoły programistów, a także firmy komercyjne, głównie startupy i małe przedsiębiorstwa. W dalszej części artykułu właśnie te elementy zostaną wykorzystane do stworzenia generycznego modelu ekosystemu informacyjnego społeczności programistycznych tworzących się i działających głównie w dużych środowiskach miejskich.

#### 4.1. Podstawowe założenia modelu generycznego

Programowanie, zwłaszcza większych systemów, jest czynnością, a właściwie procesem o dużej złożoności, wymagającym specjalistycznej wiedzy, której składnikami są: (i) zbiór tzw. kompetencji twardych, którymi głównie są: umiejętność rozumienia algorytmów, praktyczna znajomość języków programowania i związanych z nimi bibliotek komponentów programistycznych oraz dobrych praktyk, kodowania algorytmów w wybranym języku i testowania utworzonych programów, oraz (ii) tzw. kompetencje miękkie, którymi są m.in.: cierpliwość, odporność na stres i zdolność do współpracy zarówno z innymi programistami, jak i z innymi, nieinformatycznymi podmiotami. Utrzymanie na wysokim poziomie i rozwijanie kompetencji twardych wymaga silnego i wszechstronnego wsparcia informacyjnego w zakresie technik informatycznych i inspirujących studiów przypadków, zarówno tych pozytywnych, jak i negatywnych. Co do kompetencji miękkich, to wsparcie informacyjne przyda się z pewnością w zakresie technik kreatywnej pracy zespołowej nakierowanej na rozwiązywanie problemów oraz budowania, utrzymywania i rozszerzania siatki współpracy z otoczeniem. Na marginesie warto odnotować zastanawiający fakt: programiści, to często absolwenci wyższych studiów technicznych, którzy uczęszczali na zajęcia z programowania, w tym na zajęcia z zakresu inżynierii oprogramowania, poświęcone m.in. metodom i technikom tworzenia systemów informatycznych i informacyjnych, a jednocześnie – jak pokazuje praktyka – rzadko korzystają z tej wiedzy w trakcie prowadzonych przez siebie projektów, zdając się na intuicyjne sposoby organizowania pracy. Prowadzi to do nieefektywnego wykorzystania zasobów i zmniejsza szanse na osiągnięcie konkurencyjnych wyników. Słabość tę w dużym stopniu może zniwelować dołączenie programistów do istniejących sieci programistów i wykorzystanie przez nich dostępnych zasobów informacyjnych i kompetencyjnych sieci.

Jak wspomniano, społeczności programistyczne i ich związki tworzą się raczej na obszarach metropolitalnych – ich naturalnym środowiskiem są miasta, zwłaszcza miasta przekształcające swą tkankę miejską w miasto inteligentne ze wszystkimi wynikającymi z tego korzyściami, wśród których szczególną rolę odgrywają: dobrze rozwinięta infrastruktura teleinformatyczna, efektywne, wykorzystujące tę infrastrukturę i stosujące nowe metody zarządzania, władze municypalne oraz duża liczba organizacji i inicjatyw pozarządowych. Władze miejskie we współpracy z tymi organizacjami działają na rzecz tworzenia, utrzymania i rozwoju dóbr wspólnych, z których specjalne znaczenie dla

środowisk programistycznych mają ogromne zasoby otwartych danych dotyczące zjawisk, procesów, przedsięwzięć, wydarzeń, historii, miejsc i ludzi – przykładem takich danych były 53 zbiory danych (m.in. dotyczące komunikacji miejskiej, rozkładów jazdy, parkowania, sportu, edukacji, jednostek użyteczności publicznej), z których korzystali uczestnicy projektu „Dane po Warszawsku”<sup>2</sup>. Jednocześnie przedsięwzięcia podejmowane przez różne prywatne, publiczne i państwowe podmioty w modelu otwartych innowacji, nawet jeśli nie są to inicjatywy *par excellence* informatyczne, potrzebują rozwiązań i/lub wsparcia informatycznego i programistycznego, najczęściej jako specjalistycznych aplikacji. I tu właśnie jest miejsce i rola do odegrania, a właściwie już odgrywana, przez społeczności programistyczne.

Powyższe stwierdzenia upoważniają do rozszerzenia przedstawionego wyżej Jacobfeurborna modelu ekosystemu otwartej innowacyjności do postaci diagramu pokazanego na rysunku 1.



Rys. 1. Model ekosystemu otwartej innowacyjności

W modelu tym pojawiły się nowe elementy, którymi są miasto i społeczności programistyczne oraz otwarte oprogramowanie i otwarte dane, zaś miejsce fabryki innowacji zajął inkubator innowacji. O otwartym oprogramowaniu i otwartych danych traktuje bardzo duża liczba publikacji. Tutaj warto zwrócić uwagę na to, że otwarte dane, zarówno jako dane faktograficzne, jak i dane tekstowe oraz multimedialne są generowane przez uczelnie, ośrodki badawcze, laboratoria, organizacje pozarządowe, agencje (rządowe) regulujące rynki oraz przez administrację miejską, ale także przez społeczności programistyczne, które korzystając z danych pozyskanych z tych źródeł tworzą nowe zasoby danych, którym w wielu przypadkach nadają status danych otwartych. Jeśli chodzi o otwarte oprogramowanie, to jest ono tworzone głównie w uczelniach i ośrodkach badawczych, organizacjach pozarządowych oraz naturalnie w środowiskach programistycznych, niekiedy związanych z służbami informatycznymi miast.

<sup>2</sup> <http://www.danepowarszawsku.pl/>



Społeczności programistyczne jako samodzielny podmiot występujący w strukturach innowacyjnych, zwłaszcza tych związanych z funkcjonowaniem miast, są relatywnie nowym zjawiskiem. Jaskółką, która zwiastowała ich pojawienie się były hackathony, które można traktować jako prefigurację współczesnych społeczności i sieci programistycznych. Na marginesie warto wspomnieć, że za pierwszy hackathon uważa się wydarzenie OpenBSD, które miało miejsce w Calgary w 1999 r. Wikipedia pod hasłem *hackathon* podaje, że jest to

wydarzenie skierowane do programistów, podczas którego informatycy i inne osoby związane z rozwojem oprogramowania, takie jak projektanci grafiki, twórcy interfejsów i menedżerowie projektów, stają przed zadaniem rozwiązania określonego problemu związanego z projektowaniem. Hackathony odbywają się w krótkim czasie, zazwyczaj na przestrzeni dnia lub weekendu. Zadanie do wykonania ogłaszane jest w dniu rozpoczęcia konkursu. Podczas oceniania pod uwagę brana jest wyłącznie praca wykonana podczas trwania wydarzenia (Hackathon, 2016).

Hackathony były częścią ekosystemów, opisanych w sekcji trzeciej: Deutsche Telekom Developer Garden, Twilio, Open Middleware 2.0 Community oraz BIHAPI. Z pewnością rola hackathonów w tworzeniu innowacyjnych rozwiązań będzie rosła i choćby z tego tytułu będą one przedmiotem dalszych badań.

Biorąc pod uwagę podane wyżej fakty, obserwacje i wnioski, można pokusić się o stworzenie generycznego modelu ekosystemu informacyjnego dla innowacyjnych społeczności programistycznych. Trzy aspekty tych społeczności są szczególnie istotne, a mianowicie: sieciowość, otwartość i „informacyjność”. Model ten przedstawiono poniżej w dwóch aspektach: informacyjnym (źródła informacji) i procesualnym.

#### 4.2. *Informacyjny aspekt modelu generycznego*

Głównymi (a więc nie jedynymi) źródłami informacji i wiedzy, z których korzystają społeczności programistyczne są: otwarte dane, otwarte oprogramowanie oraz crowdsourcing, to bowiem te źródła są dużą i ważną częścią tego, co można nazwać „informacyjnym dobrem wspólnym”. Otwarte dane pochodzą z różnorodnych miejsc, wśród których najsilniejszymi generatorami danych są procesy zachodzące i rejestrowane informacyjnie w miastach, w różnorodnych placówkach rządowych i samorządowych, państwowych uczelniach, instytutach naukowych i laboratoriach badawczych oraz w całej rzeszy organizacji pozarządowych. Dane te tworzone i przechowywane są w różnych formatach oraz dostępne są za pomocą różnych narzędzi i procedur informatycznych, w tym procedur ochrony danych. Jest to zatem niezwykle heterogenna przestrzeń informacyjna, co bardzo utrudnia dostęp i korzystanie z niej, na co nierzadko nakłada się niechęć dysponentów danych do ich udostępnienia, pomimo istnienia formalnego obowiązku w tym względzie. Dodatkowe problemy związane z otwartymi danymi wiążą się z tym, że – po pierwsze – zazwyczaj nie są one uporządkowane i ustrukturyzowane (zbiory danych zapisane są w postaci tekstowej, np. w formacie DOCX lub PDF, czy tabelarycznej w formacie XLSX, a nie na przykład według otwartego standardu *OpenDocument*, czy otwartej relacyjnej bazy danych w rodzaju SQLite, MySQL czy PostgreSQL, czy serwera i wyszukiwarki tekstowej Sphinx) oraz – po drugie – brakiem metadanych opisujących zawartość zbiorów danych, ich pochodzenie, tryb aktualizacji, kontakt do administratora itp. Metadane nie tylko ułatwiają korzystanie ze zbiorów danych, ale bywa, że ich brak uniemożliwia korzystanie z samych danych. Trzeba podkreślić, że społeczności programistyczne korzystając z otwartych danych często same

tworzą nowe zbiory otwartych danych, które są wynikiem przetworzenia, na przykład zagregowania danych źródłowych albo wyselekcjonowania z nich podzbiorów relewantnych do określonej potrzeby informacyjnej.

Abstrahując od dyskusji na temat różnic pomiędzy otwartym oprogramowaniem, a wolnym oprogramowaniem, najważniejsze jest to, że użytkownik ma dostęp do kodu źródłowego i może dokonywać w nim modyfikacji w zależności od swoich potrzeb. Więcej, respektując licencję (zwykle dość liberalną) nałożoną na ten kod, może upowszechnić swoją wersję. W ten sposób społeczności programistyczne, udostępniając opracowane przez siebie programy, kolektywnie pracują nad rozwojem oprogramowania. Warto odnotować, że ten tryb pracy i dyfuzji oprogramowania nie wyklucza jego pewnych zastosowań komercyjnych.

Trzecim elementem informacyjnego dobra wspólnego, wchodzącym w skład modelu ekosystemu informacyjnego społeczności programistycznych, jest crowdsourcing. Ta forma pozyskiwania wiedzy z otoczenia, które na tzw. otwarte zapytanie (ang. *open call*) reaguje w postaci zbioru propozycji rozwiązań zadania sformułowanego w zapytaniu jest wykorzystywana przez społeczności programistyczne w sytuacji, gdy nie posiadają one kompetencji do rozwiązania problemu, którym się zajmują. I odwrotnie, to dany zespół programistyczny, jeśli posiada potrzebne kompetencje albo tylko dobrze uzasadnione i realistyczne pomysły, może proponować rozwiązania problemów sformułowanych przez inne podmioty.

Oto inne, ważne źródła informacji i wiedzy wykorzystywane przez społeczności programistyczne w ekosystemie informacyjnym. Niewątpliwie społeczności te dysponują własnymi zasobami informacyjnymi, swoim *know-how* i regułami zarządczymi. Elementy te są wynikiem doświadczenia i wraz z rozwojem społeczności nabierają coraz większego znaczenia w kolejnych projektach przez nie realizowanych. W przypadku większych projektów pojawia się nierzadko moment, w którym trzeba skorzystać z technik wywiadu gospodarczego w celu zidentyfikowania komponentów potrzebnych do realizacji tych projektów. Zwykle społeczność robi to siłami własnymi, ale bywa też, że korzysta z kompetencji współpracujących z nią innych podmiotów, na przykład tych, które należą do ekosystemu otwartych innowacji albo sięga po ekspertów zewnętrznych (którymi często są osoby z innych społeczności programistycznych) lub po prostu zamawia taką usługę w wyspecjalizowanej placówce. Wynikiem tych działań najczęściej są dane rynkowe, dane gospodarcze, informacje o patentach, licencjach i innych kwestiach z zakresu praw własności intelektualnej, dane o technologiach, producentach, użytkownikach itp. Ważnymi źródłami informacji są prosumenci, czyli osoby, grupy osób czy nawet firmy, które współpracują ze społecznością programistyczną już w fazie testowania, a niekiedy nawet projektowania produktu. To oni właśnie są głównymi użytkownikami wersji beta wytworzonego oprogramowania czy aplikacji. Niebagatelnym źródłem informacji, stale obecnym w pracy programistów są media społecznościowe, będące zarazem wehikułem komunikacyjnym i instrumentem promocyjnym i/lub marketingowym.

### 4.3. Procesualny aspekt modelu generycznego

Jak wspomniano, model ekosystemu informacyjnego dla społeczności programistycznych obejmuje także aspekt procesowy, czyli sposób ich tworzenia się i funkcjonowania. Społeczności takie powstają najczęściej jako grupy startupów w parkach technologicznych czy pod auspicjami uczelni wyższych. Tworzą się one także *ad hoc* na użytek realizacji dużych

projektów finansowanych z grantów przyznawanych przez agencje rządowe lub samorządowe; w Polsce są to m.in.: Narodowe Centrum Badań i Rozwoju, Polska Agencja Rozwoju Przedsiębiorczości i Urzędy Marszałkowskie. W takich przypadkach lider projektu zaprasza do udziału w nim różne zespoły programistyczne, biorąc na siebie obowiązek zdefiniowania i przydzielenia im zadań, koordynację prac oraz integrację wyników częściowych w produkt końcowy. Inicjatywę utworzenia społeczności programistycznych podejmują także korporacje we współpracy z uczelniami, czego przykładem jest społeczność Open Middleware 2.0 Community i ekosystem informacyjny BIHAPI, wspomniane w sekcji trzeciej tego artykułu.

Ekosystem BIHAPI jest o tyle interesujący, że jest dobrym przykładem środowiska do akceleracji innowacyjnych produktów informatycznych. W ramach akceleracji zwykle powstaje prototyp o statusie *proof-of-concept*, czyli rozwiązanie, które ma przekonać inwestora(ów) do zainwestowania środków finansowych w opracowanie produktu finalnego i jego komercjalizację. Najczęściej są to różnorakie aplikacje przeznaczone dla urzędzeń mobilnych takich, jak smartfony czy tablety, choć naturalnie akceleracja może dotyczyć projektowania i realizacji większych przedsięwzięć opartych na platformach cyfrowych.

Modelowo akcelerację można przedstawić jako proces składający się z następujących etapów:

- (1) Zdefiniowanie zadania innowacyjnego, czyli opracowanie specyfikacji produktu finalnego. Realizację tego zadania bierze na siebie firma lub organizacja zainteresowana utworzeniem nowego produktu.
- (2) Utworzenie ekosystemu innowacji. Jeśli zadanie innowacyjne będzie realizowane według modelu otwartych innowacji, to w składzie ekosystemu znajdują się inwestor(zy), placówki badawcze i inne wyżej wymienione podmioty.
- (3) Ogłoszenie konkursu (może on być otwarty albo zamknięty, czyli ograniczony do wybranej grupy zespołów programistycznych) na realizację zadania. Konkurs ogłasza podmiot sprawujący rolę koordynatora ekosystemu. Konkursy mogą przybierać popularną obecnie formę hackathonu.
- (4) Opracowanie produktów (najczęściej w postaci prototypów). Zespoły programistyczne, które przystąpią do konkursu tworząc produkty (aplikacje) spełniające warunki konkursu, stają się automatycznie częścią ekosystemu i uzyskują dostęp do tych jego zasobów, które udostępniają uczestnicy ekosystemu na czas realizacji zadania. Zespoły pracują niezależnie od siebie.
- (5) Ocena wyników pracy zespołów wedle kryteriów wcześniej ustalonych przez inwestorów.
- (6) Selekcja przedłożonych rozwiązań. Inwestorzy wybierają te rozwiązania (prototypy), które zamierzają przekształcić w produkty spełniające warunki komercjalizacji.
- (7) Opracowanie wersji finalnej (komercyjnej) produktów.
- (8) Komercjalizacja, czyli wprowadzenie produktów na rynek.

## 5. Zakończenie

Zaproponowany model może posłużyć jako punkt wyjścia do dalszych prac nad konstruowaniem, architekturą, funkcjami i zarządzaniem społecznościami programistycznymi, których znaczenie w tworzeniu innowacji technicznych i wspierania nowatorskich rozwiązań gospodarczych i społecznych, nabiera szczególnego znaczenia w kontekście wysiłków wyrwania się Polski z pułapki średniego rozwoju.

Warto ponownie podkreślić, że przedstawiony generyczny model ekosystemu informacyjnego innowacyjnych społeczności programistycznych, łączy w sobie trzy niezwykle ważne zjawiska współczesnej gospodarki i życia społecznego. Pierwsze z nich to sieciowy, horyzontalny charakter powiązań i interakcji w ramach zachodzących procesów w obrębie samych społeczności i procesów łączących je z otoczeniem. Drugie zjawisko, w coraz większym stopniu uświadamiane i wykorzystywane przez programistów, to rosnąca rola informacji i wiedzy w tych procesach, z naciskiem na zapewnienie otwartego i szerokiego dostępu do informacji z zachowaniem, w możliwie jak największym stopniu ochrony tożsamości użytkowników. Trzeci fenomen, który w tak dużej skali, jaki obecnie pojawił się w historii informatyki i jej zastosowań po raz pierwszy, to wola tworzenia innowacyjnych produktów i rozwiązań problemów społecznych, co wobec coraz większej presji na wykorzystanie zasobów naturalnych planety i nasilające się problemy demograficzne jest szczególnie godne odnotowania.

## Bibliografia

- Chesbrough, H. (2003). *Open Innovation: The New Imperative for Creating and Profiting from Technology*. Boston: Harvard Business School Press.
- Chesbrough H., Vanhaverberek W., West J. (2006). *Open Innovation. Researching a New paradigm*. Oxford University Press.
- Global Developer (2017). *Global Developer Population and Demographic Study 2016 V2*. Evans Data Corporation.
- Grabowski, S. (2014). What is Open Data and How to Benefit from It? *Zagadnienia Informatyki Naukowej*, 52(1), 43–51.
- Grabowski, S. (2016). Społeczności programistyczne w obszarze otwartych interfejsów programistycznych. Open API: przykłady i wnioski. W: B. Sosińska-Kalata et al. (red.), *Nauka o informacji w okresie zmian. Informatologia i humanistyka cyfrowa*. Miscellanea Informatologica Varsoviensia. vol. VIII (121–134). Warszawa: Wydaw. SBP.
- Griffiths, D., Pearson, A. W. (1973). The Organization of Applied Research and Development with Particular Reference to the Customer – Contractor Principle. *R&D Management* 3(3), 121–124.
- Hackathon (2016). Wikipedia. Wolna encyklopedia [online] [23.10.2017], <https://pl.wikipedia.org/wiki/Hackathon>
- Hartmann, D., Trott, P. (2009). Why 'Open Innovation' is Old Wine in New Bottles. *International Journal of Innovation Management* 13(4), 715–736.
- Jacobfeuerborn, B. (2012). A Knowledge Concocter to Sustain Innovation Propensity. W: Z. E. Zieliński (red.), *Rola informatyki w naukach ekonomicznych i społecznych. Innowacje i implikacje interdyscyplinarne* (40–50). Kielce: Wydaw. Wyższej Szkoły Handlowej.
- Jacobfeuerborn, B., Muraszkiewicz, M. (2012). ICT Based Information Facilities to Boost and Sustain Innovativeness. *Studia Informatica* 33(2B), 485–496.
- Jacobfeuerborn, B., Muraszkiewicz, M. (2013). A Sketchy Critique of Innovation by Advocati Dei and Diaboli. W: B. Jacobfeuerborn (ed.), *Innovating Innovation. Essays on the Intersection of Information Science and Innovation* (11–20). Warszawa: MOST Press, IINiSB UW, 11–20.
- Pearson, A. W., Green, T., Ball, D.F. (1979). A Model for Studying Organizational Effects of an Increase in the Size of R&D Projects. *IEEE Transactions on Engineering Management*, EM-26(1), 14–21
- Rogers, E. M. (1962). *Diffusion of Innovations*. Free Press of Glencoe, Macmillan Company.
- Schumpeter, J. A. (1989). *Essays: On Entrepreneurs, Innovations, Business Cycles and the Evolution of Capitalism*. New Brunswick, N.J.: Transaction Publisher.

## Modeling of Information Ecosystems for Innovative Programming Communities

### Abstract

**Purpose/Thesis:** The purpose of the paper is to propose a generic model of an information system supporting innovative communities of ICT programmers, in particular those who work in urban environments.

**Approach/Methods:** The method adopted in the study consisted in analyzing a selected set of foreign network ecosystems that boost programming communities (among others: Deutsche Telekom Development Garden and AT&T Developers Portal) and the communities co-established by Sebastian Grabowski, that is, Open Middleware 2.0 Community and BIHAPI. The analysis covered structures of these ecosystems and their *modus operandi*. Conclusions, in particular those related to the factors supporting innovation processes, helped to define a generic model of an information ecosystem for supporting innovative programming communities.

**Results and conclusions:** The examination of the analyzed ecosystems along with the authors' own experience acquired as a result of cooperation with programming communities on different projects and programs, including Open Middleware 2.0 Community and BIHAPI and numerous hackathons, proved an important role currently played and to be played in the future by the programming communities. It was discovered that particularly efficient and productive way for these communities to organize and work was the use of open platforms, with focus on open-source software, open data and open innovation.

**Application:** The proposed model of a generic information ecosystem supporting innovative programming communities provides the background for further work on the construction, architecture, functionality and management of programming communities whose role in producing technical innovations and boosting innovative economic and social solutions becomes of tremendous importance in the context of Poland's ambition to escape the middle-income trap.

**Originality/Value:** The proposed model is generic which makes it adaptable to programming environments of varying profiles and objectives. It is based on open platforms as regards data and software and puts emphasis on the work mode specific for open innovations, with an innovative approach to the programmers' work. Furthermore, the model is open which offers reliable background for further research on modelling and developing information ecosystems supporting programming communities. The value of the model discussed was confirmed in practice during Open Middleware 2.0 Community and BIHAPI programs.

### Keywords

API. Application Programming Interface. Information ecosystem. Open data. Open innovation. Open-source software. Programming community.

---

*SEBASTIAN GRABOWSKI jest dyrektorem Centrum Badawczo-Rozwojowego Orange Polska S.A. oraz Prezesem Fundacji „ArchitectsPL” zajmującej się m.in. wpływem otwartych danych na społeczeństwo. Jest współtwórcą programu Open Middleware Community zrzeszającego różne środowiska deweloperów, naukowców, dostawców technologii oraz firm IT oraz twórcą ogólnopolskiej inicjatywy poświęconej tematyce otwartych danych – BIHAPI. Jego badania dotyczą otwartych danych, interakcji pomiędzy użytkownikami sieci telekomunikacyjnych i Internetu oraz zagadnień związanych z otwartym rządem. Wśród jego ostatnich publikacji znajdują się: S. Grabowski (2014). What is Open Data and How to Benefit from It? ZIN, 52(1), 43–51; K. Litwiniuk, T. Czarnecki, S. Grabowski, J. Legierski (2012). BusStop — Telco 2.0 application supporting public transport in agglomerations. Federated Conference on Computer Science and Information Systems (FedCSIS), 9(12), 649–653; S. Grabowski, J. Legierski (2011). Interfejsy programistyczne w systemach Otwartego Rządu, PITWIN, Kielce.*

Kontakt z autorem:

[sebastian.grabowski@orange.com](mailto:sebastian.grabowski@orange.com)

Orange Labs, Centrum Badawczo-Rozwojowe

Obrzeźna 7

02-691 Warszawa

Prof. dr hab. MIECZYŚLAW MURASZKIEWICZ jest profesorem zwyczajnym w Instytucie Informatyki Politechniki Warszawskiej. Jego zainteresowania naukowe dotyczą inteligentnych systemów informacyjnych, metod reprezentacji wiedzy, informacji dla nauki oraz relacji wiążących technikę i kulturę, a także zagadnień innowacyjności. Wśród jego ostatnich publikacji znajdują się: M. Muraszekiewicz (2016). *Techniki komunikacyjne i informacyjne dla inteligentnych miast*. W: D. Gotlieb, R. Olszewski (eds.), *Smart City. Informacja przestrzenna w zarządzaniu inteligentnym miastem* (53–68). Warszawa: PWN; B. Jacobfeuerborn, M. Muraszekiewicz (2014). *Some Challenges and Trends in Information Science*. In: R. Bembenik et. al. (eds.), *Intelligent Tools for Building a Scientific Information Platform: From Research to Implementation* (3–14). Springer [Studies in Computational Intelligence No. 541]; B. Jacobfeuerborn, M. Muraszekiewicz (2013). *Media, Information Overload, and Information Science*. In: R. Bembenik et. al. (eds.), *Intelligent Tools for Building Scientific Information Platform. Advanced Architectures and Solutions* (3–13). Springer [Studies in Computational Intelligence No. 467].

Kontakt z autorem:

[m.muraszkiewicz@ii.pw.edu.pl](mailto:m.muraszkiewicz@ii.pw.edu.pl)

Instytut Informatyki

Wydział Elektroniki i Technologii Informacyjnych

Politechnika Warszawska

Nowowiejska 15/19

00-665 Warszawa



# Formaty danych strukturalnych w zasobach World Wide Web

Marcin Roszkowski

*Katedra Informatologii, Wydział Dziennikarstwa, Informacji i Bibliologii,  
Uniwersytet Warszawski*

---

## Abstrakt

**Cel/Teza:** Celem artykułu jest charakterystyka i porównanie formatów danych wykorzystywanych do strukturyzacji metadanych dotyczących treści zasobów World Wide Web w ramach systemu języka znacznikowego HTML. Przedmiotem badań są wybrane formaty danych strukturalnych (mikroformaty, RDFa, mikro dane i JSON-LD) oraz zakres ich wykorzystania na potrzeby reprezentacji informacji w środowisku WWW. Punktem wyjścia do rozważań są tezy, że rozdzielenie warstwy treści zasobów sieciowych od sposobu ich prezentacji jest jedną z fundamentalnych cech środowiska World Wide Web oraz że poziom strukturyzacji treści zasobów sieciowych jest czynnikiem determinującym zakres ich wyszukiwalności.

**Koncepcja/Metody badań:** Scharakteryzowano koncepcję formatów danych strukturalnych (FDS), która zakłada formalną reprezentację informacji o treści zasobów sieciowych bezpośrednio wewnątrz dokumentów HTML. Analizie poddano formaty danych najszerszej wykorzystane w środowisku WWW. Przyjęto również założenie, że FDS może być interpretowany jako środek ekspresji metadanych dokumentu oraz że reprezentacja informacji ma miejsce nie na poziomie samego dokumentu, ale na poziomie faktów go konstytuujących. Analiza FDS została przeprowadzona z punktu widzenia sposobów formalnej ekspresji metadanych (poziom syntaktyczny) oraz zastosowanych do opisu struktur pojęciowych wraz z ich językowymi wykładnikami (poziom semantyczny).

**Wyniki i wnioski:** FDS przeznaczone do reprezentacji treści stron internetowych zapewniają nowe możliwości strukturyzacji treści w ramach systemu znaczników języka HTML i tym samym rozszerzają zakres możliwych funkcjonalności mechanizmów wyszukiwawczych. Wyniki badań nad implementacją FDS w latach 2012–2016 pokazują wzrost zainteresowania tą formą strukturyzacji treści w zasobach WWW. Największą szczegółowość w reprezentacji wiedzy zapewnia RDFa, jednak to mikro dane wydają się być kompromisem między pożądaną siłą ekspresji a prostotą implementacji, na co wskazują wyniki badań szczegółowych uzyskanych w projekcie Web Data Commons.

**Oryginalność/Wartość poznawcza:** Przedstawione porównanie wybranych elementów FDS daje obraz ich możliwości w głębokiej strukturyzacji treści zasobów WWW, ze szczególnym uwzględnieniem wykorzystania istniejących schematów metadanych i ontologii. Analiza dokumentacji projektu Web Data Commons z lat 2014–2016 pozwala sądzić, że to mikro dane będą FDS, który będzie miał istotne znaczenie w kontekście zastosowania technologii semantycznych na potrzeby strukturyzacji treści zasobów WWW.

## Słowa kluczowe

JSON-LD. Linked Data. Metadane. Mikro dane. Mikroformaty. RDFa. Schema.org. Sieć Semantyczna.

*Otrzymano: 30 sierpnia 2017. Zrecenzowano: 16 listopada 2017. Zaakceptowano: 18 grudnia 2017.*

---



## 1. Wprowadzenie

Jedną z fundamentalnych cech World Wide Web (WWW) jest rozdzielenie warstwy prezentacji informacji od sposobu formatowania i zapisu danych ją tworzących (Guha et al., 2015). Na płaszczyźnie prezentacji informacji mamy do czynienia z interpretacją przez przeglądarki internetowe kodu stron internetowych, zapisanego najczęściej w języku HTML oraz wyświetleniem użytkownikowi treści stron WWW wraz dodatkowymi funkcjonalnościami i elementami interaktywnymi. Poziom formatowania i strukturyzacji informacji to wykorzystanie języków znacznikowych (ang. *markup language*) do organizacji stron internetowych (język HTML – HyperText Markup Language) oraz projektowania ich typografii (np. język CSS – Cascading Style Sheets). HTML jest językiem, który w założeniu ma jednak opisywać strukturę dokumentu, nie zaś jego treść. System jego znaczników odpowiada w większości za kwestie związane z formatowaniem dokumentu i sposobem prezentacji jego treści użytkownikowi za pomocą przeglądarki internetowej. Z punktu widzenia formatowania danych tworzących zawartość stron internetowych, ograniczenia języka HTML powodują, że treść tych dokumentów na poziomie semantycznym nie może być formalnie odwzorowana w celu jej automatycznego przetwarzania. Możliwości języka HTML w kontekście reprezentacji informacji o zawartości stron internetowych kończą się na metadanych dotyczących całego dokumentu, które zapisywane są w jego części nagłówkowej (sekcja <head>). Z punktu widzenia architektury sieciowych systemów informacyjnych, stosowane w nich rozwiązania bazodanowe oferują duże możliwości strukturyzacji przechowywanych danych, lecz w efekcie są one publikowane w postaci stron internetowych za pośrednictwem systemu języka HTML, gubiąc jednocześnie ich pierwotne uporządkowanie. Z punktu widzenia mechanizmów wyszukiwarek internetowych zaindeksowanie dokumentu wiąże się z automatyczną interpretacją ciągu znaków, które tworzą jego treść. Brak informacji o treści dokumentu wyrażonej w formie przetwarzalnej przez aplikację powoduje, że adekwatność indeksowania rzeczowego zależy od efektywności zastosowanych algorytmów przetwarzania języka naturalnego. Taka sama sytuacja ma miejsce w odniesieniu do narzędzi agregujących dane z WWW (takich, jak np. serwisy porównywania cen produktów). Jeżeli zastosowany w danym serwisie internetowym system zarządzania treścią (CMS – Content Management System) nie oferuje dostępu do kolekcji za pośrednictwem interfejsu programistycznego (API – Application Programming Interface), to pozyskanie wybranych treści wiąże się z koniecznością opracowania algorytmu i mechanizmu ekstrakcji danych przeznaczonych dla danego źródła internetowego.

Można więc przyjąć, że poziom strukturyzacji treści zasobów sieciowych jest czynnikiem determinującym zakres funkcjonalności wyszukiwarek internetowych, które je indeksują (Bizer et al., 2012, 3). Z drugiej strony, dążenie do zwiększenia możliwości publikowania informacji na głębokim poziomie strukturyzacji było jednym z początkowych założeń koncepcji WWW (Bizer et al., 2009, 15). Tim Berners-Lee już w 1994 r. założył, że rozwój WWW powinien uwzględnić zmianę sposobu publikowania dokumentów w sieci. Miała ona polegać na formatowaniu stron internetowych w sposób adekwatny do ich percepcji przez użytkownika oraz na zastosowaniu przetwarzalnych maszynowo informacji o charakterze semantycznym (Berners-Lee, 1994, 797). Wizja T. Bernersa-Lee doprowadziła do rozwoju koncepcji Sieci Semantycznej (ang. *Semantic Web*), w której zakładał nałożenie dodatkowej warstwy na istniejące kolekcje sieciowe, zapewniającej formalną reprezentację

informacji zapisanych w ich zasobach. Idea Sieci Semantycznej opiera się na publikowaniu w środowisku sieciowym wzajemnie powiązanych, ustrukturyzowanych zbiorów danych z wykorzystaniem przyjętych za standardowe technologii informacyjnych. Według tej koncepcji (Berners-Lee et al., 2001) w rozproszonych zasobach sieciowych reprezentacja informacji powinna mieć miejsce nie na poziomie dokumentu, lecz na poziomie faktów ekstrahowanych z jego treści. Takie podejście do tworzenia i udostępniania metadanych pozwoli na utworzenie sieci danych (ang. *Web of Data*), która bazować ma na istniejącej sieci dokumentów. Dla WWW standardem opisu zasobów sieciowych nadal pozostaje HTML, natomiast *lingua franca* Sieci Semantycznej jest język Resource Description Framework (RDF). Zaznaczyć należy, że RDF jest językiem reprezentacji wiedzy przeznaczonym do opisu zbiorów danych, nie zaś – tak jak HTML – językiem prezentacji treści stron internetowych.

Jedną z form realizacji Sieci Semantycznej jest metodyka publikowania danych o nazwie Linked Data. Zakłada ona udostępnianie całych zbiorów danych, zgodnie z przyjętymi standardami sieciowymi oraz konieczność ustalania relacji między elementami opisywanymi w tych zbiorach a zewnętrznymi źródłami informacji. Utworzona w ten sposób sieć danych nazywana jest chmurą danych powiązanych (ang. *Linked Data Cloud*). Choć jest to wiele obiecująca realizacja wizji Tima Bernersa-Lee, to problem polega na tym, że dane źródłowe muszą być konwertowane zgodnie z założeniami Linked Data oraz udostępniane w pierwszej kolejności dla aplikacji, które będą je przetwarzać. Mamy więc do czynienia z funkcjonowaniem kilku wersji tego samego źródła informacji, np. Wikipedia i jej „semantyczna wersja” DBpedia<sup>1</sup>. Istotne więc wydaje się być pytanie o status istniejących zasobów WWW w kontekście rozwoju technologii semantycznych.

Jednym z rozwiązań problemu ograniczeń HTML w zakresie formalnej strukturyzacji zawartości stron internetowych jest stosowanie formatów danych operujących tzw. semantycznymi znacznikami (ang. *semantic markup*), które stanowią rozszerzenia systemu znaczników tego języka. W odniesieniu do systemu języka znacznikowego HTML termin *znacznik semantyczny* jest rozumiany jako formalny wykładnik reprezentujący element treści dokumentu oraz jako składnik specyfikacji systemu znaczników, którego celem jest semantyczna strukturyzacja zawartości stron WWW.

## 2. Cel i metodologia badań

Celem artykułu jest charakterystyka i porównanie formatów danych, wykorzystywanych do strukturyzacji treści stron internetowych w ramach systemu języka znacznikowego HTML. Przedmiotem badań są cztery formaty danych strukturalnych (FDS; ang. *structured data formats*), aktualnie wykorzystywane do reprezentacji informacji na stronach internetowych WWW:

- (1) mikroformaty;
- (2) Resource Description Framework in Attributes (RDFa);
- (3) mikrodane;
- (4) Java Script Object Notation for Linked Data (JSON-LD).

---

<sup>1</sup> <http://dbpedia.org/>

Wybór formatów do analizy oparto na kryterium zakresu ich rozpowszechnienia w środowisku WWW. Podstawę do identyfikacji danych do analiz stanowiły wyniki badań projektu Web Data Commons<sup>2</sup>, którego celem była ekstrakcja danych ustrukturyzowanych z korpusu stron internetowych oraz analiza zakresu ich implementacji. W artykule przedstawiono również analizę wyników badań tego projektu w kontekście rozpowszechnienia wspomnianych czterech formatów danych strukturalnych.

Przyjęta metodologia badań zakłada interpretację czterech wspomnianych specyfikacji jako formatów danych, służących do tworzenia metadanych na temat treści stron internetowych i zagnieżdżania tych informacji w strukturze dokumentu HTML. Przeprowadzone analizy uwzględniały aspekt semantyczny i syntaktyczny. W pierwszym przypadku oznacza to dociekania związane z zakresem ekspresywności w danym systemie znaczników semantycznych. W drugim ujęciu, każdy z formatów danych został przedstawiony z punktu widzenia formalnych rozwiązań prezentacji metadanych zagnieżdżanych w strukturze dokumentu HTML. Dla każdej z omawianych specyfikacji wskazano zalety i wady związane z jej implementacją.

Do realizacji celu głównego opracowano zestaw szczegółowych pytań badawczych:

- (1) Jaki jest kontekst powstania danego formatu danych strukturalnych?
- (2) Jaki model danych reprezentuje dany FDS?
- (3) Czy dany FDS wprowadza nowe elementy do systemu języka HTML?
- (4) W jaki sposób konstruowane są wykładniki dla elementów metadanych dla każdego z FDS?
- (5) Czy dany FDS pozwala na wykorzystanie istniejących schematów metadanych?
- (6) Jaki jest szacunkowy zakres wykorzystania danego FDS do reprezentacji informacji na stronach internetowych?

### 3. Analiza formatów danych strukturalnych

Z punktu widzenia ewolucji koncepcji formatów danych strukturalnych, do najwcześniejszych prób opracowania specyfikacji umożliwiających zagnieżdżanie metadanych w strukturze dokumentu HTML, należą mikroformaty oraz RDFa. Rozwój tych formatów postępował równolegle do ewolucji języka HTML, który stanowić miał ich naturalne środowisko funkcjonowania, ale był zainicjowany przez mniej (mikroformaty) lub bardziej (RDFa) zorganizowane społeczności internautów. Specyfikacja mikrodanych została opracowana w 2009 r. jako odpowiedź twórców języka HTML na zainteresowanie strukturyzacją treści stron internetowych, zaś JSON-LD jest jedną z ostatnich prób standaryzacji reprezentacji informacji w ramach struktury dokumentów HTML. Przyjęta kolejność analiz FDS odpowiada zatem chronologii, w jakiej formaty te opracowano.

#### 3.1. Mikroformaty

Mikroformaty to jedna z pierwszych prób wprowadzenia danych strukturalnych do systemu języka HTML (Bizer et al., 2012, 3; Ronallo, 2012). Powstały one dzięki społecznej

<sup>2</sup> <http://www.webdatacommons.org/structureddata/>

inicjatywie, której celem było opracowanie szeregu specyfikacji znaczników semantycznych, które w założeniu w prosty sposób mogłyby zostać wykorzystane jako metadane opisujące wybrane fakty publikowane na stronach internetowych.

Celem prac nad mikroformatami było pokonanie ograniczeń języka HTML w zakresie odwzorowania treści stron internetowych, co miało przełożyć się na większą efektywność wyszukiwania tak opisanych zasobów sieciowych. Tłem dla tej inicjatywy były również prace nad realizacją koncepcji Sieci Semantycznej, które skierowane były na popularyzację języka XML w środowisku sieciowym, opublikowanie standardu reprezentacji wiedzy RDF oraz ewolucja samego języka HTML. W 2000 r. opublikowano specyfikację XHTML (Extensible Hypertext Markup Language), która nie była kolejną wersją języka HTML, lecz jego specyfikacją, zgodnie z zaleceniami XML. HTML wywodzi się bezpośrednio z języka Standard Generalized Markup Language (SGML), który reprezentuje inny model niż XML, na którym to oparto XHTML. Prace nad tym standardem prowadzono przez następne kilka lat, lecz XHTML spotkał się z krytyką twórców aplikacji sieciowych i ostatecznie porzucono ten projekt na rzecz piątej wersji języka HTML (HTML5), którą udostępniono w 2014 r. Z punktu widzenia ewolucji technologii sieciowych inicjatywa mikroformatów powstała w momencie, w którym z jednej strony, HTML radykalnie ograniczał możliwości strukturyzacji danych w obrębie strony internetowej, a z drugiej, nowe rozwiązania (XML, RDF) były na tyle skomplikowane i kosztowne, że ich potencjalne wdrożenie na potrzeby projektowania serwisów internetowych nie wchodziło wówczas w grę. Mikroformaty były więc odpowiedzią na realizację wizji Sieci Semantycznej, ale w warstwie powierzchniowej, tzn. na poziomie struktury dokumentów HTML, nie zaś baz danych.

Geneza mikroformatów sięga 2003 r., kiedy to opracowano pierwszą z jego specyfikacji – XHTML Friends Network (XFN). Był to wykaz znaczników, który służył do ekspresji relacji interpersonalnych w odniesieniu do autorów i czytelników intensywnie wówczas rozwijającej się blogosfery. XFN opracowany przez Global Multimedia Protocols Group (2003) pozwalał na specyfikację hiperłączy między blogami lub osobistymi stronami internetowymi poprzez możliwość wskazania rodzaju relacji zachodzących między ich autorami (np. relacje przyjacielskie, rodzinne, zawodowe). Dla tworzenia mikroformatów charakterystyczne było oddolne, społeczne podejście do rozwoju tego rodzaju specyfikacji w przeciwieństwie do instytucjonalnej standaryzacji schematów metadanych. Platformą wymiany informacji i dokumentacji specyfikacji mikroformatów stał się utworzony w 2005 r. serwis internetowy [microformats.org](http://microformats.org). Tak więc cały proces prac nad specyfikacją i jej publikowaniem odbywał się zdalnie. Za motto tej inicjatywy przyjęto stwierdzenie: „zaprojektowane w pierwszej kolejności dla ludzi, w drugiej dla maszyn” (ang. *designed for humans first and machines second*). Konsekwencją takiego podejścia jest pragmatyzm, który jest widoczny zarówno na płaszczyźnie specyfikacji tych formatów danych, jak i w sposobie realizacji prac nad nimi.

### 3.1.1 Poziom semantyczny

Semantyka mikroformatów opiera się na gotowych zestawach znaczników identyfikujących klasy i ich własności, zaprojektowanych i zaakceptowanych przez społeczność zgromadzoną wokół tej inicjatywy. Mamy więc do czynienia ze specyfikacjami metadanych dla różnych kategorii pojęciowych występujących w treści stron internetowych, które są przeznaczone do zagnieżdżenia w strukturze dokumentu HTML. Z punktu widzenia organizacji struktur pojęciowych, dla których formalne wykładniki zapewnia dana specyfikacja mikroformatu,

mamy do czynienia z układem hierarchicznym (Sporny, 2015), którego trzonem jest rodzaj opisywanego elementu (np. osoba, wydarzenie, produkt, recenzja), a znaczniki z nim związane reprezentują jego atrybuty i stanowią jego semantyczne otoczenie. Tabela 1 zawiera przykładowe specyfikacje mikroformatów wraz ze wskazaniem zakresu ich stosowania (zob. też Tab. 6).

Tab. 1. Przykładowe specyfikacje mikroformatów

| Mikroformat            | Zakres stosowania                                 |
|------------------------|---|
| hCard <sup>1</sup>     | metadane dla osób, organizacji                    |
| hCalendar <sup>2</sup> | metadane dla wydarzeń                             |
| hReview <sup>3</sup>   | metadane dla recenzji (produktów i usług)         |
| hAtom <sup>4</sup>     | metadane dla wpisów na blogach internetowych      |
| hProduct <sup>5</sup>  | metadane dla produktu handlowego                  |
| hReview <sup>6</sup>   | metadane dla opinii, recenzji produktu lub usługi |

<sup>1</sup> <http://microformats.org/wiki/hCard>

<sup>2</sup> <http://microformats.org/wiki/hCalendar>

<sup>3</sup> <http://microformats.org/wiki/hReview>

<sup>4</sup> <http://microformats.org/wiki/hAtom>

<sup>5</sup> <http://microformats.org/wiki/hproduct>

<sup>6</sup> <http://microformats.org/wiki/hreview>

Specyfikacja mikroformatów zakłada funkcjonowanie atrybutów prostych i ustrukturyzowanych (Tomberg & Laanpere, 2009). W pierwszym przypadku są to elementy metadanych. W drugim – mają one postać elementów, które reprezentują dany atrybut, ale jego zakres znaczeniowy można dodatkowo uszczegóławiać. Na przykład, mikroformat hCard, przeznaczony do zapisu ustrukturyzowanych informacji na temat osób i instytucji, zakłada funkcjonowanie atrybutów prostych, m.in.:

- fn: nazwa prosta
- email: adres e-mail
- tel: numer telefonu
- url: adres URL strony domowej
- bday: data urodzin

ale również elementy ustrukturyzowane, m.in.:

- adr: adres zamieszkania
  - street-address: ulica
  - locality: miejscowość
  - region: region, stan
  - postal-code: kod pocztowy
  - country-name: państwo
- n: nazwa (element ustrukturyzowany)
  - honorific-prefix: zwroty grzecznościowe (np. Pan, Pani)
  - given-name: imię
  - additional-name: drugie imię
  - family-name: nazwisko
  - honorific-suffix: tytuły (np. Dr, Prof.).

Koncepcja atrybutów ustrukturyzowanych nawiązuje do kwalifikatorów w schemacie metadanych Dublin Core<sup>3</sup>, które były przez krótki czas stosowane przed wprowadzeniem jego rozszerzonej wersji – Terminów Metadanych DCMI<sup>4</sup>. W modelu mikroformatów stosuje się koncepcję obligatoryjności i fakultatywności elementów metadanych. Specyfikacja danego mikroformatu w sposób jawny określa, które z elementów metadanych mają status obowiązkowy.

### 3.1.2 Poziom syntaktyczny

Na poziomie syntaktycznym w mikroformatach została wykorzystana podstawowa składnia HTML. Oznacza to, że nie wprowadzono w nich nowych elementów do systemu znaczników języka HTML. Specyfikacja mikroformatu zakłada funkcjonowanie klas, czyli rodzajów opisywanych obiektów, ich własności oraz zestawu reguł związanych z formalną ekspresją metadanych. Mikroformaty wykorzystują selektory języka HTML, najczęściej znaczniki reprezentujące elementy strukturalne dokumentu (np. *h1* – nagłówek pierwszego stopnia, *p* – paragraf, *a* – hiperłącze lub elementy grupujące – np. *div*), których znaczenie jest definiowane najczęściej za pomocą atrybutu znacznika *class* z wykorzystaniem słownictwa danego mikroformatu. Znacznik *class* w języku HTML jest przewidziany przede wszystkim definiowania elementów w strukturze dokumentu HTML, dla których określono specyficzny sposób ich graficznej prezentacji z wykorzystaniem Kaskadowych Arkuszy Stylów (CSS). W przypadku mikroformatów został on zaadaptowany do deklaracji rekordu danych strukturalnych. Niżej przedstawiono przykładowe formatowanie akapitu tekstu zawierającego dane na temat Jana Kowalskiego oraz jego afiliacji za pomocą mikroformatu hCard.

Wersja wyświetlana użytkownikowi:

Jan Kowalski  
 Uniwersytet Warszawski  
 jan@kowalski.pl  
 Krakowskie Przedmieście 26/28  
 Warszawa , 00-927 Polska

Wersja sformatowana

```
<div id="hcard-Jan-Kowalski" class="vcard">
<span class="fn">Jan Kowalski</span>
<div class="org">Uniwersytet Warszawski</div>
  <a class="email">jan@kowalski.pl</a>
<div class="adr">
  <div class="street-address">Krakowskie Przedmieście 26/28</div>
    <span class="locality">Warszawa</span>
    <span class="postal-code">00-927</span>
    <span class="country-name">Polska</span>
  </div>
```

Rekord metadanych na temat Jana Kowalskiego sformatowano bezpośrednio w treści dokumentu za pomocą znacznika grupującego *div*, gdzie nadano mu identyfikator *id="hcard-Jan-Kowalski"*. W dalszej części rekordu widoczne są wykładniki dla atrybutów prostych (np. *fn*, *org*) oraz ustrukturyzowanych (*adr*).

<sup>3</sup> <http://dublincore.org/documents/2000/07/11/dcmes-qualifiers/>

<sup>4</sup> <http://www.bn.org.pl/download/document/1253606451.pdf>



Zastosowanie mikroformatów uwzględnia również odwzorowanie dodatkowych informacji za pomocą znaczników HTML *rel* oraz *href*, które umożliwiają określenie relacji między elementem treści danego dokumentu a zewnętrznym zasobem sieciowym. Na przykład, dodając zapis w treści dokumentu informujący, że został on opublikowany na licencji „Uznanie autorstwa 2.0 Ogólny (CC BY 2.0)”, można formalnie odwzorować ten fakt za pomocą deklaracji z wykorzystaniem mikroformatu „license”, tworząc dodatkowo hiperłącze do treści licencji:

```
<a href="http://creativecommons.org/licenses/by/2.0/" rel="license">Opublikowano na licencji:  
Uznanie autorstwa 2.0 Ogólny (CC BY 2.0)</a>
```

### 3.1.3 Zalety i ograniczenia

Do zalet mikroformatów zalicza się ich prostotę, zarówno na płaszczyźnie interpretacji specyfikacji, jak również ich wdrażania w serwisach WWW, zgodność z funkcjonującymi standardami sieciowymi (mikroformaty wykorzystują system znaczników języka HTML), a także duże wsparcie społeczności. Krytyka mikroformatów dotyczy przede wszystkim ich niedostatecznej siły semantycznej oraz **braku formalnego modelu** reprezentacji informacji (Tomberg & Laanpere, 2009, 104). W pierwszym przypadku oznacza to małą liczbę specyfikacji i tym samym poważne ograniczenia w szczegółowości odwzorowania treści zasobów WWW. Brak formalnego modelu danych, który leżałby u podstaw mikroformatów, powoduje problemy związane z zastosowaniem automatycznego przetwarzania i analizy spójności danych.

Z punktu widzenia sposobu identyfikacji elementów i atrybutów istotne jest odnotowanie faktu, że żaden z elementów rekordu nie jest identyfikowany za pomocą standardu URI. Oznacza to, że opisywany obiekt, jego kategoryzacja oraz wszystkie elementy metadanych mają postać danych tekstowych. W konsekwencji opisywany obiekt nie ma tożsamości sieciowej (nie posiada identyfikatora sieciowego) i jest rozpoznawalny wyłącznie w kontekście dokumentu HTML. Model mikroformatów ma ograniczone możliwości związane z jawnym określeniem zasad kodowania danych tekstowych (ang. *typed literals*; np. jednostki miar i czasu) i pozwala jedynie na specyfikację języka tekstu elementu metadanych.

## 3.2. RDFa

Specyfikacja Resource Description Framework in Attributes (RDFa) została opracowana w 2004 r., a cztery lata później została opublikowana jako oficjalna rekomendacja konsorcjum W3C (Bizer et al., 2012, 8). W założeniu RDFa miał być formatem wykorzystywanym wyłącznie w ramach języka znacznikowego XHTML, ale wraz z ewolucją HTML format ten został dostosowany do aktualnie funkcjonującej wersji HTML5 (Herman et al., 2015). RDFa opracowano jako serializację<sup>5</sup> standardu reprezentacji wiedzy RDE, przeznaczoną do zagnieżdżania ustrukturyzowanych danych wewnątrz dokumentów HTML (Bizer et al., 2012, 8). RDF jest rekomendowany przez konsorcjum World Wide Web jako standard reprezentacji wiedzy w Sieci Semantycznej oparty na modelu grafowym. Wprowadzono

<sup>5</sup> Pojęcie serializacji oznacza tutaj jawny i udokumentowany formalny sposób ekspresji. Np. format MARC 21 posiada serializację w postaci specyfikacji MARC/XML, ale również formatu wymiennego ISO 2709.



w nim koncepcję tzw. trójki RDF (ang. *RDF Triple*) jako elementarnej jednostki wypowiedzi, w której wyróżnia się *przedmiot* opisu (ang. *subject*), *predykat* (ang. *predicate*; atrybut lub wykładnik relacji semantycznej) oraz *obiekt* (ang. *object*; będący wartością opisywanej cechy lub przedmiotem powiązaniem relacją reprezentowaną przez predykat). RDFa jako serializacja RDF na poziomie konceptualnym realizuje założenia RDF i pozwala na zagnieżdżanie metadanych w postaci trójek RDF wewnątrz systemu znacznikowego języka HTML. RDF jest przeznaczony do formalnej reprezentacji danych na potrzeby baz wiedzy w ramach Sieci Semantycznej, natomiast celem RDFa jest wykorzystanie ekspresywności tego modelu na potrzeby opisu zawartości stron WWW.

### 3.2.1 Poziom semantyczny

Podstawową cechą RDFa w kontekście semantycznej warstwy reprezentacji informacji jest brak z góry narzuconego zestawu elementów metadanych. Oznacza to możliwość wykorzystania dowolnego schematu metadanych lub ontologii sieciowej, które posiadają swoją formalną specyfikację. Warunek formalnej reprezentacji schematu metadanych, który ma być wykorzystany w opisie za pomocą RDFa oznacza dostęp do niego za pośrednictwem jego unikalnej sieciowej przestrzeni nazw (ang. *namespace*) i tym samym stosowanie standardu URI do identyfikacji jego elementów strukturalnych<sup>6</sup>. Jest to niewątpliwa zaleta RDFa, ponieważ pozwala na ponowne użycie (ang. *re-use*) istniejących specyfikacji metadanych i brak konieczności opracowania nowych, wyłącznie na potrzeby tego formatu danych strukturalnych. Dodatkowo w ramach jednego rekordu metadanych (zestawu trójek RDF) istnieje możliwość jednoczesnego stosowania wielu takich specyfikacji.

### 3.2.2 Poziom syntaktyczny

Na poziomie syntaktycznym w RDFa wprowadzono szereg nowych atrybutów do składni HTML, których celem jest dostosowanie modelu RDF do koncepcji zagnieżdżania metadanych w strukturze HTML. RDFa funkcjonuje w dwóch wersjach – wersji podstawowej RDFa Core 1.1<sup>7</sup> oraz wersji uproszczonej – RDFa Lite<sup>8</sup>. Różnica polega na liczbie elementów syntaktycznych, a więc tym samym na poziomie szczegółowości opisów. RDFa Lite wprowadzono w celu ułatwienia twórcom serwisów internetowych implementacji tego FDS, a kryteria doboru elementów miały charakter pragmatyczny. Dokumentacja RDFa Lite zawiera stwierdzenie, że ta specyfikacja będzie adekwatna w 80% przypadków związanych z opracowaniem semantycznych znaczników dla zawartości stron WWW. W celu ilustracji rozwiązań syntaktycznych RDFa w dalszej części tej sekcji zostaną one przedstawione na podstawie jego uproszczonej wersji. W RDFa Lite wprowadzono pięć nowych atrybutów o charakterze syntaktycznym<sup>9</sup>:

- (1) *vocab* – element systemu znacznikowego służący do identyfikacji bazowej przestrzeni nazw (URI) schematu metadanych, z którego elementy będą wykorzystywane w opisie;

<sup>6</sup> Np. schemat metadanych Dublin Core Metadata Element Set jest zarejestrowany w przestrzeni nazw – <http://purl.org/dc/elements/1.1/>, w której każdy ze zidentyfikowanych w nim atrybutów posiada swój unikalny identyfikator URI (np. tytuł – <http://purl.org/dc/elements/1.1/title>).

<sup>7</sup> <https://www.w3.org/TR/rdfa-core/>

<sup>8</sup> <https://www.w3.org/TR/rdfa-lite/>

<sup>9</sup> RDFa w wersji podstawowej wprowadza osiem nowych elementów syntaktycznych do składni HTML.

- (2) *prefix* – element systemu znacznikowego służący do wprowadzenia akronimu dla dodatkowych przestrzeni nazw stosowanych w opisie w celu skrócenia i uproszczenia sposobu odwoływania się do elementów metadanych;
- (3) *typeof* – element systemu znacznikowego służący do kategoryzacji opisywanego obiektu (np. osoba, wydarzenie). Wartość tego elementu pochodzi z wcześniej zadeklarowanej (element *vocab*) przestrzeni nazw dla schematu metadanych;
- (4) *property* – element systemu znacznikowego służący do identyfikacji odwzorowywanej własności obiektu. Nazwa tego elementu pochodzi z wcześniej zadeklarowanej (element *vocab*) przestrzeni nazw dla schematu metadanych, a jego wartość w zależności od typu atrybutu przyjmuje postać ciągu znaków lub URI;
- (5) *resource* – element systemu znacznikowego służący do identyfikacji opisywanego obiektu.

Poniżej przedstawiono formatowanie akapitu tekstu zawierającego wcześniej przywoływane dane teleadresowe na temat Jana Kowalskiego za pomocą formatu danych RDFa:

```
<div vocab="http://schema.org/" resource="Jan-Kowalski" typeof="Person">
<span property="name">Jan Kowalski</span>
<span property="email">jan@kowalski.pl</span>
<div property="affiliation" resource="UW" typeof="Organization">
  <span property="name">Uniwersytet Warszawski</span>
<div property="address" resource="UW-adres" typeof="PostalAddress">
<span property="streetAddress">Krakowskie Przedmieście 26/28</span>
  <span property="addressLocality">Warszawa</span>
  <span property="postalCode">00-927</span>
  <span property="addressCountry">Polska</span></div>
</div>
```

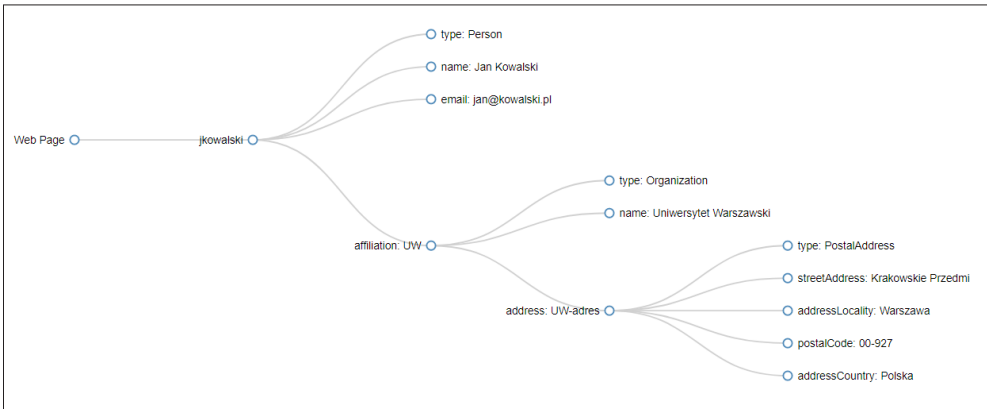
Przedstawiony powyżej zestaw metadanych w formacie RDFa rozpoczyna się od deklaracji wykorzystanego w dalszej części zewnętrznego schematu metadanych (*vocab*), którym w tym przykładzie jest specyfikacja [schema.org](http://schema.org/)<sup>10</sup>, określenia przedmiotu opisu (*resource*) oraz wskazania jego wewnętrznego identyfikatora dla osoby Jana Kowalskiego (*resource* = "Jan-Kowalski"), a także formalnej kategoryzacji opisywanego obiektu, tzn. stwierdzenia, że jest on instancją, tj. wystąpieniem klasy człowiek (*typeof* = "Person"). W dalszej części każda wartość atrybutu jest poprzedzona identyfikatorem *property* oraz wykładnikiem atrybutu ze schematu [schema.org](http://schema.org/). Przedmiotem opisu jest zadeklarowany w pierwszej linii obiekt *Jan-Kowalski*, ale zestaw metadanych jest dodatkowo ustrukturyzowany i zawiera również informacje o obiekcie *Uniwersytet Warszawski*, który reprezentuje afiliację *Jana Kowalskiego*. Obiekt ten zidentyfikowano (*resource*) i skategoryzowano jako instancję klasy organizacja (*typeof* = "Organization"). Na jego opis składają się elementy proste (nazwa – *name*) oraz ustrukturyzowane (*PostalAddress*) – szczegółową specyfikację danych adresowych przedstawia rysunek 1.

RDFa, tak jak i RDF, wyznacza pewien formalny schemat reprezentacji wiedzy. Jego słownictwo, oprócz podstawowych deklaracji ontologicznych, jest wykorzystywane głównie do organizacji składni. Do niewątpliwych zalet RDFa należy możliwość zastosowania zewnętrznych schematów metadanych i ontologii sieciowych poprzez odwoływanie się do ich zawartości za pośrednictwem URI dla danej przestrzeni nazw. Istnieje więc możliwość

<sup>10</sup> <http://schema.org/>

jednoczesnego korzystania z wielu takich specyfikacji. Pozwala to na wysoki stopień szczegółowości reprezentacji informacji, ale wiąże się z koniecznością każdorazowej weryfikacji statusu ontologicznego wykorzystanych atrybutów i klas. Poniżej przedstawiono formatowanie tego samego tekstu z zastosowaniem RDFa, gdzie użyto elementy schematu metadanych schema.org oraz ontologii Friend of a Friend (FOAF)<sup>11</sup>:

```
<div prefix="schema: http://schema.org/ foaf: http://xmlns.com/foaf/0.1/" resource="Jan-Kowalski" typeof="foaf:Person">
<span property="foaf:name">Jan Kowalski</span>
<span property="foaf:mbox">jan@kowalski.pl</span>
<div property="schema:affiliation" resource="UW" typeof="schema:Organization">
  <span property="foaf:name">Uniwersytet Warszawski</span>
<div property="schema:address" typeof="schema:PostalAddress" resource="UW-adres">
<span property="schema:streetAddress">Krakowskie Przedmieście 26/28</span>
  <span property="schema:addressLocality">Warszawa</span>
  <span property="schema:postalCode">00-927</span>
  <span property="schema:addressCountry">Polska</span></div>
</div>
```



Rys. 1. Graficzna reprezentacja deklaracji w RDFa

Ontologia FOAF jest jedną z bardziej rozpowszechnionych formalnych struktur informacji osobowej. Jest wykorzystywana do tworzenia tzw. semantycznych wizytówek sieciowych, w których treści można zawrzeć zarówno podstawowe dane osobowe, jak również specyfikować relacje z innymi ludźmi (np. znajomości). W powyższym przykładzie nie zadeklarowano bazowej przestrzeni nazw (*vocab*), lecz wykorzystano wykładnik *prefix* do wskazania na specyfikacje metadanych, które będą użyte do reprezentacji informacji. Imię i nazwisko (*foaf:name*) oraz adres e-mail (*foaf:mbox*) odwzorowano za pomocą wykładników klas (*typeof="foaf:Person"*) i atrybutów z ontologii FOAF. Ontologia ta nie pozwala jednak na dokładne odwzorowanie danych teleadresowych, stąd informacje na temat Uniwersytetu Warszawskiego zapisano za pomocą wykładników atrybutów i klas ze schematu schema.org.

<sup>11</sup> <http://xmlns.com/foaf/spec/>

### 3.2.3 Zalety i ograniczenia

Oprócz możliwości łączenia wielu formalnych specyfikacji metadanych i ontologii w ramach zestawu deklaracji, wartą podkreślenia zaletą RDFa jest operowanie identyfikatorami sieciowymi URI ze wskazanymi w opisie przestrzeni nazw. A zatem, zastosowany w przykładzie skrótowy zapis *foaf:Person* dla kategoryzacji Jana Kowalskiego, formalnie oznacza, że obiekt Jan-Kowalski opisany na danej stronie internetowej jest wystąpieniem klasy osoba w ontologii FOAF, która posiada swoje URI – [http://xmlns.com/foaf/spec/#term\\_Person](http://xmlns.com/foaf/spec/#term_Person), a obiekt UW opisany na tej samej stronie WWW jest wystąpieniem klasy organizacja w myśl schematu [schema.org](http://schema.org), która również posiada swoje unikalne URI – <http://schema.org/Organization>. Taka formalna kategoryzacja treści zasobów WWW oraz reprezentacja z wykorzystaniem schematów metadanych i ontologii z jednej strony pozwala na precyzyjne indeksowanie zasobów WWW przez wyszukiwarki internetowe, a z drugiej na agregację danych i ich konwersję do postaci baz wiedzy, np. do formatu RDF.

Zastosowanie formatu RDFa może jednak napotkać problemy po stronie pragmatyki. Format ten daje duże możliwości na płaszczyźnie semantycznej, co w warstwie syntaktycznej przejawia się dużą elastycznością. Z drugiej strony, Bethany Wetherill (2014) twierdzi, że wyniki badań wskazują na duży odsetek błędów popełnianych przez twórców serwisów WWW w praktycznym operowaniu wieloma narzędziami reprezentacji informacji oraz odwoływaniu się do nich w deklaracjach RDFa.

## 3.3. Mikrodane

Mikrodane to format danych strukturalnych, który powstał w 2009 r. w ramach prac nad piątą wersją języka HTML oraz w wyniku krytyki środowiska, dotyczącej włączenia RDFa do specyfikacji HTML (Ronallo, 2012). W założeniu mikrodane miały być alternatywnym sposobem zagnieżdżania semantycznych znaczników w strukturze dokumentów HTML (Bizer et al., 2012, 11), który da więcej możliwości niż mikroformaty, ale będzie prostszy niż RDFa. Format ten, w przeciwieństwie do dwóch omówionych wcześniej, jest elementem języka HTML. Nie stanowi on nadbudowy dla systemu znaczników HTML, tak jak to ma miejsce w przypadku mikroformatów i RDFa, lecz posiada status modułu, który rozszerza jego możliwości zgodnie z założeniami specyfikacji HTML5 o reprezentację danych strukturalnych wewnątrz struktury stron WWW (Sikos, 2015, 35). Format mikrodane powstał w momencie, w którym środowisko twórców serwisów internetowych zainteresowane wprowadzaniem danych strukturalnych miało już kilkuletnie doświadczenie zarówno w adaptacji mikroformatów jak i RDFa. Jest to więc wynik z jednej strony, świadomej ewolucji języka HTML, a z drugiej, refleksji o charakterze pragmatycznym, w której pod uwagę wzięto potencjalne i realne trudności związane z wdrażaniem szczególnie RDFa. I chociaż dla technologii semantycznych format RDF ma kluczowe znaczenie, to jego serializacja w postaci RDFa nie spotkała się z entuzjazmem projektantów serwisów WWW, chociaż z perspektywy kilku lat funkcjonowania mikrodanych, to RDFa uznaje się za standard dojrzalszy (Wetherill, 2014).

### 3.3.1 Poziom semantyczny

W modelu konceptualnym, leżącym u podstaw mikrodanych, zakłada się funkcjonowanie zbioru par atrybut-wartość nazywanych tutaj obiektami (ang. *item*). Dany obiekt posiada

dodatkowo deklarację ontologiczną (*item type*) wskazującą na kategorię pojęciową, którą reprezentuje opisywany zasób informacyjny (np. człowiek, organizacja, książka) oraz może zawierać tzw. globalny identyfikator (*item id*), który powinien być wyrażony za pomocą standardu URI. Pierwszy element z pary atrybut-wartość określa własność (ang. *property*) opisywanego obiektu. Może ona przyjąć jedną lub wiele wartości. Wartości są odwzorowywane za pomocą ciągu znaków (ang. *string*) lub mają postać złożoną i tworzą je zestawy par atrybut-wartość. Istnieje również możliwość wyrażania wartości danego atrybutu poprzez odwołanie do innego obiektu (McCathie Nevile & Brickley, 2017). Przykładem własności może być oznaczenie odpowiedzialności dla dokumentu, które może mieć tyle wartości, ile było podmiotów zaangażowanych w jego utworzenie. Atrybut ten może przyjąć wartość prostą w postaci ciągu znaków (np. „Jan Kowalski”) lub złożoną w postaci odwołania do obiektu o unikalnym identyfikatorze, w którego opisie będzie kolejny zestaw par atrybut-wartość (np. imię: Jan; nazwisko: Kowalski).

Format mikrodane, tak jak RDFa, nie narzuca określonego zestawu metadanych. Podobnie jak w przypadku RDFa, i tutaj można wykorzystać istniejące specyfikacje metadanych i ontologii sieciowych, które posiadają zarejestrowane przestrzenie nazw. Duży udział w powstaniu mikrodanych miała inicjatywa przedstawicieli największych dostawców wyszukiwarek internetowych – Google, Bing, Yahoo i Yandex, w ramach prac nad wspólną specyfikacją metadanych przeznaczoną dla formatów danych strukturalnych. Efektem tej współpracy jest projekt schema.org, którego celem było opracowanie prostej specyfikacji metadanych o szerokim zakresie stosowania do opisu najczęściej występujących typów obiektów w treści stron internetowych (np. osoby, miejsca, wydarzenia, produkty) z wykorzystaniem formatów danych strukturalnych. Schema.org opublikowano w 2011 r. i obecnie liczy on 638 klas i 965 atrybutów i relacji (Guha et al., 2015). Specyfikacja ta została opracowana dla twórców serwisów internetowych jako stosunkowo proste narzędzie do semantycznej strukturyzacji treści stron internetowych w celu optymalizacji ich indeksowania przez wyszukiwarki internetowe i tym samym w celu zwiększenia trafności wyszukiwania informacji. Tym samym, schema.org stał się niejako pierwszym wyborem dla formatu mikrodane na płaszczyźnie semantycznej.

### 3.3.2 Poziom syntaktyczny

Mikrodane jako immanentny element strukturalny języka znacznikowego HTML5 wprowadza do jego systemu pięć nowych elementów (McCathie Nevile & Brickley, 2017):

- (1) *itemscope* – element systemu znacznikowego służący do identyfikacji nowego obiektu, którego opis będzie wyrażony za pomocą par atrybut-wartość;
- (2) *itemtype* – element systemu znacznikowego służący do kategoryzacji opisywanego obiektu (np. osoba, wydarzenie);
- (3) *itemid* – element systemu znacznikowego służący do identyfikacji opisywanego obiektu z wykorzystaniem identyfikatora URI;
- (4) *itemprop* – element systemu znacznikowego służący do identyfikacji odwzorowywanej własności obiektu;
- (5) *itemref* – element systemu znacznikowego służący do grupowania par atrybut-wartość w celu organizacji deklaracji wewnątrz opisu.

Najważniejsze elementy mikrodanych przeznaczone do odwzorowania informacji to *itemscope*, *itemtype* oraz *itemprop*. Poniżej przedstawiono formatowanie akapitu tekstu

zawierającego wcześniej przywoływane dane teled adresowe na temat Jana Kowalskiego za pomocą mikrodanych z wykorzystaniem elementów metadanych ze schematu `schema.org`:

```
<div itemscope itemtype="http://schema.org/Person"
  itemid="http://przyklad.pl/JKowalski" >
  <span itemprop="name">Jan Kowalski</span>
  <span itemprop="email">jan@kowalski.pl</span>
  <div itemprop="affiliation" itemscope itemtype="http://schema.org/Organization" itemid="http://
  przyklad.pl/UW" >
    <span itemprop="name">Uniwersytet Warszawski</span>
    <div itemprop="address" itemscope itemtype="http://schema.org/PostalAddress">
      <span itemprop="streetAddress">Krakowskie Przedmieście 26/28</span>
      <span itemprop="addressLocality">Warszawa</span>
      <span itemprop="postalCode">00-927</span>
      <span itemprop="addressCountry">Polska</span></div>
    </div>
```

Przedmiotem opisu w przykładzie jest obiekt (*itemscope*) o identyfikatorze (*itemid*) – `http://przyklad.pl/JKowalski`. Deklaracja rozpoczynająca opis zawiera również informacje o statusie ontologicznym obiektu poprzez wskazanie, że jest on instancją klasy człowiek (*itemtype* = "`http://schema.org/Person`") zidentyfikowanej w schemacie `schema.org`. W dalszej części występują pary atrybut-wartość odwzorowujące imię i nazwisko oraz adres e-mail. Afiliacja Jana Kowalskiego została odwzorowana poprzez ustalenie relacji między obiektem Jan Kowalski (*itemid* = "`http://przyklad.pl/JKowalski`") a obiektem Uniwersytet Warszawski (*itemid* = "`http://przyklad.pl/UW`"), której wykładnikiem jest atrybut *affiliation*. Uniwersytet Warszawski zyskał w tym opisie własną tożsamość poprzez nadanie mu (lub odwołanie się do jego) identyfikatora oraz został skategoryzowany jako wystąpienie klasy organizacja (*itemtype* = "`http://schema.org/Organization`"). Informacje na temat instytucji afiliującej zostały odwzorowane w sposób prosty (nazwa) oraz poprzez wyodrębnienie obiektu złożonego adresu, który w ontologii `schema.org` reprezentuje klasa `http://schema.org/PostalAddress`. Taki zabieg zastosowano zarówno w przypadku użycia `schema.org` w RDFa jak i w mikrodanych. Z formalnego punktu widzenia `schema.org` nie pozwala na użycie atrybutów teled adresowych – nazwa ulicy (*streetAddress*), lokalizacja (*addressLocality*), kod pocztowy (*postalCode*) oraz państwo (*addressCountry*) bezpośrednio do opisu organizacji, którą w tym przypadku reprezentuje Uniwersytet Warszawski. Są to własności zdefiniowane dla klasy adres pocztowy (*PostalAddress*), która służy do reprezentacji wartości atrybutu adres (*address*). Specyfikacja `schema.org` pozwala bowiem na dwa sposoby odwzorowania atrybutu adres. Pierwszym z nich jest nieustrukturyzowany ciąg znaków, który zawiera dane teled adresowe, drugim – wersja ustrukturyzowana z wykorzystaniem typu klasy adres pocztowy (*PostalAddress*). Stosując pierwsze rozwiązanie, dane te mogą przyjąć następujące formatowanie:

```
<div itemscope itemtype="http://schema.org/Person"
  itemid="http://przyklad.pl/JKowalski" >
  <span itemprop="name">Jan Kowalski</span>
  <span itemprop="email">jan@kowalski.pl</span>
  <div itemprop="affiliation" itemscope itemtype="http://schema.org/Organization" itemid="http://
  przyklad.pl/UW" >
    <span itemprop="name">Uniwersytet Warszawski</span>
    <span itemprop="address"> Krakowskie Przedmieście 26/28, 00-927 Warszawa, Polska</span>
  </div>
```



### 3.3.3 Zalety i ograniczenia

Mikrodane jako format danych strukturalnych postrzegane są jako wynik konsensusu między dążeniem do pogłębionej strukturyzacji treści stron internetowych a prostotą systemu znaczników semantycznych. Tomberg i Laanpere (2009) traktują mikrodane jako hybrydę, która powstała w wyniku połączenia rozwiązań zaczerpniętych z mikroformatów, RDFa i HTML5. Nie ulega jednak wątpliwości, że wsparcie mikrodanych przez Google, Bing, Yahoo i Yandex poprzez rekomendację tego FDS dla implementacji schema.org pozwala sądzić, że będzie on coraz szerzej stosowany do strukturyzacji treści stron internetowych. Do zalet mikrodanych należy możliwość stosowania wielu schematów metadanych i ontologii oraz wykorzystanie identyfikatorów sieciowych (URI) dla opisywanych obiektów. Model mikrodanych nawiązuje do modelu RDF, który leży u podstaw RDFa, co traktuje się jako zaletę, z którą wiąże się możliwość stosunkowo łatwego generowania trójek RDF z tak ustrukturyzowanej zawartości stron WWW na potrzeby tworzenia lub wzbogacania semantycznych baz danych (Sporny, 2015; Tomberg & Laanpere, 2009; Wetherill, 2014). Prostota, czyli to, co jedni (np. Guha et al., 2015) postrzegają jako zaletę mikrodanych, przez innych (np. Wetherill, 2014) jest traktowana również jako ograniczenie. Większość krytyki mikrodanych jest skoncentrowana na fakcie, że pragmatyczne pobudki i mocne powiązanie ze schema.org powodują, iż format ten nie nadaje się do opisów złożonych o dużym stopniu analityczności, np. publikacji naukowych czy treści z sektora e-government.

## 3.4. JSON-LD

Format danych strukturalnych JSON-LD (Java Script Object Notation for Linked Data) powstał w odpowiedzi na zainteresowanie środowiska twórców usług sieciowych (aplikacje sieciowe, serwisy internetowe) implementacją metodyki publikowania danych według założeń Linked Data. Punktem wyjścia było poszukiwanie adekwatnego rozwiązania w postaci rozszerzenia możliwości ekspresji formatu Java Script Object Notation (JSON). Jest on szeroko stosowanym formatem danych, który zapewnia efektywną reprezentację danych przesyłanych z serwera obsługującego dany serwis WWW do przeglądarki internetowej, wyświetlającej jego zawartość i zapewniającej interakcję z użytkownikiem. Do zalet formatu JSON należy prostota w tworzeniu reprezentacji danych oraz ich interpretacji, zarówno przez ludzi, jak i aplikacje je przetwarzające. Obsługa danych w formacie JSON nie wymaga instalacji dodatkowych rozszerzeń w przypadku wykorzystania tego formatu na potrzeby serwisów internetowych. Dane przesyłane do przeglądarki w postaci kodu zagnieżdżonego w strukturze dokumentu HTML są przetwarzane przez skrypty w języku Java Script, które realizują określone funkcjonalności.

JSON pozwala na definicję typów danych (np. dane liczbowe, ciąg znaków) oraz oferuje własną notację. W założeniu ekspresja danych przyjmuje formę zbioru par atrybut-wartość, która jest zamknięta w nawiasy klamrowe, pełniące również funkcję znaków delimitacyjnych dla obiektów o złożonej strukturze. Przywoływany w tekście przykład danych teleadresowych dla Jana Kowalskiego w uproszczonej postaci formatu JSON przedstawiono poniżej:



```
{
  „imię-nazwisko”: „Jan Kowalski”,
  „e-mail”: „jan@kowalski.pl”,
  „afiliacja”: {
    „nazwa”: „Uniwersytet Warszawski”,
    „ulica”: „Krakowskie Przedmieście 26/28”,
    „miasto”: „Warszawa”,
    „państwo”: „Polska”,
    „kod-pocztowy”: „00-927”}
}
```

Z punktu widzenia tematu tego artykułu istotny jest kontekst semantyczny odwzorowywany w formatach danych strukturalnych. Chodzi o możliwość formalnej specyfikacji (za pośrednictwem URI przestrzeni nazw) schematów metadanych czy ontologii, wykorzystywanych do ekspresji danych i stosowania semantycznych znaczników. JSON w podstawowej formie nie daje takich możliwości, jak np. RDFa (wykładniki *prefix* oraz *vocab*) czy mikrodane (wykładnik *itemtype*). Problem ten został rozwiązany w postaci rozszerzenia JSON w formie specyfikacji JSON-LD. W przedstawionym przykładzie występują wykładniki elementów metadanych, które wyrażono za pomocą języka naturalnego (np. e-mail, ulica, miasto), co jest zrozumiałe dla człowieka, jednak z punktu widzenia przetwarzania, a szczególnie wymiany danych, powoduje problemy z ich formalną interpretacją.

Prace nad formatem danych JSON-LD doprowadziły do opublikowania w 2012 r. pierwszej wersji roboczej jego specyfikacji. Dwa lata później konsorcjum WWW przyjęło i opublikowało oficjalną rekomendację JSON-LD 1.0 o nazwie *A JSON-based Serialization for Linked Data*.

### 3.4.1 Poziom semantyczny

Na poziomie koncepcyjnym JSON-LD realizuje model formatu danych JSON, który ma postać grafu skierowanego, którego węzły tworzą elementy danych (np. ciąg znaków, liczba, URI), a krawędzie są wykładnikami atrybutów i relacji (Sporny et al. 2014). Format JSON-LD pełni funkcję serializacji modelu RDF, tak jak to ma miejsce w odniesieniu do RDFa. JSON-LD, tak jak RDFa i mikrodane, pozwala na stosowanie formalnych specyfikacji metadanych i ontologii w celu reprezentacji informacji. JSON-LD pozwala również na jednoczesne stosowanie wielu schematów metadanych w ramach reprezentacji tego samego zasobu informacyjnego. Możliwość odwoływania się do zewnętrznych specyfikacji metadanych została zapewniona poprzez wprowadzenie dodatkowych elementów do składni JSON-LD, co zapewnia zgodność z wytycznymi publikowania danych w modelu Linked Data.

### 3.4.2 Poziom syntaktyczny

Na poziomie syntaktycznym JSON-LD nie wprowadza żadnych elementów do systemu języka HTML. Zagnieżdżanie danych w formacie JSON czy JSON-LD w dokumencie HTML polega na zastosowaniu skryptu w języku Java Script, który albo pobiera dane z serwera, albo odwołuje się do danych przesłanych wraz ze stroną HTML, które zapisano w sekcji odpowiedzialnej za skrypty (<script>). Jest to więc zgoła inne podejście do zastosowania FDS niż w przypadku mikroformatów, RDFa i mikrodanych, gdzie formaty te są wykorzystywane do zagnieżdżania danych bezpośrednio w treści dokumentu.

JSON-LD jako rozszerzenie JSON wprowadza szereg nowych elementów syntaktycznych, które pozwalają na publikowanie danych w modelu Linked Data wewnątrz struktury dokumentu HTML. Cztery najważniejsze z nich to:

- (1) *@context* – element systemu znacznikowego służący do identyfikacji bazowej przestrzeni nazw (URI) schematu metadanych, z którego elementy będą wykorzystywane w opisie;
- (2) *@id* – element systemu znacznikowego służący do identyfikacji opisywanego obiektu z wykorzystaniem identyfikatora URI;
- (3) *@type* – element systemu znacznikowego służący do kategoryzacji opisywanego obiektu (np. osoba, wydarzenie);
- (4) *@vocab* – element systemu znacznikowego służący do deklaracji dodatkowych przestrzeni nazw, które rozszerzają możliwości formalnej ekspresji metadanych.

Poniżej przedstawiono uproszczoną wersję formatu JSON-LD dla przywoływanego wcześniej przykładu danych teleadresowych dla Jana Kowalskiego:

```
{
  «@context»: «http://schema.org/»,
  «@type»: «Person»,
  "name": "Jan Kowalski",
  "affiliation": {
    "@context": "http://schema.org/",
    „@type”: „Organization”,
    „name”: „Uniwersytet Warszawski”,
    „address”: {
      „@type”: „PostalAddress”,
      „streetAddress”: „Krakowskie Przedmieście 26/28”,
      „addressLocality”: „Warszawa”,
      „postalCode”: „00-927”,
      „addressCountry”: „Polska”}
    }
  }
}
```

Mamy tutaj do czynienia z reprezentacją informacji, której przedmiotem jest Jan Kowalski formalnie skategoryzowany jako instancja klasy człowiek w ontologii schema.org. Tak jak w poprzednich przykładach afiliacja Jana Kowalskiego została odwzorowana za pomocą prostych elementów metadanych (nazwa) oraz obiektu złożonego odwzorowującego dane teleadresowe Uniwersytetu Warszawskiego. Zdefiniowanie kontekstu poprzez użycie deklaracji „@context”: „http://schema.org/” powoduje, że wszystkie zastosowane wykładniki metadanych, chociaż zapisywane w skróconej postaci (np. „name”), są formalnie interpretowane za pośrednictwem URI w zadeklarowanej przestrzeni nazw (np. http://schema.org/name).

Nadanie tożsamości sieciowej osobie Jana Kowalskiego w przytoczonym przykładzie polega na zastosowaniu elementu *@id* i przeformułowaniu deklaracji zawierającej kontekst:

```
{
  „@context”:
  {
    „@vocab”: „http://schema.org/”
  },
  „@id”: „ http://przyklad.pl/Kowalski”,
  „@type”: „Person”,
  „name”: „Jan Kowalski”,
  ...
}
```

W tym przypadku formalny kontekst semantyczny deklaracji, wyrażonej za pomocą formatu JSON-LD, został wyrażony poprzez zdefiniowanie przedmiotu odniesienia (elementów schematu schema.org) za pośrednictwem wykładnika „@vocab”.

Specyfikacja JSON-LD zawiera szczegółową charakterystykę elementów tego formatu danych, które pozwalają na bardzo szczegółową reprezentację danych oraz konstrukcję elementów złożonych. Zasadniczą cechą JSON-LD jest zbieżność koncepcyjna z modelem RDF oraz operowanie URI zarówno dla elementów zastosowanego schematu metadanych, jak i dla obiektów zdefiniowanych na danej stronie internetowej.

### 3.4.3 Zalety i ograniczenia

Do podstawowych zalet tego formatu danych należy z pewnością zaliczyć jego rozpowszechnienie wśród twórców aplikacji i usług sieciowych oraz prostotę w użyciu. Dane w formacie JSON-LD nie są zapisywane bezpośrednio w treści strony WWW, lecz w osobnej sekcji przeznaczony dla zagnieżdżania skryptów. W przypadku interpretacji JSON-LD nie jako formatu wymiany danych, lecz jako dodatkowej warstwy semantycznej odwzorowywanej w ramach struktury dokumentu HTML, można to traktować zarówno jako zaletę jak i mankament. W pierwszym ujęciu mamy ewidentne rozdzielenie treści strony WWW od metadanych, w drugim – mamy do czynienia z powielaniem pewnych elementów treści.

## 3.5. Podsumowanie

Przedstawiona charakterystyka czterech formatów danych strukturalnych daje obraz przede wszystkim sposobu konstrukcji i ekspresji metadanych w ramach systemu języka znacznikowego HTML, kładąc nacisk na rozwiązania syntaktyczne oraz otwartość na implementację istniejących schematów metadanych i ontologii sieciowych, stosowanych w kolekcjach cyfrowych. Mikroformaty były pierwszą próbą wzbogacenia struktury stron WWW o znaczniki semantyczne pozwalające na głębszą strukturyzację zawartości. RDFa jest wynikiem próby przeniesienia koncepcji reprezentacji wiedzy standardu RDF stosowanego w bazach wiedzy na poziom stron WWW, zaś mikrodane wydają się być kompromisem między prostotą mikroformatów, a szczegółową strukturyzacją charakterystyczną dla RDFa. JSON-LD to z kolei „ukłon” konsorcjum WWW w stronę twórców aplikacji sieciowych, dla których format JSON jest naturalnym środowiskiem danych. Tabela 2 zawiera porównanie czterech opisanych formatów danych strukturalnych z uwzględnieniem wybranych cech. Manu Sporny (2015) opracował szczegółowe porównanie mikroformatów, RDFa oraz mikrodanych, zawierające ich pogłębione analizy.

Z punktu widzenia ingerencji w system języka znacznikowego HTML najwięcej nowych elementów syntaktycznych wprowadza RDFa, co niewątpliwie ma wpływ na siłę ekspresji tego FDS. Zarówno mikro dane, RDFa jak i JSON-LD są otwarte na implementację zewnętrznych schematów metadanych i ontologii sieciowych w kontekście doboru narzędzi reprezentacji informacji adekwatnego dla opisywanej kolekcji. W przypadku mikro danych wiąże się to jednak w praktyce ze stosowaniem specyfikacji schema.org, może nawet nie tyle ze względu na siłę semantyczną tego schematu metadanych, co jego rekomendację przez dostawców wyszukiwarek internetowych i tym samym – dążenie do „lepszej widoczności” tak zaindeksowanych zasobów sieciowych w rezultatach wyszukiwania. Mikroformaty są ściśle powiązane z konkretnymi specyfikacjami metadanych. Zarówno mikro dane, jak i RDFa, pozwalają na odwoływanie się do opisywanych obiektów oraz ich atrybutów za pośrednictwem ich unikalnych URI. Daje to większą możliwość standaryzacji i tym samym gotowość na agregację danych z wielu źródeł.

Tab. 2. Porównanie formatów danych strukturalnych;  
źródła: Sporny, 2015 oraz badania własne

| Cecha   | Mikroformaty | Mikro dane       | RDFa   | JSON-LD     |
|---|--------------|------------------|--------|-------------|
| Liczba nowych elementów syntaktycznych HTML   | 0            | 5                | 8      | Nie dotyczy |
| Powiązanie z konkretnym schematem metadanych  | Tak          | Nie <sup>1</sup> | Nie    | Nie         |
| Identyfikacja opisywanego obiektu za pomocą URI   | Nie          | Tak              | Tak    | Tak         |
| Formalna kategoryzacja opisywanego zasobu informacyjnego z wykorzystaniem URI                     | Nie          | Tak              | Tak    | Tak         |
| Możliwość jednoczesnego stosowania wielu schematów metadanych w ramach jednego zestawu deklaracji | Nie          | Tak              | Tak    | Tak         |
| Trudność w implementacji  | Niska        | Średnia          | Wysoka | Niska       |

<sup>1</sup> Formalnie mikro dane nie są związane z konkretnym schematem metadanych, ale w praktyce ich stosowanie wiąże się z użyciem schema.org

Trudność w implementacji odnosi się nie tyle do aspektu technicznego, co do pragmatyki, czyli zastosowania w procesie adnotowania treści stron internetowych przez ich twórców. W takim ujęciu RDFa jest wskazywane jako stosunkowo trudne we wdrożeniu, przy średniej ocenie mikro danych i prostocie mikroformatów oraz JSON-LD.

Przedstawione FDS można również analizować z punktu widzenia ich genezy. Mikroformaty to bezsprzecznie inicjatywa twórców serwisów internetowych oraz rozwijającej się ówczesnie blogosfery. RDFa to pomysł na uproszczenie tworzenia baz wiedzy z wykorzystaniem RDF na rzecz strukturyzacji bezpośrednio treści stron internetowych i próba implementacji technologii semantycznych do projektowania serwisów WWW. JSON-LD to z kolei odpowiedź konsorcjum WWW na potrzeby twórców aplikacji sieciowych. W końcu mikro dane to głos twórców języka HTML w dyskusji nad adekwatnym formatem danych strukturalnych.

#### 4. Zakres wykorzystania formatów danych strukturalnych w World Wide Web

Z dotychczas przedstawionych rozważań wyłania się obraz czterech formatów danych przeznaczonych do formalnej strukturyzacji treści stron WWW, z których mikro dane, RDFa i JSON-LD operują zbliżonymi środkami ekspresji. Obecność wielu standardów jest weryfikowana w praktyce przez ich realne wykorzystanie przez społeczność, do której są skierowane. Na wybór danego rozwiązania oprócz jego efektywności ma wpływ wiele czynników, do których na pewno można zaliczyć poziom ich trudności oraz stosunek poniesionych nakładów do realnych korzyści wynikających z ich zastosowania. Richard K. Bergman (2011) stwierdził, że twórcy serwisów WWW będą wybierali formaty, które są łatwe do zrozumienia, posiadają kompletną dokumentację i są tym samym łatwe do implementacji. Zdaniem Bergmana takim wyborem są mikro dane. Chociaż przywoływany pogląd pochodzi z 2011 r., to zaprezentowane poniżej wyniki badań dotyczących zakresu wykorzystania FDS wskazują na jego aktualność.

Celem tej części artykułu jest określenie zakresu wykorzystania mikroformatów, RDFa, mikro danych oraz JSON-LD w zasobach WWW w celu strukturyzacji treści stron internetowych. Przedmiotem przeprowadzonych badań była dokumentacja zbioru danych badawczych, która powstała w ramach projektu Web Data Commons<sup>12</sup>. Projekt ten został zainicjowany w 2012 r. w ramach współpracy między niemieckimi instytucjami badawczymi – Wolnym Uniwersytetem Berlina (Freie Universität Berlin) oraz politechniką Karlsruher Institut für Technologie. Celem projektu było cykliczne ekstrakowanie ze zbioru Common Crawl<sup>13</sup> (największego i publicznie dostępnego korpusu zasobów sieciowych) informacji o stronach internetowych, w których ich twórcy wykorzystali FDS jako narzędzia formalnej strukturyzacji informacji. Do projektu włączono dodatkowo graf hiperłączy oraz kolekcję tabel, które derywowano z dokumentów HTML. Analizy zakresu wykorzystania wspomnianych formatów danych są prowadzone co roku począwszy od 2009 r. (z wyłączeniem 2011 r.). Od 2015 r. badania obejmują również format JSON-LD. Ostatnie badanie przeprowadzone w ramach projektu Web Data Commons zawiera dane reprezentujące stan na październik 2016 r.

Tab. 3. Ogólne statystyki badania Web Data Commons z 2016 r.

| Zmienna  | Wartość        |
|--|----------------|
| Liczba przetworzonych URL                                      | 3 181 199 447  |
| Liczba URL zawierających deklaracje                            | 1 242 727 852  |
| Liczba przetworzonych adresów domenowych (PLD)                 | 34 076 469     |
| Liczba domen (PLD), których zasoby zawierają dane strukturalne | 5 638 796      |
| Liczba zdefiniowanych obiektów                                 | 9 590 731 005  |
| Liczba deklaracji  | 44 242 655 138 |

Metodologia projektu Web Data Commons zakłada automatyczną rejestrację informacji nie tylko o zastosowanym formacie danych strukturalnych na danej stronie WWW, ale

<sup>12</sup> <http://webdatacommons.org/>

<sup>13</sup> <http://commoncrawl.org/>

również ekstrahowanie całego zbioru danych ustrukturyzowanych z określonego źródła sieciowego<sup>14</sup>. Według danych z października 2016 r. przeanalizowano ponad 3.2 mld adresów URL z czego ponad 1.2 mld stron internetowych zawierały formalnie specyfikowane deklaracje świadczące o obecności danych ustrukturyzowanych (Tab. 3). Łącznie przeanalizowano ponad 54 terabajty danych.

Przez deklarację rozumie się tutaj elementarną jednostkę wypowiedzi z wykorzystaniem metadanych w układzie <obiekt>-<atribut>-<wartość>, którą w badaniu określa się mianem trójki (ang. *triple*), co nawiązuje do modelu danych RDF. Z uogólnionych danych przedstawionych w tabeli 3 wynika, że dane ustrukturyzowane były obecne prawie w 40% stron internetowych poddanych badaniu. Na potrzeby metodologii badań wprowadzono pojęcie domeny poziomu płatnego (ang. *Pay-Level Domain*; PLD), przez które rozumie się subdomenę wyodrębnioną w ramach domeny najwyższego poziomu (np. .pl, .gov, .edu)<sup>15</sup>. Jest to więc indywidualna nazwa domenowa, do której używania należy nabyć prawa. PLD wykorzystano jako zmienną do analiz przede wszystkim ilościowych, tzn. do określenia zakresu wykorzystania danych ustrukturyzowanych w ramach określonego serwisu internetowego. Z danych przedstawionych w tabeli 3 wynika, że dane ustrukturyzowane były obecne w 16% źródeł internetowych. Przez obiekt zdefiniowany (ang. *typed entity*) rozumie się obiekt, który został formalnie skategoryzowany, tzn. istnieje deklaracja (trójka), która zawiera informacje o typie obiektu, który on reprezentuje (np. wpis na blogu, przepis kulinarny, osoba, książka, itp.)<sup>16</sup>. Z udostępnionych danych wynika więc, że na jednej stronie internetowej, zawierającej dane ustrukturyzowane obecne było średnio 35 deklaracji wskazujących na własności średnio osiem zdefiniowanych obiektów.

Tab. 4. Porównanie zmian w wartościach zmiennych analizowanych w badaniach z lat 2012–2015

|  | 2013    | 2014   | 2015   | 2016   |
|--|---------|--------|--------|--------|
| Odsetek przetworzonych URL             | -25.98  | -9.47  | -12.10 | +44.34 |
| Odsetek URL zawierających deklaracje   | +58.64  | +5.87  | -12.68 | +56.43 |
| Odsetek przetworzonych domen           | -68.40  | +22.11 | -8.04  | +57.71 |
| Odsetek domen zawierających deklaracje | -22.15  | +52.95 | +0.08  | +51.68 |
| Odsetek zdefiniowanych obiektów        | +135.42 | +29.35 | +10.72 | +36.32 |
| Odsetek deklaracji                     | +134.55 | +18.81 | +19.00 | +44.90 |

W tabeli 4 zestawiono aktualne dane dotyczące próby badawczej i zakresu wykorzystania omawianych formatów danych ze stanem z lat ubiegłych. Tabela prezentuje zmiany w wartościach omawianych wcześniej zmiennych, które zaszły w stosunku do roku poprzedniego.

<sup>14</sup> Dane badawcze z listopada 2016 r. to zbiór ponad 9000 plików o łącznej wielkości 967 gigabajtów. Tym samym istnieje możliwość pobrania całego zbioru danych i przeprowadzania na nim dalszych analiz. Dane badawcze z października 2016 r. są dostępne pod adresem [http://webdatacommons.org/structureddata/2016-10/stats/how\\_to\\_get\\_the\\_data.html](http://webdatacommons.org/structureddata/2016-10/stats/how_to_get_the_data.html)

<sup>15</sup> Np. stronę internetową <http://www.przyklad.com.pl/zasoby/strona1.html> oraz <http://www.przyklad.com.pl/zasoby/strona5.html> kwalifikowano do jednej PLD – <http://przyklad.com.pl>

<sup>16</sup> W przypadku mikroformatów jest to wystąpienie atrybutu `class=typ`, RDFa – `typeof=typ`, mikro dane: `itemtype=typ`, JSON-LD – `@type=typ`.



W ciągu ostatnich lat analizowane były zbiory o różnej wielkości; można zauważyć ciągły wzrost liczby zarówno deklaracji, jak i formalnie zdefiniowanych obiektów, o których informacje są zgarniędżane na stronach WWW. Jest to szczególnie widoczne dla 2016 r.

Największy wzrost odsetka stron WWW zawierających formalnie specyfikowane deklaracje na temat ich treści w stosunku do roku wcześniejszego zaobserwowano w 2013 r. (ponad 58%), przy jednocześnie mniejszej próbie badawczej (spadek o ponad 25%). W tym roku odnotowano również wzrost o ponad 130% zarówno odsetka deklaracji, jak i formalnie zdefiniowanych elementów w treści stron WWW. W latach 2014–2015 widoczny jest nieznaczny wzrost zainteresowania publikowaniem danych ustrukturyzowanych w treści stron WWW, natomiast dane za 2016 r. pokazują duży skok w odniesieniu do wszystkich zmiennych.

Tab. 5. Średnia liczba zdefiniowanych obiektów i deklaracji w treści stron internetowych wykorzystujących formaty danych strukturalnych z lat 2012–2016

|                                      | 2012 | 2013 | 2014 | 2015 | 2016 |
|--------------------------------------|------|------|------|------|------|
| Średnia liczba obiektów na stronie   | 5    | 7    | 9    | 11   | 8    |
| Średnia liczba deklaracji na stronie | 20   | 29   | 33   | 45   | 36   |

Z punktu widzenia ilości informacji specyfikowanych za pomocą omawianych formatów danych (Tab. 5) od 2012 r. utrzymuje się tendencja wzrostowa. Dotyczy to zarówno średniej liczby zdefiniowanych obiektów, jak i liczby deklaracji. Spadek w 2016 r. nie oddaje tego stanu, ponieważ zbiór badanych stron WWW zwiększył się prawie o połowę.

Tab. 6. Wyniki badania Web Data Commons z października 2016<sup>17</sup>

| Format                     | PLD       | URL         | Obiekty zdefiniowane | Deklaracje     |
|----------------------------|-----------|-------------|----------------------|----------------|
| html-microdata             | 2 537 539 | 901 118 191 | 6 872 341 887        | 34 637 805 559 |
| html-embedded-jsonld       | 2 116 755 | 111 411 049 | 385 731 201          | 1 880 721 886  |
| html-mf-hcard <sup>1</sup> | 1 668 039 | 159 748 255 | 1 614 688 960        | 4 600 477 456  |
| html-rdfa                  | 93 883    | 311 533 110 | 511 555 208          | 2 216 933 416  |
| html-mf-xfn                | 195 595   | 24 242 546  | 48 011 285           | 300 764 344    |
| html-mf-adr                | 188 755   | 27 697 569  | 80 039 476           | 259 718 235    |
| html-mf-geo                | 238       | 6 151 013   | 14 644 289           | 25 733 274     |
| html-mf-hcalendar          | 22 313    | 3 450 075   | 33 962 568           | 177 931 362    |
| html-mf-hreview            | 16 984    | 4 551 011   | 13 680 480           | 79 631 745     |
| html-mf-hlisting           | 471       | 37 418      | 9 578 853            | 36 521 676     |
| html-mf-hrecipe            | 2 923     | 755 544     | 5 695 917            | 24 347 685     |
| html-mf2-h-adr             | 1 415     | 1 362       | 26 659               | 726 401        |
| html-mf-hresume            | 168       | 2 961       | 7 106                | 22 262         |
| html-mf-species            | 95        | 170 516     | 527 185              | 1 319 837      |

<sup>1</sup> html-mf-hcard: format dla danych kontaktowych; html-mf-xfn: format danych do specyfikacji relacji interpersonalnych; html-mf-adr oraz html-mf2-h-adr: formaty danych teleadresowych; html-mf-geo: format danych dla lokalizacji geograficznej; html-mf-hcalendar: format danych dla wydarzeń; html-mf-hreview: format danych dla recenzji; html-mf-hlisting: format danych dla ofert handlowych; html-mf-hrecipe: format danych dla przepisów kulinarnych; html-mf-hresume: format danych dla Curriculum Vitae; html-mf-species: format danych dla reprezentacji informacji z obszaru taksonomii biologicznych.

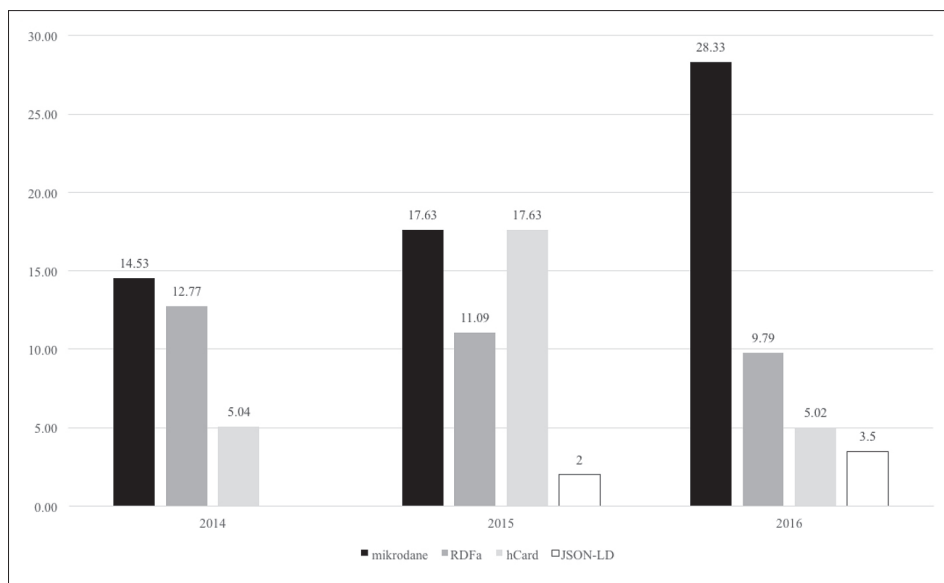
<sup>17</sup> <http://www.webdatacommons.org/structureddata/2016-10/stats/stats.html>



Aktualny obraz zakresu strukturyzacji informacji wewnątrz dokumentów HTML zawierają dane z października 2016 r. W tabeli 6 przedstawiono rozkład wartości omawianych wcześniej zmiennych w odniesieniu do mikroformatów (html-mf-xxx), mikro danych (html-microdata), RDFa (html-rdfa) oraz formatu JSON-LD (html-embedded-jsonld). Z punktu widzenia liczby stron internetowych (URL), w których treści obecne są te formaty największy odsetek wskazań odnosi się do zastosowania specyfikacji mikro danych (58%) i RDFa (20%). Pozostałe ok. 15% wskazań dotyczy poszczególnych specyfikacji mikroformatów, a ok. 7% wskazań – stron WWW zawierających dane strukturalne w formacie JSON-LD.

Z punktu widzenia PLD największy odsetek odnotowano jednocześnie dla mikro danych (31.3%) oraz mikroformatów (31.2%). Standard RDFa jest obecny tylko w ok. 15% serwisów internetowych. Z punktu widzenia liczby deklaracji, a więc ilości informacji formalnie specyfikowanych w strukturze stron internetowych, w ponad połowie przypadków (54.2%) są to deklaracje z wykorzystaniem standardu mikro dane. W ok. 37 % przypadków metadane wyrażone zostały za pomocą mikroformatów, a w ok. 6% – w standardzie RDFa.

Rysunek 2 zawiera zestawienie danych wskazujących na realne wykorzystanie FDS w latach 2014–2016. Do analiz wybrano mikro dane, RDFa, JSON-LD oraz najczęściej stosowaną specyfikację mikroformatów – hCard (zob. Tab. 6). Przez realne wykorzystanie FDS rozumie się tutaj stosunek liczby adresów URL, w których zidentyfikowano wystąpienie danego FDS do całego zbioru adresów URL, który został poddany badaniu w danym roku.



Rys. 2. Wykorzystanie FDS w latach 2014–2016

Z danych przedstawionych na rysunku 2 wynika, że w zbiorze danych badawczych można zaobserwować wzrost liczby implementacji mikro danych w ostatnich trzech latach. Prawie co trzecia strona WWW w zbiorze danych badawczych za 2016 r. zawierała formatowanie z wykorzystaniem mikro danych. Format RDFa w kolejnych latach był stosowany na podobnym poziomie ok. 10%, natomiast implementacja mikroformatu hCard

wykazuje wahania i wzrost w 2015, a następnie 50% spadek w kolejnym roku. Można zatem postawić więc pytanie, czy spadek ten miał wpływ na wzrost zastosowania mikrodanych i przyjęcie przez twórców serwisów internetowych tego formatu za podstawowy sposób strukturyzacji treści stron internetowych. W zaprezentowanych analizach widoczne jest również niewielkie zainteresowanie implementacją formatu JSON-LD, które odnotowano w 2015 r., ale wartość tej zmiennej jest stosunkowo mała.

## 5. Zakończenie

Formaty danych strukturalnych przeznaczone do reprezentacji treści stron internetowych zapewniają nowe możliwości strukturyzacji treści w ramach systemu znaczników języka HTML i tym samym rozszerzają zakres możliwych funkcjonalności mechanizmów wyszukiwawczych, zarówno implementowanych wewnątrz serwisów WWW, jak i wyszukiwarek internetowych o zasięgu globalnym. Taki sposób reprezentacji informacji w porównaniu z metodami reprezentacji wiedzy w Sieci Semantycznej może być określany mianem „płytkiej semantyki” (zob. Hitzler et al., 2012) i wpisuje się w koncepcję wyszukiwania semantycznego (ang. *semantic search*; *entity search*). „Płytki” charakter semantyki odwzorowywanej za pośrednictwem FDS polega tutaj na zapewnieniu modelu danych i środków syntaktycznych do tworzenia stosunkowo prostych deklaracji na temat faktów zapisanych w treści stron internetowych, z wykorzystaniem formalnie specyfikowanych schematów metadanych oraz elementów ontologii sieciowych. Oznacza to przyjęcie określonego konsensusu w procesie reprezentacji informacji, który jest zdeterminowany pragmatycznie. Znajdująca się na drugim biegunie „głęboka semantyka” jest metaforą sieciowych baz wiedzy operujących ontologiami, gdzie zastosowane formaty danych (np. RDF, OWL) pozwalają zarówno na szczegółową reprezentację informacji, jak i operowanie językiem logiki w celu wnioskowania nowych faktów.

Można więc przyjąć, że implementacja FDS ma na celu zwiększenie efektywności wyszukiwania zasobów sieciowych przez zastosowanie prostych technologii semantycznych, których celem jest formalna identyfikacja i kategoryzacja elementów treści stron WWW oraz odwzorowanie podstawowych faktów na ich temat z wykorzystaniem środków syntaktycznych danego FDS. Taki sposób reprezentacji informacji w zasobach WWW można również interpretować w kontekście semantycznych adnotacji (ang. *semantic annotation*), przez które rozumie się zarówno proces dodawania automatycznie przetwarzalnych metadanych w treści dokumentu, jak i jego efekt, czyli zbiór formalnie specyfikowanych faktów ekstrahowanych z jego treści (Oren et al. 2006). Na wybór określonego formatu, oprócz przytoczonych wcześniej kryteriów, powinna mieć wpływ również odpowiedź na pytanie, czy jest on wspierany przez globalne wyszukiwarki internetowe. Obecność FDS w zasobach WWW można zaobserwować, np. w postaci formy prezentacji rezultatów wyszukiwania m.in. w wyszukiwarkach Bing oraz Google. W przypadku Google mamy do czynienia z generowaniem listy adresów zasobów sieciowych spełniających kryteria wyszukiwania oraz często prezentacją zagregowanych faktów na temat przedmiotu wyszukiwania, które prezentowane są bezpośrednio pod nazwą zasobu sieciowego w rezultatach wyszukiwania oraz w panelu bocznym w interfejsie graficznym przeglądarki internetowej. Wykorzystanie globalnego identyfikatora sieciowego pozwala bowiem na agregację danych na temat

danego obiektu w rozproszonym środowisku sieciowym World Wide Web. Taką praktykę stosują dostawcy wspomnianych wyszukiwarek internetowych, co w przypadku Google przejawia się generowaniem wyników rozszerzonych, czy to w postaci dodatkowych informacji o wyszukanych zasobach sieciowych prezentowanych bezpośrednio w rezultatach wyszukiwania (tzw. *rich snippets*), czy też w postaci ustrukturyzowanych metadanych na temat obiektu zidentyfikowanego przez algorytm Google jako przedmiot wyszukiwania, które prezentowane są m.in. w panelu bocznym (Rys. 3).

The image shows a search result for the movie 'Dunkierka' (Dunkirk) on Google.pl. It is divided into two main sections: a left sidebar and a right main panel.

**Left Sidebar:**

- Godziny seansów: Dunkierka** (CET)
- Buttons for 'Dzisiaj' and 'Jutro'.
- Filter tabs: 'Wszystkie', 'Rano', 'Popołudniu', 'Wieczorem', 'Późnym wieczorem'.
- Listing of theaters and showtimes:
  - Helios Piotrków Trybunalski - Pokaz na mapie, 21:30
  - Helios Belchatow - Pokaz na mapie, 21:15
  - Multikino Silver Screen Łódź - Pokaz na mapie, 15:15, 17:40, 20:05
- Link to 'Więcej seansów'.
- Summary: 'Dunkierka (2017) - Filmweb', rating 7.8/10, and a short description.
- Wikipedia links for 'Dunkierka' and 'Dunkierka (film 2017)'.

**Right Main Panel:**

- Dunkierka** (2017 - Dramat filmowy) - 2 godziny.
- Movie poster image.
- Summary: 'Dunkierka – amerykańsko-brytyjsko-francusko-holenderski film wojenny w reżyserii Christophera Nolana, którego premiera miała miejsce 13 lipca 2017 roku w Londynie. Wikipedia'.
- Data premiery:** 13 lipca 2017 (Londyn)
- Reżyseria:** Christopher Nolan
- Autor muzyki:** Hans Zimmer
- Operator:** Hoyte van Hoytema
- Scenariusz:** Christopher Nolan
- Obsada:** Harry Styles, Tom Hardy, Cillian Murphy, Mark Rylance, Kenneth Branagh.
- Podobne wyszukiwania:** Dunkirk, Interstellar, Mroczny rycerz powstaje, Incepcja, Mroczny rycerz.

Rys. 3. Przykład rezultatów wyszukiwania w wyszukiwarce Google.pl dla zapytania „Dunkierka”

Na rysunku 3 przedstawiono fragment prezentacji rezultatów wyszukiwania w wyszukiwarce Google.pl dla zapytania „Dunkierka”. Wykaz wyników oraz elementów ustrukturyzowanych prezentowanych użytkownikowi pozwala na postawienie wniosku, że mechanizm tej wyszukiwarki na podstawie zastosowanych algorytmów przetwarzania własnej bazy, zidentyfikował przedmiot wyszukiwania jako film pt. „Dunkierka”. W panelu bocznym zaprezentowano wybrane fakty na temat tego filmu oraz dodatkowo na podstawie geolokalizacji komputera, z którego wysłano zapytanie, wskazano na kina w pobliżu, w których wyświetlany jest ten film. Zaprezentowane informacje o charakterze faktograficznym są wynikiem agregacji danych, które przeprowadził mechanizm indeksujący i z dużą pewnością można założyć, że informacje te są zapisane z wykorzystaniem formatów danych strukturalnych<sup>18</sup>.

Publikowanie informacji w WWW z wykorzystaniem formatów danych strukturalnych z pewnością stanowi również szansę dla bibliotek i repozytoriów cyfrowych z jednej strony,

<sup>18</sup> Obecność danych strukturalnych można zweryfikować za pomocą wielu aplikacji dostępnych online. Jedną z nich jest „Narzędzie do testowania uporządkowanych danych” oferowane przez Google. Zob. np. test zawartości rekordu dla filmu pt. „Dunkierka” w serwisie Filmweb.pl – <https://search.google.com/structured-data/testing-tool/u/0/#url=http%3A%2F%2Fwww.filmweb.pl%2Ffilm%2FDunkierka-2017-681141>

na lepszą „widoczność” tych kolekcji bezpośrednio w rezultatach wyszukiwania w wyszukiwarkach globalnych, a z drugiej, na włączenie ich do globalnej sieci danych (ang. *Web of Data*) poprzez wykorzystanie stosunkowo prostych technologii semantycznych. Wybór konkretnego formatu danych strukturalnych ma tutaj oczywiście znaczenie, ale wydaje się że mając na uwadze zachowania i kompetencje informacyjne użytkowników, za ważniejszą uznać trzeba świadomość konieczności implementacji takich technologii.

## Bibliografia

- Bergman, M. (2011). Structured Web Gets Massive Boost [online]. AI3[20.08. 2017], <http://www.mkbergman.com/962/structured-web-gets-massive-boost/>
- Berners-Lee, T. (1994). The World-Wide Web. *Communications of the ACM* 1, 37(8), 792–799.
- Berners-Lee, T., Hendler, J., Lassila, O. (2001). The Semantic Web [online]. *Scientific American* (May 17), [20.08.2017], <http://www.scientificamerican.com/article.cfm?id=the-semantic-web>
- Bizer, C., Heath, T., Berners-Lee, T. (2009). Linked Data – The Story So Far. *International Journal on Semantic Web and Information Systems* [online], 5(3), [20.08.2017], <http://tomheath.com/papers/bizer-heath-berners-lee-ijswis-linked-data.pdf>
- Bizer, C., Mendes, P. N., Jentzsch, A. (2012). Topology of the Web of Data. In: R. De Virgilio, F. Guerra & Y. Velegarakis (eds.), *Semantic Search over the Web* (3–29). Berlin, Heidelberg: Springer, <http://doi.org/10.1007/978-3-642-25008>
- Guha, R. V., Brickley, D., Macbeth, S. (2015). Schema.org: Evolution of Structured Data on the Web. *ACMQUEUE* [online], 9(13). [20.08.2017], <http://queue.acm.org/detail.cfm?id=2857276>
- Herman, I., Adida, B., Sporny, M. (2015). *RDFa 1.1 Primer – Third Edition. Rich Structured Data Markup for Web Documents* [online]. W3C [20.08.2017], <https://www.w3.org/TR/rdfa-primer/>
- Hitzler, P., Janowicz, K., Berg-Cross, G., Sheth, A., Finin, T., Cru, I. (2012). *Semantic Aspects of EarthCube* [online]. EarthCube [20.08.2017], <https://www.earthcube.org/document/2012/semantic-aspects-earthcube>
- McCarthy Nevele, C., Brickley, D. (2017). *HTML Microdata* [online]. W3C Working Draft 26 June 2017, [20.08.2017] <https://www.w3.org/TR/microdata/>
- Oren, E., Möller, K. H., Scerri, S., Handschuh, S., Sintek, M. (2006). *What Are Semantic Annotations?* [online]. Prof. Siegfried Handschuh [20.08.2017] <http://www.siegfried-handschuh.net/pub/2006/whatissemannot2006.pdf>
- Ronallo, J. (2012). HTML5 Microdata and Schema.org. *The Code4Lib Journal* [online], (16). [20.08.2017], <http://journal.code4lib.org/articles/6400>
- Sikos, L. F. (2015). *Mastering Structured Data on the Semantic Web: From HTML5 Microdata to Linked Open Data*. Berkeley, CA: Apress.
- Sporny, M. (2015). An Uber-Comparison of RDFa, Microdata and Microformats [online]. Internet Archive Wayback Machine [20.08.2017], <https://web.archive.org/web/20160329022408/http://manu.sporny.org/2011/uber-comparison-rdfa-md-uf/>
- Sporny, M., Kellogg, G., Lanthaler, M. (2014). JSON-LD 1.0. *A JSON-based Serialization for Linked Data* [online]. W3C [20.08.2017], <https://www.w3.org/TR/json-ld/>
- Tomberg, V., Laanpere, M. (2009). RDFa versus Microformats: Exploring the Potential for Semantic Interoperability of Mash-up Personal Learning Environments [online]. In: F. Wild, M. Kalz, M. Palmer & D. Muller (eds.), *Mash-Up Personal Learning Environments. Proc. of the 2nd Workshop MUPPLE'09, Nice, France, September 29, 2009, CEUR* (102–109). CEUR. [20.08.2017], <http://ceur-ws.org/Vol-506>
- Wetherill, B. (2014). RDFa and Microdata. *Library Philosophy and Practice (E-Journal)* [online], 1151, 19. [20.08.2017], <http://digitalcommons.unl.edu/libphilprac/1151/>

## Structured Data Formats for World Wide Web

### Abstract

**Purpose/Thesis:** The aim of this paper is the analysis and comparison of data formats for the content representation of Web pages embedded in HTML structure. The subjects of investigation are four structured data formats: microformats, RDFa, microdata and JSON-LD and their implementation on the Web.

**Approach/Methods:** The starting points for the investigation are two statements. The first one is that the separation between content and presentation layer is one of important features of the World Wide Web and the second refers to the fact that the structure level of Web content is the determining factor for the types of functionality that search engines can provide. These two approaches offer the background for the concept of structured data formats aimed at the formal representation of Web page content using HTML language system. The subjects were selected based on the scope of their implementation on the Web. The analysis was based on the assumption that structured data formats may be investigated from the metadata perspective with the premise that the annotation act is not made on the document level but is related to the facts that constitute the content. The study on structured data formats is based on semantic and syntactic analysis of their features.

**Results and conclusions:** Structured data formats for the content representation of Web pages provide new methods for knowledge representation by means of HTML language and thus extend the functionalities of both locally implemented and global search mechanisms. The results of the survey conducted in the years 2012–2016 indicate the growth of the interest in the semantic representation of Web pages. RDFa represents a high level of specificity but microdata seem to be the consensus between the desired expressiveness and the ease of implementation, confirmed with the results of Web Data Commons project.

**Originality/Value:** The comparison of selected features of four structured data formats offers a clear picture of their capability for deep content annotations with metadata schemes and ontologies. The results from Web Data Commons project for the period 2014–2016 indicate that microdata and schema.org will play an important role in the domain of applying semantic technologies to Web page content representation.

### Keywords

JSON-LD. Linked Data. Metadata. Microdata. Microformats. RDFa. Schema.org. Semantic Web.

---

*Dr MARCIN ROSZKOWSKI jest adiunktem w Katedrze Informatologii na Wydziale Dziennikarstwa, Informacji i Bibliologii Uniwersytetu Warszawskiego. Jest członkiem International Society for Knowledge Organization oraz Komitetu ds. ontologii w projekcie DBpedia. Jego zainteresowania naukowe obejmują problematykę organizacji wiedzy i reprezentacji informacji w środowisku sieciowym, ze szczególnym uwzględnieniem modelowania conceptualnego systemów informacyjnych oraz metadanych i ontologii sieciowych. Najważniejsze publikacje: B. Sosińska-Kalata, M. Roszkowski (2016). Organizacja informacji i wiedzy. W: W. Babik (red.), Nauka o informacji (305–358). Warszawa: Wydaw. SBP; M. Roszkowski, W. Mustafa El Hadi (2016). The Role of Digital Libraries as Virtual Research Environments for the Digital Humanities. In: J. A. C. Guimarães, S. Oliveira Milani, & V. Dodebei (eds.), Advances in Knowledge Organization (Vol. 15). Ergon Verlag, 392–402; M. Roszkowski (2016). Kartoteka haseł wzorcowych jako usługa sieciowa – automatyczna identyfikacja nazw osobowych z wykorzystaniem kartoteki VIAF. W: J. Woźniak-Kasperek & J. Franke (red.), Bibliografia – teoria, praktyka, dydaktyka (203–222). Warszawa: Wydaw. SBP.*

*Kontakt z autorem*

*m.roszkowski@uw.edu.pl*

*Katedra Informatologii*

*Wydział Dziennikarstwa, Informacji i Bibliologii*

*Uniwersytet Warszawski*

*ul. Nowy Świat 69*

*00–046 Warszawa*

---

# Bibliometryczna metodologia prognozowania i oceny rozwoju dyscyplin naukowych. Analiza piśmiennictwa.

## Część 2. Badania porównawcze, hybrydowe, statystyczne, analizy dokumentów patentowych, ścieżek rozwoju dyscyplin oraz pozostałe oryginalne podejścia metodologiczne

Łukasz Opaliński

*Biblioteka Politechniki Rzeszowskiej im. I. Łukasiewicza*

---

### Abstrakt

**Cel/Teza:** Celem pracy jest dokonanie krytycznego przeglądu literatury naukowej dotyczącej metod ilościowych stosowanych w obszarze bibliometrycznego prognozowania przyszłego rozwoju dyscyplin naukowych. Przez „przeгляд krytyczny” rozumie się wskazanie i omówienie mankamentów prezentowanych w literaturze przedmiotu podejść do rozpatrywanego problemu.

**Koncepcja/Metody badań:** Przyjęta w pracy metoda polega na analizie literatury przedmiotu, stworzeniu typologii wykorzystywanych w ramach bibliometrii metod prognozowania i oceny rozwoju dyscyplin, podkreśleniu ograniczeń stosowanych w tym zakresie metod oraz wyciągnięciu wniosków końcowych. Postarano się również zaznaczyć obszary, w których przejawia się praktyczna stosowalność prognoz tempa i kierunku rozwoju dyscyplin nauki.

**Wyniki i wnioski:** Głównym wnioskiem wypływającym z analizy literatury jest bardzo duże zróżnicowanie ilościowych metod przewidywania rozwoju dyscyplin, a także fakt, że często są one stosowane we wzajemnej izolacji. Po drugie, niemal każde z proponowanych podejść do omawianego problemu posiada różnego rodzaju wady, które niekorzystnie wpływają na uzyskiwane w jego ramach wyniki. Po trzecie, wydaje się, że szczególnie potencjał prognostyczny tkwi w metodzie statystycznej analizy szeregów czasowych, chociaż w praktyce badawczej metoda ta była wykorzystywana stosunkowo rzadko. Po czwarte, odnotowano wysoki stopień interdyscyplinarności metod bibliometrycznych, które często są modyfikacjami lub zapożyczeniami metod typowych dla matematyki, informatyki, statystyki matematycznej i ekonometrii. Ostatnim wnioskiem jest stwierdzenie, że prognozowanie odbywa się najczęściej przy użyciu specjalnie sprofilowanej analizy cytowań lub analizy liczb publikacji, które ogłoszono w danej dyscyplinie w różnych latach. U podstaw metodologii opartej na analizie cytowań leży konieczność zaakceptowania funkcjonującego w socjologii wiedzy naukowej tzw. paradygmatu normatywnego, jako epistemologicznego uzasadnienia dla badania rozkładów cytowań. Metodologia oparta na ocenie liczb publikacji dyscyplinarnych jest z kolei związana z toczącą się w informatologii dyskusją na temat tematycznej relewancji publikacji naukowych.

**Oryginalność/Wartość poznawcza:** Oryginalność i wartość poznawcza pracy polega na podjęciu próby przybliżenia rozległości i stopnia zaawansowania badań prowadzonych nad przewidywaniem i oceną stanu rozwoju dyscyplin na świecie. Badania te są realizowane w polskim środowisku naukowym stosunkowo rzadko, w związku z czym wydaje się, że przegląd przyjmowanej w ich ramach metodologii mógłby przyczynić się do wzbudzenia większego zainteresowania tą problematyką.



Zaakcentowanie słabych stron poszczególnych metod jest w tym kontekście podstawą do dalszego modyfikowania i udoskonalania poszczególnych metod. W obrębie całego spektrum bibliometrii istnieje bowiem jeszcze wiele przestrzeni dla tego rodzaju przedsięwzięć, ponieważ obecnie o żadnej z metod nie można powiedzieć, że idealnie spełnia ona postawione przed nią zadanie i jest „ostatnim słowem” w ramach całego nurtu prognostycznego.

### **Słowa kluczowe**

Bibliometria. Metody ilościowe w informatologii. Prognozowanie. Rozwój dyscyplin naukowych.

*Otrzymano: 3 stycznia 2017. Zrecenzowano: 10 lipca 2017. Zaakceptowano: 5 sierpnia 2017.*

## **1. Wprowadzenie**

Niniejszy artykuł stanowi drugą część i zarazem kontynuację analizy piśmiennictwa skupiającego się wokół problemu bibliometrycznego prognozowania i oceny bieżącego stanu rozwoju dyscyplin naukowych. W części pierwszej, omówiono publikacje pionierów tego kierunku badań oraz postarano się przedstawić, w jaki sposób analizy te „dojrzały”, aż do osiągnięcia kształtu, w jakim są one najczęściej prowadzone obecnie. W drugiej części, położono z kolei nacisk na metody alternatywne wobec metod tradycyjnych, na badania prowadzone w ujęciu porównawczym oraz na analizy głównych ścieżek rozwoju dyscyplin naukowych, którym nadano we współczesnym ujęciu charakter prognostyczny, jakiego brak znamionował wczesną postać tej metody bibliometrycznej.

Warto w tym miejscu dodać, że potrzeba realizacji badań porównawczych wynika przede wszystkim z braku szerszego konsensusu co do tego, która z metod tradycyjnych daje w efekcie jej zastosowania wyniki odpowiadające najbliższemu faktycznemu stanowi rzeczy. Inaczej mówiąc, motywem podejmowania badań komparatywnych jest niepewność co do tego, która z metod podstawowych (tj. metoda powiązań bibliograficznych, metoda współcycowań i metoda współwystępowania terminów naukowych) szybciej, dokładniej i pełniej identyfikuje fronty badawcze nadające bieg rozwojowi poszczególnych, wąskich obszarów problemowych. Przyczyn tego zjawiska można upatrywać najprawdopodobniej w fakcie, że wiele badań prognostycznych prowadzonych jest we wzajemnej izolacji, w oparciu o zróżnicowany materiał empiryczny, jak również w sytuacji braku obiektywnego, uniwersalnego punktu odniesienia dla wyników i wniosków wyciąganych w ramach tych badań. Większość bowiem tego, co możemy powiedzieć o rzeczywistych, istniejących realnie prądach intelektualnych kształtujących obraz współczesnej nauki pochodzi właśnie ze wspomnianych badań bibliometrycznych. W tym kontekście podkreśla się często m.in. subiektywność i niepewność ocen eksperckich, metody delfickiej, historycznych analogii czy ogólniej – metod jakościowych, które mogłyby ewentualnie stanowić alternatywne źródło wiedzy o stanie rozwoju nauki i jej perspektywach, a zarazem wspomniany punkt odniesienia dla wyników uzyskiwanych poprzez wykorzystywanie metod ilościowych (zob. np.: Daim et al., 2006, 981–985, 999; Jun et al., 2012). Wskazuje się też niekiedy na potrzebę integracji obydwu typów podejść metodologicznych do problemu prognozowania rozwoju nauki i technologii i podkreśla zarazem istniejący wciąż brak jednomyślności w tym zakresie (zob. np. Haegeman et al., 2013).

Cel drugiej części pracy pozostaje taki sam co w części pierwszej. Jest nim mianowicie dokonanie krytycznego przeglądu literatury naukowej dotyczącej metod ilościowych, stosowanych w obszarze bibliometrycznego prognozowania rozwoju dyscyplin naukowych, wraz

ze wskazaniem mankamentów poszczególnych podejść wykorzystywanych przez autorów zajmujących się tego rodzaju zagadnieniami. Niezmieniona jest również metoda badań, czyli analiza piśmiennictwa i wnioskowanie w oparciu o przedstawiane w niej główne tezy. Niniejszą, drugą część artykułu rozpoczyna przegląd badań porównawczych, w którym nawiązuje się bezpośrednio do wspomnianych wcześniej i wyjaśnionych szczegółowo w części pierwszej metod klasycznych (podstawowych).

## 2. Badania porównawcze

Kolejny typ badań to prace, w których bezpośrednio porównano ze sobą efektywność różnych metod grupowania dokumentów z wykorzystaniem omówionych wcześniej miar podobieństwa publikacji, tj. relacji współcytowania, relacji powiązania bibliograficznego oraz relacji cytowań bezpośrednich. Bo Jarneving (2005) porównał ze sobą pierwszą i drugą z wymienionych metod, w zakresie zdolności do odwzorowywania frontu badawczego. Jarneving porównywał ze sobą profile tematyczne dwóch grup publikacji. Pierwszą z nich uzyskał za pomocą standardowej techniki generowania powiązań bibliograficznych, uzupełnioną o normalizację siły powiązania ze względu na różnice w liczbie opisów występujących w bibliografiach załącznikowych (Jarneving, 2005, 250). Drugą grupę stanowił z kolei nie sam zbiór publikacji powiązanych relacją współcytowania (ang. *co-citation cluster*), ale zbiór publikacji powołujących się (cytujących), na którykolwiek element tego zbioru prac powiązanych współcytowaniami. Dzięki temu rozróżnieniu (tj. klastr publikacji współcytowanych versus publikacje cytujące ten klastr) można było porównać publikacje wydane w podobnym przedziale czasu (Jarneving, 2005, 246–247). Bez wprowadzenia tego rozróżnienia, które dostrzegł też m.in. Olle Persson w 1994 r., pomiędzy publikacjami powiązanymi bibliograficznie, a publikacjami powiązanymi relacją współcytowania istniałaby czasowa luka. Powiązania bibliograficzne można bowiem ustalić nawet dla dokumentów wydanych w roku bieżącym (np. 2016), podczas gdy utworzenie się relacji współcytowania wymaga czasu potrzebnego na akumulację dostatecznej liczby cytowań pojawiających się pewien czas po wydaniu prac, które te cytowania zyskują i dopiero potem są ze sobą wiązane współcytowaniem. Jako punkt odniesienia dla tego porównania przyjęto, jak już powiedziano, profil tematyczny uzyskanych klastrów. Określono go za pomocą słów zawartych w tytułach zgrupowanych publikacji. Słowa te poddano normalizacji leksykalnej (tj. usunięto „słowa stopujące”, zastosowano kontrolę synonimów i homonimów oraz normalizację ze względu na liczbę słów zawartych w całym klastrze itd.) (Jarneving, 2005, 252). Wyniki porównania okazały się jednak niekonkluzywne. Jarneving stwierdził, że sugerują one wprawdzie, iż ma się do czynienia z dwoma różnymi frontami badawczymi (pod względem charakterystyk treściowych). Nie można jednak było ostatecznie potwierdzić tego wniosku z uwagi na problem w ścisłym zdefiniowaniu poziomu, na którym dwa tak złożone obiekty jak klastry można określić jako podobne. W związku z tym Jarneving wskazał na potrzebę dalszych analiz porównawczych w warunkach kontrolowanych w wyższym stopniu, niż miało to miejsce w przypadku jego własnego badania<sup>1</sup> (Jarneving, 2005, 260–262).

<sup>1</sup> Kontrola ta mogłaby polegać np. na zastosowaniu ściśle określonego, dokładnie tego samego zbioru cytowanych pozycji bibliograficznych w przypadku konstruowania porównywanych obiektów, tj. klastrów prac powiązanych relacją współcytowania i relacją powiązania bibliograficznego (Jarneving, 2005, 262).

Naoki Shibata i in. (2009) badali w ujęciu porównawczym charakterystyki siatek cytoowań uzyskanych za pomocą metod powiązań bibliograficznych, współcytoowań i cytowań bezpośrednich pod kątem tego, które z tych podejść daje najlepsze rezultaty w zakresie detekcji najaktualniejszych frontów badawczych. Badanie przeprowadzono w ramach trzech specjalności usytuowanych na pograniczu fizyki i inżynierii materiałowej (Shibata et al., 2009, 572). Wskazane trzy metody porównano pod względem trzech aspektów wytworzonych przez nie sieci cytoowań: widoczności frontu (ang. *visibility*) mierzonej rozmiarami wykrytych klastrów, szybkości przesuwania się frontu (ang. *speed*) mierzonej średnim wiekiem publikacji w klastrze oraz relewancją/spójnością tematyczną frontu (ang. *topological relevance*) mierzoną siłą wzajemnych powiązań między elementami klastrów, a inaczej mówiąc tzw. gęstością klastrów (Shibata et al., 2009, 572, 574). Zdaniem Shibaty i in. najefektywniejsza jest ta metoda, która wykryłaby największe i najbardziej spójne klastry na jak najwcześniejszym etapie ich życia (Shibata et al., 2009, 573). Zastosowaną metodą wydzielenia klastrów była metoda tzw. grupowania topologicznego (ang. *topological clustering*) Marka Newmana (2004). Algorytm Newmana opierał się na takim dzieleniu wyjściowego zbioru elementów (publikacji) na mniejsze podzbiory, aby uzyskać strukturę sieciową, w której podzbiory są złożone z elementów powiązanych ze sobą ściśle (podzbiory mają dużą gęstość wewnętrzną), a powiązania pomiędzy elementami należącymi do różnych podzbiorów są dużo słabsze. Miarą tego, jak gęsty jest klaster jest tu tzw. modularność (ang. *modularity*) oznaczana symbolem „Q”. Określa ona różnicę między gęstością klastrów wygenerowanych przez algorytm, a gęstością przypadkową, tj. taką, której można byłoby oczekiwać przy całkowicie przypadkowym rozłożeniu się połączeń między wszystkimi elementami zbioru wyjściowego. Algorytm tak dzieli zbiór wyjściowy, aby zoptymalizować wartość „Q”, czyli wybrać najlepszy spośród wszystkich możliwych podziałów zbioru wyjściowego (Newman, 2004, 1–2; Shibata et al., 2008, 763; Shibata et al., 2009, 573). W wyniku porównania ustalono, że najlepszą metodą wykrywania frontów było cytowanie bezpośrednie. Metoda ta okazała się zdolna do najwcześniejszego wykrycia największych i najspójniejszych nowo powstałych klastrów i stwarzała zarazem najmniejsze ryzyko „przeoczenia” frontu (nieumyślnego wykluczenia go z klastra) (Shibata et al., 2009, 575, 577, 578–579). Najgorzej wypadła metoda współcytoowań, głównie ze względu na istnienie opóźnienia czasowego koniecznego do akumulacji cytoowań wokół publikacji (Shibata et al., 2009, 577). Gęstości klastrów wytworzonych w ramach tej metody były ponadto mniejsze, stwarzały wrażenie bardziej przypadkowych. Podobnie wyglądała spójność klastrów uformowanych przy użyciu metody powiązań bibliograficznych. Przewagą tej ostatniej metody nad techniką współcytoowań była natomiast wrażliwość na czas powstawania połączeń w sieci – w przypadku powiązań bibliograficznych nie występuje bowiem opóźnienie czasowe. Wraz z powiększaniem się (ewolucją) dyscypliny, przyłączaniem do niej kolejnych publikacji, metoda ta dawała jednak w wyniku coraz bardziej przypadkowe (mało spójne) klastry (Shibata et al., 2009, 575–579).

Badanie analogiczne do Shibaty i in. przeprowadzili rok później Kevin Boyack i Richard Klavans (2010). Główne różnice w stosunku do badania Shibaty i in. to włączenie w obszar zainteresowania dodatkowego wariantu metody współcytoowań (tzw. *co-citation analysis*) oraz wykorzystanie innego kryterium porównawczego. Pierwszym kryterium była – tak jak u Shibaty i in. – koherencja tekstualna uzyskanych klastrów, podczas gdy drugim była niestosowana dotąd, oryginalna propozycja Boyacka i Klavansa oparta na zamieszczanych

w artykułach informacjach na temat realizacji założeń (zadań) określonych grantów badawczych<sup>2</sup> (ang. *grant-to-article linkages*) (Boyack & Klavans, 2010, 2390, 2399). Wspomniany dodatkowy wariant metody współcytowań polegał na próbie usunięcia problemu opóźnienia czasowego poprzez dołączanie do wyodrębnionych z użyciem standardowego współcytowania klastrów artykułów bieżących (tj. artykułów spoza samych klastrów, które jednak cytują pozycje należące do klastrów). Decyzję podejmowano na podstawie tego, które z artykułów składających się na dane klastry były cytowane (przywoływane w bibliografiach załącznikowych) przez artykuły bieżące. Zabieg ten stanowił więc swego rodzaju złożenie metody powiązań bibliograficznych z metodą współcytowań (Boyack & Klavans, 2010, 2391, 2397). Uzyskane wyniki wskazały na następujące zależności:

- (1) Koherencja tekstualna klastrów była najwyższa dla metody złożonej, następnie kolejno spadała dla metod powiązań bibliograficznych, współcytowań i cytowań bezpośrednich (Boyack & Klavans, 2010, 2400).
- (2) Przewaga metody złożonej nad pozostałymi uwidoczniła się także w przypadku drugiego kryterium, chociaż metoda współcytowań osiągnęła w tym aspekcie niemal równie dobry wynik. Obie one znacznie przewyższyły w tym względzie pozostałe dwie metody (Boyack & Klavans, 2010, 2401).
- (3) Metoda złożona była ogólnie najlepsza pod każdym względem za wyjątkiem tzw. zasięgu (ang. *coverage* – wskaźnika procentowego określającego jaki odsetek zbioru wyjściowego został w ogóle zakwalifikowany do jakiegokolwiek klastra). Zasięg był o niecałe 2% wyższy w przypadku metody współcytowań.
- (4) Na drugim miejscu znalazła się metoda powiązań bibliograficznych – cechował ją niemal równie wysoki zasięg oraz precyzja (mierzoną metodą połączeń publikacji z grantami) w zakresie zdolności reprezentowania frontu badawczego.
- (5) Metoda współcytowań uplasowała się na miejscu trzecim, a metoda cytowań bezpośrednich okazała się najgorsza (Boyack & Klavans, 2010, 2402).

Mu-Hsuan Huang i Chia-Ping Chang (2015) otrzymali rezultat przypominający wyniki Boyacka i Klavansa w zakresie porównania skuteczności metody współcytowań z metodą powiązań bibliograficznych. Motywem podjęcia badania przez tych autorów był zaznaczony przez nich, istniejący wciąż brak jedności w środowisku naukowym co do tego, która z metod dokładniej i szybciej identyfikuje aktualne i przyszłe trendy rozwojowe w obrębie dyscyplin naukowych (Huang & Chang, 2015, 2043, 2049). Analiza porównawcza Huang i Chang objęła wysoko cytowane publikacje (20% prac najwyżej cytowanych w każdym kolejnym roku analizy) z lat 2000–2004 i 2005–2009, należące do pogranicza obszarów elektroniki i materiałoznawstwa (Huang & Chang, 2015, 2044). Minimalnym progiem siły współcytowania oraz siły powiązania bibliograficznego było pięć współcytowań i pięć wspólnych pozycji bibliograficznych (Huang & Chang, 2015, 2045). Po zastosowaniu operacji grupowania hierarchicznego zestawiono ze sobą jej wyniki, które przybrały postać zespołu klastrów jako odpowiedników frontów badawczych wraz z nadanymi im nazwami (przynależnością do pewnych kategorii tematycznych) i charakterystykami zmieniającymi

<sup>2</sup> Dane do przeprowadzenia tego porównania składały się z dopasowanych do siebie numerów grantów i identyfikatorów artykułów z bazy MEDLINE. Założeniem tego kryterium było uznanie, że artykuły, które realizowały te same granty (wykorzystywały środki finansowe z tych samych grantów), powinny skupić się w tych samych klastrach (Boyack & Klavans, 2010, 2399).

się dynamicznie pomiędzy pierwszym a drugim podokresem badania. Rozpatrywane charakterystyki to rozmiar frontu, zmiana tego rozmiaru (rozrost lub zmniejszanie się) lub nagłe pojawianie się albo też całkowity zanik frontu (Huang & Chang, 2015, 2049–2051). Ostatecznie Huang i Chang wyciągnęli wniosek, według którego metoda powiązań bibliograficznych była w stanie wykryć większą liczbę frontów badawczych na wcześniejszym etapie ich kształtowania się niż metoda współcytowań, co wskazało jednoznacznie na jej przewagę nad metodą współcytowań (Huang & Chang, 2015, 2052, 2054).

### 3. Badania hybrydowe (mieszane)

Badaniami „hybrydowymi” lub mieszanymi nazwano te, w ramach których połączono poszczególne (dwie lub więcej) metody wykrywania formowania się nowych lub przekształcania się pewnych klastrów publikacji, jako struktur wytyczających granice frontów badawczych.

Peter Van Den Besselaar i Gaston Heimeriks (2006) zauważyli np., że naukowcy jednocześnie posługują się pewnymi terminami celem opisu przedmiotu własnych badań oraz odnoszą się do literatury przedmiotu (cytując ją), celem wskazania tradycji badawczej, do której przynależą ich własna praca. Pojawiająca się w publikacjach terminologia nabiera w związku z tym specyficznego znaczenia w kontekście zestawu cytowanych pozycji bibliograficznych. Wychodząc z tego założenia, zaproponowano kombinację metody *co-word* z metodą powiązań bibliograficznych, zgodnie z którą wskaźnikiem podobieństwa dwóch publikacji jest liczba współwystąpień określonych terminów oraz, jednocześnie, pozycji przywoływanych w bibliografiach załącznikowych (ang. *word-reference combinations*) (Van Den Besselaar & Heimeriks, 2006, 381, 384). Zdaniem autorów ich propozycja wyeliminowała problem arbitralności w ustalaniu progu minimalnej siły związku pomiędzy dokumentami, która dopuszcza przyłączenie ich do klastra. Uwzględnia się tutaj bowiem wszystkie dokumenty, które charakteryzują się przynajmniej jednym współwystąpieniem zarówno terminu/frazy, jak i przypisu bibliograficznego odsyłającego do tej samej pracy (Van Den Besselaar & Heimeriks, 2006, 391).

Podobnie, połączenie metody tekstualnej z powiązaniem bibliograficznymi zasugerował Wolfgang Glänzel (2012). Jego zdaniem metoda tekstualna powinna być stosowana jako technika nadawania nazw (technika terminologicznego kodowania) klastrom wyodrębnionym w oparciu o metodę powiązań bibliograficznych oraz jako równorzędna miara siły związków między dokumentami tworzącymi klastry. Jednym z sygnałów wyłonienia się nowej specjalności jest w tym ujęciu, np. przesunięcie leksykalnego „punktu ciężkości” klastra (ang. *topic shift*) w kierunku innej niż dotychczasowa problematyki, co uwidacznia się właśnie w sferze tekstualnej (Glänzel, 2012, 195–196). Ponadto połączenie obu metod daje sposobność dostosowywania badań do specyfiki badanej dyscypliny. Przykładowo, w naukach społecznych i humanistycznych, w których cytowalność artykułów z czasopism jest na ogół niższa niż w naukach przyrodniczych i technicznych, można wprowadzić „ważenie” elementów bibliograficznych i tekstualnych. Element tekstualny mógłby posiadać większą wagę (pierwszeństwo) w obrębie nauk społecznych i humanistycznych. Co więcej, podobieństwo tekstualne może w takich przypadkach zastąpić „brakujące” powiązania o naturze bibliograficznej (tj. powiązania bibliograficzne w sensie teorii Kesslera) (Glänzel, 2012, 195–197, 200).



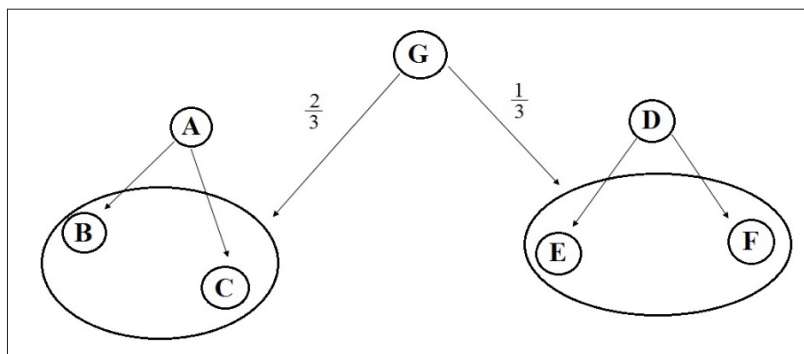
To samo hybrydowe kryterium podobieństwa publikacji wykorzystali Glänzel i Bart Thijs (2012). Było ono podstawą do ustalenia siły połączeń pomiędzy klastrami publikacji w jednym okresie a tzw. publikacjami kluczowymi (ang. *core documents*) w innym, późniejszym okresie rozpatrywanym w ramach analizy dynamiki i transformacji struktur klastrowych. Publikacje kluczowe w obrębie jednego klastra są publikacjami najsilniej ze sobą związanymi (najbliższymi sobie wg ustalonego kryterium podobieństwa), tj. dzielą ze sobą najwięcej jednostek bibliograficznych. Inaczej mówiąc, muszą one posiadać pewną wymaganą liczbę połączeń o określonej sile i powinny stanowić nie więcej niż 1% całego klastra. Można uważać je za reprezentantów klastra, a celem ich identyfikacji jest redukcja „szumu” (ang. *noise*), który wywołują publikacje powiązane luźniej z pracami centralnymi w obrębie klastra, leżące (na mapie) dalej od centrum i będące mniej relewantnymi dla tematyki przewodniej klastra (Glänzel & Thijs, 2012, 402–403; Glänzel, 2012, 196). Ustalenie charakteru związków między klastrami uformowanymi w pewnym czasie a dokumentami kluczowymi w innym czasie (w rozłącznych okresach obserwacji), do czego niezbędne jest wspomniane kryterium, było w ujęciu Glänzela i Thijsa głównym problemem w poszukiwaniu nowo powstających specjalności w dyscyplinie. Związki te polegają na wyłanianiu się nowych klastrow, zmianach rozmiaru klastrow, łączeniu się klastrow lub ich rozpadzie (Glänzel & Thijs, 2012, 401, 405–410). Rozpoznanie tych zależności pozwala wskazać, które obszary problemowe są aktualnie w fazie najintensywniejszego rozwoju i które z nich dyktują kierunki przesuwania się granic wiedzy w dyscyplinie jako całości.

Boyack i Klavans (2014) użyli metody hybrydowej nieco innego rodzaju w ramach przedsięwzięcia obejmującego utworzenie dynamicznej i globalnej mapy nauki za lata 1996–2011 na podstawie danych z bazy Scopus. Jak stwierdzili autorzy, „mapowanie” nauki jest dwuetapowe, tj. składa się z klasyfikacji (grupowania, podziału wyjściowego zbioru publikacji na klastry) i dalszej wizualizacji sieci cytowań pomiędzy nimi (Boyack & Klavans, 2014, 671). Operację grupowania oparto na złożeniu techniki generowania klastrow publikacji współcytowanych z techniką połączeń bibliograficznych. Ponieważ klastry publikacji współcytowanych nie zawierają publikacji najnowszych (leżą w przeszłości w stosunku do prac wydanych w roku bieżącym – w tym przypadku w 2011 r.), podjęto decyzję o uzupełnieniu zawartości tych klastrow o publikacje najnowsze. Jeżeli artykuł wydany w 2011 r. miałby np. 10 pozycji w bibliografii załącznikowej, które występowałyby zarazem w innym z klastrow publikacji współcytowanych, zostałby on przyłączony do tego właśnie klastra. Dopuszczalne jest tu także „przydzielanie ułamkowe” (ang. *fractional assignment*). Gdyby artykuł ten miał siedem pozycji bibliograficznych wspólnych z jednym klastrem, a trzy pozycje wspólne z drugim, zostałby on przydzielony w 70% do tego pierwszego, a w 30% do drugiego (Boyack & Klavans, 2014, 673–674, 677). Jest to ta sama operacja (złożenie metod), którą autorzy rozpatrywali w 2010 r. pod nazwą *co-citation analysis* i którą przedstawili poglądowo na rysunku (zob. Rys. 1).

Rysunek 1 przedstawia przypadek, w którym mamy do czynienia z sytuacją, w której artykuł „A” powołuje się na artykuły „B” i „C”, a artykuł „D” cytuje artykuły „E” i „F”. Publikacje cytowane przez artykuł „A” tworzą więc dwuelementowy klaster publikacji współcytowanych {B, C}, a publikacje cytowane przez artykuł „D” klaster {E, F}. Artykuły „A”, „G” i „D” wydano w 2011 r., podczas gdy artykuły „B”, „C”, „E” i „F” wcześniej. Artykuł „G” ma przykładowo tylko trzy pozycje w bibliografii załącznikowej: „B”, „C” i „E”. Stąd, może on zostać przypisany częściowo do klastra {B, C} i częściowo do klastra {E, F}, chociaż



pierwotnie nie wchodził on w skład żadnego z nich ponieważ nie był z żadnym z nich współcytowany (żadna publikacja nie cytowała jednocześnie artykułu „G” i np. artykułu „B”).



Rys. 1. Ułamkowe kwalifikowanie publikacji wydanej w roku bieżącym do klastrów publikacji utworzonych w oparciu o metodę współcytowania;  
źródło: oprac. własne na podstawie Boyack & Klavans, 2010, 2390

Druga modyfikacja metodologiczna Boyacka i Klavansa pojawiła się na etapie wizualizowania związków zachodzących między wygenerowanymi już klastrami, czyli na etapie rysowania map danych dyscyplin. Na tym etapie konieczne jest posłużenie się jakąś miarą wzajemnego podobieństwa (bliskości) klastrów (ang. *cluster-cluster similarity*) (Boyack & Klavans, 2014, 675). Stosowanie różnych miar doprowadza w konsekwencji do odmiennych wizualizacji. Boyack i Klavans dążyli do takiego układu mapy (ang. *cluster layout*), w którym widniałyby luźno ze sobą powiązane „wyspy” (ang. *sets of islands connected by pathways*) (Boyack & Klavans, 2014, 675–676). Jest to standardowy wymóg stawiany przed wynikami operacji podziału (grupowania) zbiorów: podział „dobry” to podział, który zapewnia uformowanie podzbiorów gęstych (o silnych połączeniach wewnętrznych) i zarazem dobrze wyodrębnionych od pozostałych (tj. takich, które posiadają znacznie słabsze połączenia „zewnętrzne” w sensie połączeń z innymi podzbiorami). Kryterium podobieństwa Boyacka i Klavansa miało charakter tekstualny – klastry reprezentowały frazy zawierające fragmenty tytułów i abstraktów publikacji zgrupowanych w ich ramach. Kryterium to miało postać wyrażenia matematycznego, w której pierwszorzędą rolę odgrywała częstość występowania określonych terminów w poszczególnych obiektach (klastrach) (Boyack & Klavans, 2014, 675–676).

Small, Boyack i Klavans (2014) zaproponowali kolejną metodę hybrydową mającą na celu umożliwienie wykrywania nowo powstających obszarów problemowych jeszcze w trakcie ich formowania się. Zadanie to stanowiło w ich ocenie przeciwieństwo analiz retrospektywnych, które, „patrzac wstecz”, były w stanie wskazać specjalności nowo powstające, wyłonione np. pięć lub więcej lat przed momentem realizacji badania (Small et al., 2014, 1452). Istotą metody Smalla i in. było dwukrotne zgrupowanie (podział na klastry) tych samych danych o publikacjach i ich cytowaniach, przy użyciu dwóch różnych kryteriów podobieństwa – relacji współcytowania oraz relacji cytowań bezpośrednich. Uzyskuje się w ten sposób dwa modele (mapy) oparte na innych sposobach organizacji danych i cechujące się inną charakterystyką. Następnie oba modele są na siebie „nakładane” (ang. *intersected*) w celu

zidentyfikowania obszarów (ang. *clusters of activity*), które w obu modelach w najbardziej wyraźny sposób manifestują cechy nowości/oryginalności (ang. *novelty, newness*) i wzrostu (ang. *growth*) (Small et al., 2014, 1452). Cechy te uważane są bowiem powszechnie za nieodłącznie związane z nowo wyłaniającymi się specjalnościami dyscyplinarnymi (Small et al., 2014, 1451). Ilościową techniką porównywania wyodrębnionych klastrów zaobserwowanych po nałożeniu na siebie obu modeli, pod względem osiągniętego przez nie stadium rozwojowego, była tzw. funkcja EP (ang. *Emergence Potential*). Pozwalała ona na „nominowanie” pewnych klastrów jako najlepszych kandydatów do miana załączków nowych specjalności (ang. *the most emergent*) (Small et al., 2014, 1453). Najpoważniejszym problemem w tej metodologii było wykryte opóźnienie czasowe pomiędzy zaistnieniem wydarzenia krytycznego dla uformowania się nowej specjalności (np. odkrycia), a jej wyłonieniem się i możliwością jej detekcji. Opóźnienie to, według szacunku Smalla i in. wynosiło 2.3 roku i było zbyt duże, aby można było wykorzystać w praktyce wiedzę o formowaniu się nowego obszaru (np. podczas podejmowania decyzji o finansowaniu badań) (Small et al., 2014, 1462). Dopiero znaczna minimalizacja tego opóźnienia przekształciłaby metodę z *de facto* opisowej w metodę autentycznie prognostyczną, w związku z czym konieczne jest dalsze poszukiwanie odpowiednich do tego sposobów adaptacji metody (Small et al., 2014, 1463).

Chengliang Liu i Quinchang Gui (2016), badając dyscyplinę geografii transportu pod kątem dokonujących się w niej przemian pomiędzy rokiem 1982 a 2014, połączyli w ramach kompleksowego podejścia technikę „mapowania” współcytowań autorów i współcytowań artykułów i czasopism, technikę analizy współwystąpień słów kluczowych, analizę przyrostu liczby publikacji w czasie wraz z liczbą ich cytowań oraz analizę geoprzestrzenną (ang. *geospatial analysis*). Ten ostatni element miał za zadanie odwzorować międzynarodową sieć współpracy naukowej w dyscyplinie poprzez wskazanie jej centrów (miast, regionów, instytucji badawczych i „niewidzialnych fakultetów”) oraz narodowościowego rozkładu zajmujących się daną tematyką autorów. Z tym zamiarem wykorzystano pakiety oprogramowania komputerowego: *CiteSpace*, *BibExcel* i *ArcGIS*, tworząc odrębne mapy dla poszczególnych, rozpatrywanych aspektów (Liu & Gui, 2016, 161–162, 164–165, 167, 170). Z punktu widzenia niniejszego przeglądu najistotniejszymi zadaniami podjętymi przez Liu i Guiego było identyfikowanie centralnych dla dyscypliny frontów badawczych i ich dynamiki, za pomocą narzędzia *CiteSpace*. Zapewniło ono autorom możliwość detekcji „eksplozji terminów” (ang. *burst terms*) wskazujących główne „trajektorie badawcze” (ang. *research trajectories*), czyli zagadnienia, na których w przeważającej mierze skupiła się uwaga społeczności naukowej. Na powstałej mapie miały one postać punktów połączonych krawędziami o różnej wielkości. Krawędzie reprezentowały relację współwystępowania terminów i siłę (skalę) tej relacji, a rozmiar punktu – częstość wykorzystywania terminów (popularność) w różnych okresach. Terminy, których częstość występowania wzrosła gwałtownie dopiero niedawno, zostały uznane za sygnał pojawiania się nowych kierunków badań w dyscyplinie (ang. *research hotspots/emerging hot issues*) (Liu & Gui, 2016, 173–175). Wykorzystując kolejną funkcję *CiteSpace*, zbadano wzorce współcytowań indywidualnych publikacji i dokonujące się w tym zakresie przemiany w różnych przedziałach czasu. Szczególną uwagę zwrócono na „eksplozje” cytowań pewnych prac (ang. *bursting references*), których tematyka w dużym stopniu wpłynęła na kierunek ekspansji geografii transportu (Liu & Gui, 2016, 175–178).

Czasowy przyrost liczby publikacji o określonej problematyce, obok analiz współcytowania publikacji i współwystępowania terminów, był także przedmiotem badania Pinga Liu

i in. (2016). Materiałem badawczym było 500 najwyżej cytowanych artykułów z zakresu fizyki, które ukazały się w latach 2000–2015. Liu i in. (2016), podobnie jak Liu i Gui (2016), obserwowali nie tylko najnowsze trendy badawczo-rozwojowe, ale też i inne właściwości charakterystyczne dla wybranej specjalności, takie jak kluczowe czasopisma, afiliacje autorów i kraje wiodące prym w zakresie produktywności i siły wpływu na całokształt badań w dyscyplinie (Liu et al., 2016, 1593–1595). Drugim podobieństwem do procesu badawczego Liu i Guiego (2016) było wyodrębnianie (w postaci osobnych map) klastrów prac współcytowanych i klastrów współwystępujących ze sobą słów kluczowych. Te publikacje, które znalazły się w centrum klastra (sieci współcytowań), tj. posiadały najwyższy wskaźnik tzw. „centralności pośredniczącej”<sup>3</sup> (ang. *betweenness centrality*) określono mianem „publikacji rewolucyjnych” (ang. *revolutionary publications*), a więc publikacji nadających kierunek dalszej aktywności społeczności badaczy i symbolizujących kluczowe dla danej specjalności osiągnięcia (Liu et al., 2016, 1598). Zmierzone też średni wiek publikacji wchodzących w skład klastrów, a następnie przeprowadzono podział (grupowanie) zbioru stu najczęściej występujących w parze słów kluczowych (Liu et al., 2016, 1600). W kolejnym etapie badania wytypowano terminy o najwyższym stopniu „centralności” i najwyższej częstości występowania w całym badanym zbiorze oraz klaster terminów o najniższej średniej wieku publikacji, z których pochodziły wyróżnione słowa kluczowe (Liu et al., 2016, 1600–1601).

Jeszcze jedną oryginalną, hybrydową metodę przewidywania przyszłych związków i transformacji w sieciach typu *co-word* przedstawili Nazim Choudhury i Shahadat Uddin (2016). Metoda grupowania współwystępujących, autorskich słów kluczowych (reprezentujących konkretne koncepcje) w podzbiory, została uzupełniona o techniki zaczerpnięte z tzw. statystycznej analizy szeregów czasowych<sup>4</sup> (ang. *time series approach*) (Choudhury & Uddin, 2016, 747–748). Pierwszym krokiem procedury Choudhury’ego i Uddina było wygenerowanie zestawu słów kluczowych dla trzech wybranych obszarów tematycznych i ich normalizacja (usunięcie znaków interpunkcyjnych, znaków diakrytycznych, zamiana liczby mnogiej rzeczowników na pojedynczą, ujednoczenie synonimów itd.). Ze słów kluczowych utworzono następnie mapy ich współwystępowania we wszystkich latach z przedziału 2010–2014 (przedział „T” reprezentujący „fazę próbną” – ang. *training phase*) i w roku 2015 (okres „T+1”) jako okresie testowym (ang. *test phase*). Dla jednego obszaru tematycznego (ze wszystkich trzech wybranych) mapy odwzorowywały relacje pomiędzy terminami dla lat 2008–2012 („T”), a rolę okresu testowego pełniły dane za rok 2013 („T+1”)

<sup>3</sup> „Centralność pośrednicząca” określana także w języku polskim jako „centralność oparta na pośrednictwie” lub „wskaźnik pośrednictwa” jest wskaźnikiem typowym dla analiz grafów i sieci (w tym także sieci cytowań – ang. *citation networks*) i określa, jak wiele najkrótszych ścieżek (połączeń między dwoma dowolnymi wierzchołkami w sieci) przebiega przez wierzchołek, dla którego częstość ta jest wyznaczana. Miernik ten można interpretować jako współczynnik wpływu, który wywiera dany punkt (artykuł) na przepływ informacji między elementami sieci (Shibata et al., 2007, 877; zob. też: Leydesdorff, 2007, 1304–1305).

<sup>4</sup> Analiza szeregów czasowych to jedna z powszechnie stosowanych w naukach ekonomicznych i społecznych metod prognozowania i oceny dynamiki zjawisk masowych. Analizy te polegają na symulacji przyszłego zachowania się wybranych zjawisk w oparciu o znane dane na temat przebiegu tych zjawisk w teraźniejszości i w przeszłości. Przykładowe dane, które są przedmiotem stosowania takich metod to np. ceny lub skala eksportu towarów, kursy walut, liczba ludności w danym regionie lub kraju itd. Szereg czasowy ma postać ciągu wartości liczbowych badanego zjawiska w kolejnych jednostkach czasu (zob. np.: Sobczyk, 2015, 307–356).

(Choudhury & Uddin, 2016, 753–755). Następnie zastosowano kilka matematycznych wskaźników<sup>5</sup> opisujących dynamikę struktury sieciowej w okresie „T”, z podziałem na pięć jednorocznych okresów obserwacji, aby przewidzieć związki (współwystąpienia) pewnych terminów (lub też ich brak) oraz trafność tych przewidywań w okresie „T+1” (Choudhury & Uddin, 2016, 748–751, 755–756). Z wyliczonych wskaźników zbudowano serie czasowe, które reprezentowały strukturę sieci w kolejnych latach i posłużyły za dane wejściowe dla trzech różnych technik prognostycznych: tzw. wygładzania wykładniczego, średniej ruchomej połączonej z procesem autoregresyjnym (ARIMA) oraz błędzenia losowego (Choudhury & Uddin, 2016, 751–752, 772; nt. metody ARIMA zob. np.: Montgomery et al., 2008, 231–286; Shumway & Stoffer, 2011, 83–171; Zeliaś et al., 2013, 234–243; nt. błędzenia losowego zob. np.: Aczel, 2007, 621–622; Klafter & Sokolov, 2015). Model pozwalał na przewidzenie trzech mechanizmów formowania się nowych połączeń między słowami kluczowymi: połączenia między słowami nowymi (tj. istniejącymi tylko w fazie testowej „T+1”), połączenia między słowami nowymi a „starymi” (tj. istniejącymi w okresie „T” oraz w okresie „T+1”) oraz nowe, nieistniejące dotąd połączenia między słowami „starymi” (obecnymi tylko w okresie „T”) (Choudhury & Uddin, 2016, 757, 772–773).

Chong-Chen Wang i Yuh-Shan Ho (2016) badali bieżące trendy obecne w obszarze chemii polimerów wykorzystując w tym celu technikę scharakteryzowaną w podrozdziale siódmym części pierwszej (przy okazji omawiania pracy Zhenga i in. (2016)) w połączeniu z obserwacją przyrostu liczby publikacji w czasie, dat ich wydania oraz przypisanych im słów kluczowych (autorskich i „dodatkowych”, dostępnych w WoS – *Key Words Plus*) (Wang & Ho, 2016, 483, 485–486). Słowa kluczowe zostały dodatkowo uzupełnione przez „słowa wspierające” (ang. *supporting words*), które zasugerowali eksperci zaznajomieni z wybranym obszarem nauki. Analizę słów, zawężono tylko do słów występujących w tytule, abstrakcie lub polu słów kluczowych WoS (ang. *front page filter*) (Wang & Ho, 2016, 483). Konkluzje co do przeważających w rozpatrywanej specjalności tendencji wyciągnięto po skonstruowaniu „klastrow słownictwa” (ang. *word clusters*) i dopełnieniu analizy za pomocą dokładniejszego przeglądu treści publikacji najwyżej cytowanych (od momentu wydania aż do roku 2014), które znalazły się w ramach wyszukanego zbioru (Wang & Ho, 2016, 494–504).

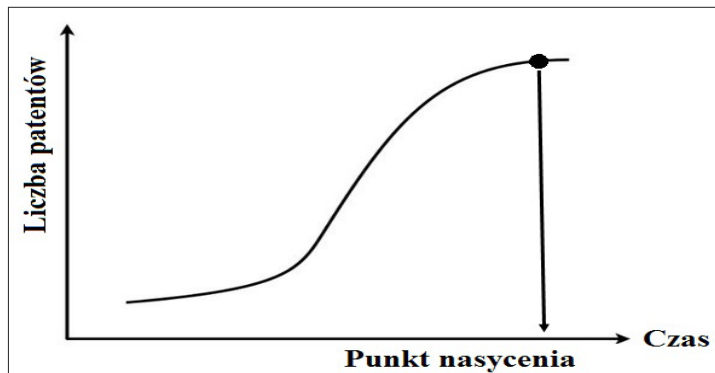
#### 4. Publikacje patentowe

W obrębie tej podgrupy podsumowano przedsięwzięcia, które za materiał badawczy obrały dokumenty patentowe. Często podkreślaną cechą patentów, która decyduje o ich podatności na badania zorientowane prognostycznie jest fakt, iż są to źródła piśmiennicze bezpośrednio odzwierciedlające technologiczny postęp i jego charakter (m. in. tempo czy kierunek). Warunkiem przyznania prawa patentowego jest bowiem nowatorskość (ang. *novelty*) i postępowość (ang. *inventive step*) (Debackere et al., 2002, 213), a same patenty – niejako

<sup>5</sup> Wskaźniki te były dwójakiego rodzaju: lokalne i globalne. Wszystkie one miały charakter topologiczny, tj. reprezentowały strukturalne podobieństwo par wierzchołków sieci (grafu), które zależało wyłącznie od struktury sieci jako całości. Przykładowym wskaźnikiem lokalnym było tu np. podobieństwo mierzone liczbą „wspólnych sąsiadów” dwóch wierzchołków w sieci (ang. *common neighbours*). Szczegółowe omówienie wszystkich wykorzystanych wskaźników wymagałoby jednak przypuszczalnie odrębnego opracowania. Więcej szczegółów nt. konstrukcji wskaźników podali sami Choudhury i Uddin (2016, 748–751).

z definicji – są bezpośrednim świadectwem nowych osiągnięć naukowych, które stymulują dalszy rozwój danej gałęzi techniki (Yoon & Park, 2007, 592; zob. też: Ena et al., 2016, 1025).

W pierwszej kolejności należy w tym kontekście wspomnieć o pracy Tugrula Daima i in. (2006), którzy proponowali zastosowanie połączenia kilku metod badawczych do analizy liczby patentów odnoszących się do pewnych technologii, ogłoszonych w różnym czasie. Metody te można podzielić na ilościowe (bibliometria, dopasowanie funkcji do danych empirycznych i jej dalsza ekstrapolacja) oraz jakościowe (perspektywa organizacyjna i personalna – budowanie scenariuszy w oparciu o opinie ekspertów, modelowanie dynamiki systemów w oparciu o technikę Jaya Forreстера stworzoną w latach 60. XX w., analizy analogii historycznych, analizy rynku, ryzyka i preferencji konsumenckich) (Daim et al., 2006, 981–985, 999). Zakłada się tutaj, że różne metody należy stosować do różnych etapów projektowania, wdrażania i komercjalizacji nowych technologii. Patenty odzwierciedlają etap rozwoju technologii, który następuje po etapie badań podstawowych i przed etapem zastosowań technologii w praktyce i uwidaczniania się jej wpływu na społeczeństwo. Przyrost liczby patentów w czasie można modelować funkcjami o kształcie zbliżonym do litery „S” (m.in. funkcją Gompertza, Fishera-Pry’ego, Voltera-Lotka, Pearla-Reeda) nazywanymi też funkcjami logistycznymi (zob. np.: Kucharavy & De Guio, 2011, 559–561; Price, 1967, 27–28), dzięki którym można przewidzieć ramy czasowe wczesnego etapu rozwoju technologii, wzmożenia prac nad jej dopracowaniem oraz osiągnięcia pewnego maksimum, które wyznacza moment „dojrzałości” technologii i przejścia do kolejnych etapów jej cyklu życiowego (zob. Rys. 2) (Daim et al., 2006, 990–991, 999–1000; zob. też: Lee et al., 2012, 817).



Rys. 2. Przykładowa krzywa logistyczna opisująca hipotetyczne tempo przyrostu liczby patentów w czasie; źródło: oprac. własne na podstawie Daim, T. et al., 2006, 986

Jak zaznaczyli Daim i in., problematycznym aspektem metody jest fakt, że analiza danych bibliometrycznych na temat liczby patentów jest w stanie dostarczyć jedynie danych historycznych, które mogą nie być jeszcze dostępne w przypadku najwcześniejszego etapu cyklu życiowego technologii nowo powstających (Daim et al., 2006, 1009).

Osmo Kuusi i Martin Meyer (2007) także poddali analizie patenty, które rozpatrywali w aspekcie wielu konkurujących ze sobą tzw. technologicznych paradygmatów funkcjonujących w obrębie tej samej specjalności. Paradygmaty te określono też jako „wizje”



(ang. *guiding image*; niem. *leitbild*), które stanowiły ogólne, podzielane przez całą społeczność wartości, oczekiwania, specjalistyczny język i cele badawcze, oferowały długoterminowe metody ich realizacji i dostarczały podstaw dla podobnie zorientowanych prac prowadzonych w jednym kierunku w obrębie odmiennych dyscyplin badawczych. Można je było również przyrównać do koncepcji samoorganizacji nauki funkcjonującej jako interpersonalny „stabilizator” wysiłków poszczególnych naukowców lub ogólniej – dynamiki postępu w nowo powstałej specjalizacji. Bibliometryczna analiza klastrów stosunkowo niedawno opublikowanych patentów (ich rosnącej w czasie liczby, powiązań bibliograficznych pomiędzy nimi oraz wykorzystywanego w nich słownictwa), była w tej perspektywie metodą identyfikacji różnych, obiecujących „wizji” na wczesnym etapie ich kształtowania się (Kuusi & Meyer, 2007, 760–763, 765–767, 774–775).

Inne, oryginalne podejście do bibliometrycznej analizy patentów zaprezentowali Byungun Yoon i Yongtae Park (2007). Podejście to polegało na badaniu terminologicznej morfologii patentów i wykrywaniu nieistniejących jeszcze, ale możliwych połączeń między słowami kluczowymi nadawanymi patentom na podstawie występowania w ich treści słownictwa charakteryzującego się najwyższą częstością (Yoon & Park, 2007, 592). Morfologię („mapę leksyki”) patentu można zademonstrować na przykładzie (zob. Tab. 1).

Tab. 1. Morfologia hipotetycznego zbioru patentów w postaci wyodrębnionych z nich, najczęściej pojawiających się w ich treści słów kluczowych; źródło: oprac. własne na podstawie Yoon & Park, 2007, 593

| Typ materiału     | Technika wytwarzania | Rodzaj więzi | Forma    |
|-------------------|----------------------|--------------|----------|
| Glina             | Ekstrakcja           | Ciepło       | Okrąg    |
| Metal             | Wytapianie           | Chemikalia   | Linia    |
| Tworzywo sztuczne | Prasowanie           | Magnetyzm    | Sześcián |

Morfologię pewnego patentu określa więc szereg terminów, które w powyższym przypadku można streścić jako „chemiczną ekstrakcję metalu do formy cząstek sześciennych”. To znaczy, że patent ten ma więcej słów kluczowych skojarzonych z „metalem” niż z „gliną” czy „tworzywem sztucznym” itd., ale w innych patentach konfiguracja może wyglądać inaczej. Konfiguracje „wolne”, tj. niezrealizowane dotąd w żadnym patencie mogą następnie zostać wyodrębnione i wykorzystane do wskazania prawdopodobnych kierunków dalszego rozwoju specjalności (Yoon & Park, 2007, 593). Uzupełnieniem tej metody może być analiza cytowań patentów o danej morfologii. Jako słabe strony swojej metodologii autorzy wskazali m.in. na ewentualność, że dotychczasowe słownictwo może nie oddawać w pełni kierunków rozwoju radykalnie nowych technologii. Ponadto problematyczne aspekty metody to np. możliwość, że niektóre niezrealizowane połączenia mogą okazać się fizycznie niemożliwe do uzyskania w praktyce, a także niezdolność metody do wykroczenia poza obszar, który obejmują analizowane dokumenty patentowe, jak również istnienie problemu związanego z opóźnieniem czasowym, które wywołuje konieczność oczekiwania na pojawienie się cytowań patentów. Odnotowano również potrzebę dopracowania i usystematyzowania sposobu pozyskiwania najodpowiedniejszych słów kluczowych dla konkretnych patentów (Yoon & Park, 2007, 597). Ssu-Han Chen i in. (2012), wykorzystali standardowe techniki analizy sieci cytowań



oraz algorytm grupujący publikacje patentowe wraz z analizą terminologii, celem śledzenia kierunków i trendów obserwowalnych w przesuwaniu się „frontu technologicznego” (Chen et al., 2012, 1100, 1103). Szczególny nacisk położono na transformowanie się struktur klastrów (ang. *community transition*) w kolejnych, zachodzących na siebie okresach obserwacji. Transformacje te miały postać różnego rodzaju „przejsć” (ang. *transitions*) z jednego okresu w drugi („narodziny”, „śmierć”, rozgałęzienie lub łączenie się klastrów), o których można było mówić kiedy dwa klastry w różnych okresach obserwacji zawierały wiele tych samych dokumentów patentowych. Zaznaczono ponadto potrzebę dostosowania długości okresu obserwacji (tj. szerokości przedziałów czasowych, na które dzieli się cały badany okres) do specyfiki danej specjalności. Zaproponowano w tym celu tzw. miarę TCT (ang. *technology cycle time*), która była różnicą pomiędzy datami wydania patentów cytujących i patentów cytowanych należących do pewnego klastra (Chen et al., 2012, 1102–1104). Część analizy stanowiła także hierarchizacja słów kluczowych opisujących klastry, tj. nadawanie im wag, które pokazywały jak dużą „siłę” posiadał dany termin w obrębie swojego klastra, jak dobrze odpowiadał on jego ogólnej tematyce i czy mógł on zostać przypisany także i innym klastram (Chen et al., 2012, 1199–1101, 1103).

Péter Érdi i in. (2013) wysunęli propozycję kolejnej, alternatywnej wobec pozostałych metody badania dokumentów patentowych pod kątem prognozowania nadchodzących trendów w rozwoju technologii. Rolę predyktora w metodzie autorów spełniał tzw. wektor cytowań patentów (ang. *citation vector*), który można było zdefiniować dla każdego patentu (Érdi et al., 2013, 226). Konstrukcja wektora cytowań opierała się na cytowaniach bezpośrednich. Dla każdego patentu wyznaczono sumę wszystkich pozyskanych cytowań<sup>6</sup>, które pochodziły od patentów należących do 36 różnych kategorii tematycznych ustalonych przez *US Patent & Trademark Office* (USP&TO). Cytowania „przychodzące” były dodatkowo wazone ze względu na to ile cytowań „rozesłał” patent cytujący (tj. jak obszerną posiadał bibliografię załącznikową). Waga była tym większa, im mniejsza była liczba takich cytowań „rozesłanych” (Érdi et al., 2013, 232). Wektor składał się zatem z 36 liczb pokazujących, jak wiele cytowań otrzymał dany patent od patentów należących do każdej kolejnej kategorii tematycznej USP&TO. Komponenty wektora cytowań patentu demonstrowały więc, jak duży wpływ wywarł on na inne (niż swoją własną) kategorie tematyczne, czyli jak wpływowo okazał się patent poza swoją własną kategorią (Érdi et al., 2013, 233–234). Obserwowano następnie jak relacje te zmieniały się w czasie, co pozwalało wytypować zbiory patentów podobnych pod wskazanym wyżej względem, których siła wpływu poza własną kategorią tematyczną wzrastała w czasie. Rozumienie wynalazczości prezentowane w ramach pracy Érdiego i in. polegało więc na jej postrzeganiu jako procesu rekombinacji technologii istniejących już wcześniej (Érdi et al., 2013, 238). Niedostatki tej metodologii to z kolei m.in. istnienie opóźnienia czasowego pomiędzy faktycznymi narodzinami nowej technologii a jej zaistnieniem w patentowych bazach danych i wzrostem jej cytawalności, fakt manualnego kategoryzowania patentów przez USP&TO, co narażało metodę na zarzut pewnej subiektywności oraz niejednorodność różnych obszarów technologii pod względem liczby ogłoszonych patentów, ich szans na pozyskanie cytowań oraz średniej liczby cytowań na jeden patent. Ostatnia z wymienionych cech miała wpływ na precyzję i zasadność ważenia cytowań zastosowanego przez autorów (Érdi et al., 2013, 238).

<sup>6</sup> Jeżeli patent nie otrzymał żadnych cytowań, jego wektor jest równy zeru.

Podobne zarzuty wobec analiz dokumentów patentowych w zakresie prognozowania rozwoju technologii stawiali też Koenraad Debackere i in. (2002). Wspomnieli oni m.in. o fakcie zróżnicowania różnych specjalności (oraz krajów), pod względem swojej podatności na generowanie nowych patentów, różnicach pomiędzy odmiennymi systemami (klasyfikacjami, przepisami) regulującymi przyznawanie patentów, co powoduje trudność w porównywaniu ich ze sobą, a także o pewnych problematycznych cechach tzw. zachowań naukowców w zakresie zgłaszania wynalazków (ang. *patenting behavior*) (Debackere et al., 2002, 216–217). Przez zachowania tego rodzaju rozumie się osobiste motywy i podejmowanie indywidualnych decyzji leżących u podstaw zgłaszania patentów do odpowiednich urzędów. Przykładowo, nie wszystkie patenty są stosowane w praktyce i nie wszystkie prowadzą do rzeczywistych innowacji. Nie wszystkie wynalazki dają się opatentować i nie wszystkie wynalazki, które dają się opatentować są faktycznie chronione prawem patentowym. Istnieją też różne powody, dla których nie wszystkie przedsiębiorstwa decydują się na patentowanie swoich odkryć. Mogą to być np. powody takie, jak zbyt duża trudność w wykazaniu nowatorskiego charakteru wynalazku, niechęć do ujawniania odkrycia, wysokie koszty wdrożenia wynalazku, nieefektywność prawa patentowego w zakresie szybkości i kosztów związanych z imitowaniem pewnych wynalazków. Niektóre wynalazki są ponadto patentowane „seryjnie”, w postaci więcej niż jednego dokumentu patentowego (Debackere et al., 2002, 217). Ogólnie mówiąc, zachowania badaczy w zakresie patentowania swoich wynalazków są silnie uzależnione od sektora przemysłu oraz od typu wynalazku. Istnieją również problemy związane z samymi systemami patentowymi (np. brak retrospektywnego korygowania wykorzystywanych kategorii tematycznych), problemy związane z ogromną różnorodnością wartości technologicznej i ekonomicznej wynalazków, a także z procesem cytowania patentów (np. brak ścisłych danych nt. odnawiania patentów w celu utrzymania w mocy prawa wyłącznego) (Debackere et al., 2002, 219–222).

## 5. Analizy głównych ścieżek rozwoju dyscypliny

Jest to kolejna z metod służących zademonstrowaniu sposobu, w jaki przebiegał historyczny rozwój danej dyscypliny oraz wskazaniu kierunku, jaki może on przybrać w przyszłości. Posłużyli się nią np. John Liu i Louis Lu (2012), identyfikując główne kanały przepływu informacji w obrębie literatury dotyczącej indeksu Hirsha, za pomocą metody opartej na pomysłach Hummona i Doreiana (1989) (zob. podrozdział trzeciej części pierwszej). Metoda ta została niejako rozszerzona przez Liu i Lu, przede wszystkim dzięki nadaniu jej wymiaru globalnego (w przeciwieństwie do oryginalnej metody, którą charakteryzował wymiar lokalny), dwukierunkowego (tj. analizowaniu ścieżek prowadzących nie tylko od dokumentów cytowanych do cytujących, ale też ścieżek wiodących od dokumentów cytujących do cytowanych) i dopuszczeniu możliwości wyodrębnienia więcej niż jednej „głównej ścieżki” przepływu informacji w danej dyscyplinie lub subdyscyplinie (Liu & Lu, 2012, 531–534). Ocena przyszłych perspektyw dyscypliny sprowadzała się w tym ujęciu do porównania podobieństwa wyodrębnionych ścieżek. Wysokie podobieństwo dwóch (lub większej liczby) ścieżek sygnalizowałoby koncentrację uwagi badaczy na publikacjach, które wytyczyły te ścieżki, podczas gdy brak podobieństwa sugerowałby, że rozwój dyscypliny przebiega wielokierunkowo, w sposób rozproszony, oznaczający brak jednolitej

trajektorii rozwojowej (Liu & Lu, 2012, 532, 537). Spośród wad proponowanego podejścia można wymienić np. wysoką zależność wyników od wybranych bibliograficznych źródeł informacji nt. publikacji z danego zakresu, brak uwzględnienia zróżnicowania stopnia relewancji tematycznej pomiędzy pracami cytowanymi a cytującymi (tj. milczące przyjęcie normatywnego założenia, że wszystkie cytowania są sobie równe), okazjonalne przypadki wyłączenia poza „główną ścieżkę” prac wyjątkowo wysoko cytowanych w danej dyscyplinie oraz faworyzowanie artykułów przeglądowych o pokaźnych bibliografiach załącznikowych i często wysokiej liczbie cytowań, które jednak nie przyczyniają się w bezpośredni sposób do przesuwania granic wiedzy w dyscyplinie (Liu & Lu, 2012, 538–540).

Praca Liu i Lu zainspirowała stworzenie kolejnej modyfikacji metody Hummona i Doreiana, w postaci tzw. map trajektorii konceptualnych (ang. *conceptual trajectory maps*), którą zademonstrowali Yi-Ning Tu i Shu-Lan Hsu (2016). Modyfikacja ta polegała przede wszystkim na uzupełnieniu „głównej ścieżki” publikacji o prace, które wprawdzie początkowo znalazły się poza tą ścieżką, ale wykazywały wystarczający stopień podobieństwa treściowego do prac leżących na ścieżce. Jako element podwyższający stopień podobieństwa dwóch różnych publikacji uwzględniono też daty ich wydania. Podobieństwo treściowe oraz podobieństwo pod względem czasu powstania publikacji<sup>7</sup> oceniono za pomocą wskaźnika podobieństwa kosinusowego (zob. podrozdział czwarty części pierwszej). Następnie uzyskanym tym sposobem grupom artykułów wytyczających przebieg ścieżki przypisano określone terminy zaczerpnięte z tekstów tych artykułów, które stanowiły reprezentację przedstawianych przez nie koncepcji (Tu & Hsu, 2016, 2020–2023). Zdaniem Tu i Hsu zbudowane w streszczony powyżej sposób mapy danych dyscyplin lub subdyscyplin mogą zostać wykorzystane w celu identyfikacji tematyki wartościowej i „trwałej” w danym obszarze oraz tematyki „ulotnej” (ang. *transient*), jak również w celu śledzenia narodzin i rozwoju zagadnień nowo powstających oraz w celu badania całokształtu cyklu życiowego danej problematyki w jej otoczeniu informacyjnym (Tu & Hsu, 2016, 2024, 2030).

Metoda Hummona i Doreiana była ponadto punktem wyjścia do analizy aktualnych frontów badawczych w obszarze z pogranicza prawa i finansowości, zrealizowanej przez Vincenta Ma i Johna Liu (2016). „Główne ścieżki” wybranego obszaru problemowego zostały wyodrębnione w ramach utworzonych uprzednio klastrów, przy konstrukcji których posłużono się algorytmem grupującym Newmana z 2004 r. (zob. podrozdział drugi). Rozszerzoną wersję metody identyfikacji ścieżek zaczerpnięto z kolei od Liu i Lu (2012). W dalszej kolejności uzupełniono ją o wybór słów kluczowych, które zostały wykorzystane jako deskryptory frontów badawczych (Ma & Liu, 2016, 36–37, 41, 44). Ma i Liu przyznali zarazem, że wyniki analizy „głównych ścieżek” są dynamiczne w tym sensie, że mogłyby zmienić się, gdyby do rozpatrywanego, wyjściowego zbioru publikacji dołączono publikacje najnowsze (nie dostępne jeszcze w momencie realizacji badania). Dlatego swoje wyniki autorzy określili jako „migawkę” (ang. *snapshot*), która niekoniecznie musiała odzwierciedlać trendy stabilne w czasie i obowiązujące w dalszym ciągu w zbadanym obszarze nauki (Ma & Liu, 2016, 49).

<sup>7</sup> W tym zakresie rozpatrywano tylko prace, które powstały w odstępie nie większym i nie mniejszym niż rok od opublikowania pracy, do której te pierwsze były przyrównywane. Np. jeżeli praca leżąca na ścieżce została wydana w 2006 r., wyznaczano jej podobieństwo do prac wydanych w przedziale 2005–2007 (Tu & Hsu, 2016, 2022).

Wyniki badania, które w wyższym stopniu akcentowało dynamikę dyscyplin naukowych, przedstawił Erjia Yan i Qi Yu (2016). Wykorzystana przez nich metoda nie była tą samą, która stanowiła podstawę trzech analiz streszczonych powyżej chociaż wydaje się, że można zaliczyć ją do metod pokrewnych. Metodą tą było tzw. maksymalne drzewo rozpinające (MST – ang. *maximum spanning tree*), a jej zastosowanie do sieci cytowań pozwoliło autorom na zlokalizowanie dyscyplinarnych ścieżek dyfuzji wiedzy, co było też czołowym założeniem opracowań posługujących się metodą identyfikacji „głównych ścieżek” (Yan & Yu, 2016, 1944). Danymi empirycznymi były w tym przypadku cytowania pomiędzy czasopismami (ang. *journal-to-journal citations*) oraz kategorie tematyczne czasopism zapożyczone z bazy *Scopus* (Yan & Yu, 2016, 1944–1945). Dynamiczny aspekt analizy z użyciem MST opierał się na wprowadzeniu pięciu dwuletnich okresów obserwacji cytowań. Dokładniej mówiąc, dane wyjściowe obejmowały cytowania odsyłające do czasopism wydanych dwa lata wcześniej, a więc cytowania, które pojawiły się w 1999 r. i odsyłały do prac z 1997 r., cytowania z 2002 r. odsyłające do prac z 2000 r., i cytowania z lat 2005, 2008 i 2011 odsyłające odpowiednio do prac z lat 2003, 2006 i 2009 (Yan & Yu, 2016, 1944, 1952). Sama metoda MST polega natomiast na wyodrębnieniu z całej sieci pewnej podsięci połączeń stanowiących niejako jej esencję, tj. najmniejszą z możliwych podsięci połączeń o najwyższej wadze (mierzonej odpowiednio znormalizowaną liczbą cytowań pomiędzy czasopismami), która nie zawiera połączeń cyklicznych<sup>8</sup> i jednocześnie pozwala utrzymać przepływ informacji. Usunięcie jakiegokolwiek krawędzi z takiej podsięci skutkowałoby zaburzeniem tego przepływu. Podsięć MST ma dokładnie tyle samo wierzchołków, co sieć wyjściowa i liczbę połączeń równą, co najwyżej liczbie wierzchołków minus jeden, co wskazuje, że tworzenie MST jest swego rodzaju sposobem na „oczyszczanie” sieci cytowań i obnażenie kognitywnego „kręgosłupa” (ang. *backbone*) dyscypliny (Yan & Yu, 2016, 1946–1947, 1950). Połączenie MST z kilkoma okresami obserwacji cytowań stworzyło możliwość monitorowania zmian w strukturze sieci cytowań zachodzących w niej z biegiem czasu. Konkretnie mowa tu o monitorowaniu siły wpływu, jaką wywierają na siebie wzajemnie różne dyscypliny nauki. O wpływie tym można było mówić w kategoriach „eksportu” i „importu” informacji oraz potencjału informacyjnego, który dyscypliny mają do zaferowania i z którego wyrastają nowe odgałęzienia dyscyplin (ang. *child branches*). Odgałęzienia te mogą być nie tylko specjalizacjami nowo powstającymi, ale mogą też być szerokimi strukturami dyscyplinarnymi, które z czasem zmieniają swoją pozycję na mapie nauki. Przykładowo, w latach 2000/2002 materiałoznawstwo było gałęzią fizyki („importerem wiedzy”), a w latach 2011/2009 było już „eksporterem” informacji przepływającej do dyscyplin fizycznych (Yan & Yu, 2016, 1950, 1952). Sekwencja map dawała ponadto sposobność odnotowania wielkoskalowych tendencji w ustalaniu się zależności pomiędzy głównymi obszarami mapy. Zaobserwowano np., że pomiędzy rokiem 1997 a 2011 dyscypliny takie jak biznes, materiałoznawstwo, informatyka i dyscypliny inżynierskie zajmują coraz bardziej dominujące pozycje i stają się głównymi „eksporterami” informacji (Yan & Yu, 2016, 1954). Ograniczenia zastosowanej metodologii wynikały przede wszystkim z charakterystyk zbioru danych empirycznych. Były nimi po pierwsze, sztuczne – w ocenie

<sup>8</sup> Połączeniem cyklicznym byłoby takie odwołanie, które zaczynałoby się i kończyło w tym samym wierzchołku. Miałoby ono miejsce gdyby, np. czasopismo A cytowało czasopismo B, które z kolei cytowałoby czasopismo C cytujące czasopismo A (A→B→C→A).

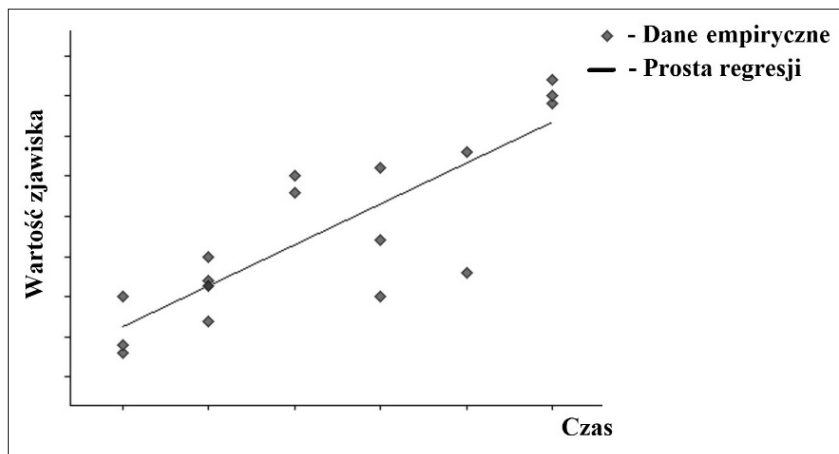
autorów – kategorie tematyczne *All Science Journal Classification*, a po drugie przyjęta jednostka analizy, którą były czasopisma i dyscypliny naukowe, a nie indywidualne publikacje. Przyjęcie jednostki usytuowanej na niższym poziomie agregacji danych umożliwiłoby, zdaniem autorów, precyzyjniejszą interpretację otrzymanych wyników. Analiza Yana i Yu była ponadto pozbawiona możliwości uwzględnienia oddziałujących na naukę czynników socjotechnicznych (np. polityki finansowania badań czy ewolucji sieci współpracy naukowej) (Yan & Yu, 2016, 1950, 1953).

## 6. Analiza szeregów czasowych

Kolejną grupą publikacji jest ta, w której autorzy posługiwali się metodą bliższą naukom społecznym i ekonomicznym (np. analizami socjologicznymi, ekonomicznymi lub finansowymi), niż naukom ścisłym (informatyce i matematyce), w przeciwieństwie do wielu streszczonych powyżej algorytmicznych metod tworzenia klastrów publikacji wzajemnie powiązanych. Ścisłej mówiąc, pod nazwą „analiza szeregów czasowych” kryje się cały zespół różnych metod statystycznej analizy dynamiki zjawisk masowych, których wspólnym celem jest przewidywanie przyszłego zachowania się zjawisk zachodzących w czasie za pomocą identyfikacji głównych trendów rozwojowych zjawiska, jego sezonowych lub okresowych odchyień od trendu i oceny niepewności towarzyszącej prognozie. W tym samym zakresie stosuje się ponadto techniki analityczne tzw. dekompozycji szeregu czasowego, które polegają na sformułowaniu równania opisującego ilościowo przebieg zjawiska z rozróżnieniem trendu głównego, składowej wahań cyklicznych lub sezonowych oraz składowej odpowiadającej wahaniom przypadkowym (zob. np.: Aczel, 2007, 621–667; Sobczyk, 2015, 307–356; Sobczyk, 2008, 35–87). O niektórych z tych metod wspomniano już w podrozdziale 3 przy okazji omawiania pracy Choudhury’ego i Uddina (2016), w której metody tego rodzaju pełniły funkcję uzupełniającą wobec metody analizy struktur sieciowych typu *co-word*. Podobnie jak w tym przypadku, także i w publikacji Yuena-Hsiena Tsenga i in. (2009) analiza trendu wyznaczającego przypuszczalny kierunek przyszłego rozwoju nowo powstałych specjalizacji, została poprzedzona identyfikacją i wizualizacją struktury dyscypliny w postaci mapy tego samego typu (*co-word*). Identyfikacja ta odbywała się przy użyciu standardowych technik grupowania wyjściowego zbioru publikacji w klastry (Tseng et al., 2009, 75, 80). Każdy klaster został następnie podzielony na podzbiory publikacji o tym samym wieku. To znaczy, że każdy klaster został przedstawiony jako szereg wartości odpowiadających liczbom tworzących go publikacji o wieku jednego roku, dwóch lat itd. Na przykład, ciąg postaci {25, 38, 47, 40...} oznaczałby, że w pewnym klastrze znalazło się 25 publikacji wydanych nie wcześniej niż rok przed momentem badania, 38 publikacji wydanych nie wcześniej niż dwa lata przed chwilą badania itd. W ten sposób powstały serie czasowe, które poddano dalszej analizie statystycznej (Tseng et al., 2009, 75). W ramach tej analizy porównano ze sobą sześć różnych wskaźników statystycznych stanowiących predyktory trendu rozwojowego (ang. *trend prediction index*) (Tseng et al., 2009, 77). Pierwszym z indeksów było średnie tempo przyrostu procentowego (ang. API – *average percentage of increase*), które informuje o tym, o ile wyższy lub niższy był średnio poziom badanego zjawiska w danym okresie, w stosunku do okresu bezpośrednio go poprzedzającego. Drugim indeksem był tzw. współczynnik kierunkowy prostej regresji (ang. SLP – *slope of linear regression*). Prosta



regresji to linia prosta, która stanowi najlepsze z możliwych dopasowanie do danych empirycznych, a kąt jej wznoszenia się lub opadania wraz ze wzrostem wartości na osi czasu określa, jak szybkie jest tempo wzrostu lub spadku wartości jakiejś cechy danego zjawiska (w tym przypadku – jak szybkie jest tempo wzrostu lub spadku liczby publikacji wraz ze wzrostem ich wieku w obrębie klastra). Współczynnik kierunkowy prostej jest z kolei czynnikiem determinowanym ściśle przez wspomniany kąt, pod którym prosta opada (lub wznosi się)<sup>9</sup>. Na rysunku 3 przedstawiono ilustrację dopasowania wznoszącej się prostej do hipotetycznych danych doświadczalnych.



Rys. 3. Prosta regresji dopasowana do hipotetycznego zbioru danych empirycznych

Pozostałe cztery indeksy stanowiły warianty dwóch pierwszych (Tseng et al., 2009, 77). Po zakończeniu etapu realizacji badania empirycznego wybranego zbioru prac, Tseng i in. zarekomendowali metodę SLP jako dającą najwyższą zgodność z zewnętrznymi źródłami informacji nt. problematyki zyskującej na znaczeniu w środowisku naukowym. Metoda ta wykazała się najwyższą efektywnością w zróżnicowanych warunkach badawczych, tj. warunkach uwzględniających różne przedziały czasowe analizy, różne okresy prognozowania, różne dyscypliny i różne, pod względem wielkości, zbiory analizowanych dokumentów (Tseng et al., 2009, 88–89).

Günter Krampen i in. (2011) wykorzystali alternatywną, wobec wskazanych powyżej, metodę prognozowania przyszłego przyrostu liczby publikacji w dyscyplinie psychologii. Metodą tą było tzw. wygładzanie wykładnicze (ang. *exponential smoothing*), w ramach którego ekstrapoluje się w przyszłość trend zaobserwowany w momencie badania i opisujący dane rozciągające się w przeszłość. Obserwacjom zjawiska zachodzącego w latach ubiegłych (a konkretnie w latach 1977–2008) przypisano w ramach tej metody wagi, które są większe dla obserwacji z lat bliższych współczesności, niż dla obserwacji starszych, przy czym waga obserwacji maleje wykładniczo wraz ze wzrostem dystansu czasowego. Inaczej mówiąc, stosując wygładzenie wykładnicze, obserwacje najnowsze wywierają najsilniejszy wpływ

<sup>9</sup> Więcej szczegółów na temat prostej regresji można znaleźć np. w podręczniku Johna Taylora (2012, 216–229), Barbary Łapkowskiej-Baster (2009, 19–31) lub Nicholasa Binghama i Johna Fry'ego (2010).



na wartości prognozowane. W związku z tym, jeśli trend (główne przyczyny determinujące zjawisko) zmieniłby się w jakimś punkcie w przeszłości, wartości zjawiska sprzed tej zmiany będą wywierać mniejszy wpływ na prognozę, niż wartości zjawiska zaobserwowane po zajściu zmiany (Krampen et al., 2011, 688–689; zob. też: Choudhury & Uddin, 2016, 751–752). Waga obserwacji coraz starszych może przy tym spadać szybko lub wolno, a jej konkretne wartości określa się przyjmując współczynnik (parametr wagowy) występujący w wyrażeniu przez które mnoży się wartość każdej obserwacji (zob. np.: Aczel, 2007, 642; Sobczyk, 2008, 121). Za problematyczny aspekt tej metody można przypuszczalnie uznać fakt, że użyte przez Krampena i in. proste wygładzanie wykładnicze<sup>10</sup> stosuje się na ogół do takich szeregów czasowych, które nie mają wyraźnego trendu ani wahań sezonowych (Aczel, 2007, 641, 647; Sobczyk, 2008, 122, 138; Snarska, 2011, 218), a więc są szeregami w dużym stopniu nieregularnymi co powoduje, że prognoza może być obciążona znacznym błędem. W literaturze przedmiotu można też napotkać stwierdzenie, według którego ocena dynamiki dyscyplin w oparciu o same jedynie trendy publikacyjne (tj. liczbę publikacji z danego zakresu w różnych latach) jest niewłaściwa z uwagi na to, że różne typy wydawnicze publikacji przyczyniają się do rozwoju dyscyplin w niejednakowy sposób lub ze względu na to, że całkowita liczba publikacji w dyscyplinach rośnie na ogół wykładniczo, co może przesłaniać węższe i bardziej wyspecjalizowane trendy (Chi et al., 2006, 69; Sitarz et al., 2012, 4; Tseng et al., 2009, 75).

Robert Sitarz i in. (2012) zaproponowali własną metodę analizy krótkoterminowych trendów w nauce, w oparciu o słowa kluczowe i klastry zbudowane przy użyciu techniki współwystępowania słów kluczowych. Klastry, traktowane jako odpowiedniki subdyscyplin lub obszarów problemowych, stają się następnie podstawą budowy szeregów czasowych, w których kolejne wartości odpowiadają liczbie dokumentów wydanych w każdym kolejnym roku analizy i zawartych wewnątrz klastra (Sitarz et al., 2012, 7–10). Metoda identyfikacji trendu była w tym ujęciu złożeniem trzech prostszych metod: trendu liniowego (ang. *trend line method*), metody oscylacyjnej (ang. *oscillator method*) i metody oscylatora stochastycznego (ang. *stochastic oscillator method*) (Sitarz et al., 2012, 11–14). Metoda trendu liniowego polega na przeprowadzeniu linii prostej łączącej minima (lub maksima) na wykresie wartości obserwacji tworzących szereg czasowy. Dopuszczalna jest tu zarazem więcej niż jedna prosta opisująca dane doświadczalne. Trend liniowy może więc składać się z np. dwóch prostych w układzie współrzędnych, które przecinają się w jakimś punkcie (Sitarz et al., 2012, 11–12). Metoda oscylacyjna polega na oszacowaniu różnicy pomiędzy liczbą publikacji w danym roku „t”, a liczbą publikacji w „x” poprzednich latach. Wylicza się tę różnicę dla wielu kolejnych lat i wynik przedstawia na jednym wykresie (Sitarz et al., 2012, 12–13). Oscylator stochastyczny jest z kolei techniką analityczną, która umożliwia przewidzenie nasilenia się trendu po pewnym momencie, jego spadek lub zmianę kierunku trendu (zob.: Sitarz et al., 2012, 14; zob. też: Krzywda, 2010, 111–113). Wykryty trend uznaje się za istotny kiedy obie metody oscylatorowe oraz metoda liniowa wysyłają ten sam sygnał, tj. gdy ich wskazania pokrywają się (Sitarz et al., 2012, 20). Jako ostateczny

<sup>10</sup> Do prognozowania wartości szeregów czasowych cechujących się wyraźniejszym trendem lub wahaniami sezonowymi stosuje się bardziej wyrafinowane metody tzw. wykładniczego wygładzania adaptacyjnego, np. metodę wygładzania Holta, metodę wygładzania Browna, metodę wygładzania Wintersa lub metodę wyznaczania tzw. trendu pelzającego (zob. np.: Aczel, 2007, 645, 647; Cipra & Hanzák, 2008, 385; Krawiec, 2014; Malina, 1994; Snarska, 2011, 212–244; Sobczyk, 2008, 97–152).

rezultat uzyskuje się fragmentację całego cyklu życiowego pewnego zbioru publikacji na krótsze podokresy, które mogą charakteryzować się własnym trendem (wznoszącym lub opadającym w różnych przedziałach czasu). Autorzy wskazali na możliwość dalszego dopracowania metody poprzez np. zwiększenie jej wrażliwości w aspekcie, w którym jest ona zależna od doboru właściwych słów kluczowych. Zaznaczono też potrzebę dokładniejszej determinacji wartości progowych w trakcie wyodrębniania klastrów (jak wiele współwystąpień słów kluczowych musi zaistnieć, aby przypisać pewną pracę do konkretnego klastra?) oraz potrzebę automatyzacji, tj. zapewnienia metodzie zdolności do analizy dużych zbiorów danych (Sitarz et al., 2012, 27–28).

## 7. Pozostałe publikacje

W tej sekcji zebrano i skrótkowo scharakteryzowano relewantne do celu i przedmiotu niniejszej pracy publikacje, które były na tyle oryginalnymi propozycjami, że nie zmieściły się w obrębie żadnej z wymienionych wcześniej podgrup.

Steven Morris (2005) zaproponował tzw. matematyczny model CAPE (ang. *cumulative advantage by paper with exemplars*), który wyjaśniał i odtwarzał m.in. proces wyłaniania się wysoko cytowanych „publikacji przykładowych/wzorcowych” (ang. *exemplar references*)<sup>11</sup> natychmiast po ukształtowaniu się nowo powstałej specjalizacji badawczej (Morris, 2005, 1251, 1266, 1272). Konceptualizacja narodzin nowych specjalności Morrisa była zgodna z teorią postępu w nauce Thomasa Kuhna (2001), tj. uznano w jej ramach, że długie okresy tzw. nauki normalnej są przerywane krótkimi okresami rewolucji naukowych (tzw. przewroty kopernikańskie). Ich konsekwencją jest zmiana obowiązującego w nauce paradygmatu i dokonywanie nowych odkryć w ramach nowego paradygmatu, które doprowadzają do powstania nowych specjalności (Morris, 2005, 1251). Model Morrisa był w stanie zrekonstruować etapy narodzin nowych specjalności takie, jak ogłoszenie odkrycia w formie publikacji, proces narastania cytawalności takich publikacji, pojawienie się prac testujących odkrycie empirycznie i potwierdzających jego słuszność, rozprzestrzenianie się wiedzy o odkryciu poza granicami jego macierzystej dyscypliny naukowej oraz wejście w fazę konsolidacji, którą wyznaczają powstające publikacje przeglądowe i monograficzne standaryzujące, kodyfikujące i popularyzujące odkrycie. Rekonstrukcja obejmowała również dynamikę nawarstwiania się cytowań i generowania powiązań bibliograficznych i współcytowań pomiędzy publikacjami (Morris, 2005, 1251–1252, 1272). Do wad swojego modelu Morris zaliczył natomiast jego niezdolność do odwzorowania relacji współcytowania publikacji pochodzących z różnych specjalności dyscyplinarnych, jak również niezdolność do odzwierciedlenia specjalnego przypadku wykształcania się nowych specjalności, jakim jest rozszczepianie się szerszych specjalności na mniejsze „subspecjalności”. Niedostatkami modelu było ponadto przyjęcie założenia, że tempo pojawiania się publikacji wzorcowych w dyscyplinie maleje wykładniczo wraz z upływem czasu, podczas gdy w rzeczywistości

<sup>11</sup> „Publikacje wzorcowe” to według Morrisa prace, w których zaprezentowano rozwiązane uprzednio problemy naukowe i które dostarczają tym samym fundamentalnych definicji problemów, które nie doczekały się jeszcze rozwiązania oraz metod, za pomocą których należy problemy te badać i rozwiązywać. Ogólnie mówiąc publikacje te wyznaczają standardy „naukowości” i standardy technik walidacji problemów istniejących w ramach aktualnie obowiązującego paradygmatu (Morris, 2005, 1251, 1253).

tempo to jest najprawdopodobniej wysoce zależne od konkretnej sytuacji, dyscypliny i specyfiki zachowań naukowców w zakresie cytowań literatury źródłowej (ang. *case-specific*). Do mankamentów swojego podejścia Morris zaliczył ponadto przyjęcie ustalonej (niezmiennej, utrzymującej się na stałym poziomie) „siły przyciągania” cytowań (ang. *attractiveness*) przez publikacje wzorcowe, w postaci stałego parametru wprowadzonego w matematyczne sformułowanie modelu (Morris, 2005, 1273).

Hanning Guo i in. (2011) wysunęli propozycję „wieloskładnikowej” metody identyfikacji nowo powstających specjalności naukowych (ang. *mixed-indicators model*). Ich koncepcja polegała na połączeniu trzech wskaźników cząstkowych, które wspólnie sygnalizowałyby wyłonienie się nowego obszaru problemowego. Wskaźnikami tymi były: (1) gwałtowny wzrost częstości użycia określonych terminów w dyscyplinie (ang. *word burst*), który poprzedza wejście takich terminów do powszechnego użytku, (2) pojawienie się nowych autorów w obrębie nowej specjalności, których przyciągnął potencjał nowego obszaru badań oraz (3) cytowanie pozycji pochodzących z różnorodnych dyscyplin (wysoce międzydyscyplinarne pozycje bibliograficzne cytowane w nowej subdyscyplinie). U podstaw ostatniego z trzech wymienionych wskaźników tkwiło założenie, że nowe obszary wyrastają z obszarów już istniejących i oddzielają się od nich czerpiąc jednocześnie wiedzę z innych dyscyplin. Drugim powodem wzrostu interdyscyplinarności przypisów był tutaj brak publikacji specyficznych wyłącznie dla takiej nowo powstałej specjalności na wczesnym etapie jej rozwoju (Guo et al., 2011, 423, 429, 431–432).

Ling-Chu Lee i in. (2012) realizowali cele predykcyjne za pomocą trzech koncepcji metodologicznych. Pierwszą z nich było zastosowanie tzw. metody Delphi z zamiarem zebrania opinii ekspertów i zidentyfikowania kluczowych zagadnień w badanym obszarze nauki. Drugą była analiza bibliometryczna, poprzez którą stworzono opis trendu wzrostowego w liczbie publikacji odpowiadających danej tematyce, co miało za zadanie ustalenie aktualnego etapu rozwoju, na którym znajdowała się dana grupa zagadnień. Trzecia koncepcja polegała na analizie literatury przedmiotu za pomocą tzw. indeksu RCA (ang. *revealed cumulative advantage*). Indeks ten pomaga wyodrębnić obszary tematyczne, które znajdują się aktualnie na porównywalnym etapie rozwoju na świecie i w jakimś konkretnym kraju (którym – w przypadku badania Lee i in. – był Tajwan) (Lee et al., 2012, 817, 820, 827; zob. też: Verbeek et al., 2002, 202). Indeks ten wskazuje, jaki jest udział kraju w dorobku światowym w pewnej dyscyplinie, w porównaniu z udziałem tego kraju w ogólnoświatowej produkcji piśmienniczej. Inaczej mówiąc, wskaźnik demonstruje, czy kraj angażuje się w danej dyscyplinie w naukę światową i w jakim stopniu, w stosunku do stopnia jego zaangażowania i aktywności w innych dyscyplinach. Wartość indeksu większa od „1” świadczy o tym, że kraj specjalizuje się w dyscyplinie (udział publikacji z niego się wywodzących jest duży w ramach dorobku ogólnoświatowego). Wartość indeksu mniejsza od „1” wskazuje, że pozycja danego kraju w dyscyplinie jest mało znacząca. Wartość równa „1” oznacza natomiast pozycję neutralną (Lee et al., 2012, 818–819, 821; zob. też: Verbeek et al., 2002, 202). Metodę prognozowania i zarazem podstawę sformułowania strategii rozwojowej kraju stanowiło ostatecznie połączenie wartości wskaźnika RCA z etapami rozwoju danej problematyki w postaci tabeli zawierającej wszystkie możliwe kombinacje wysokich i niskich wartości wskaźnika RCA z etapami rozwoju takimi jak etap „embrionalny”, nowo powstały, rosnący oraz dojrzały (ang. *embryonic, emergent, growing, mature*) (Lee et al., 2012, 821). Za kluczową część wytyczającą plany rozwoju kraju uznano te obszary nauki,

które znalazły się w fazie embrionalnej o wysokim wskaźniku RCA. Znamionowały one te kwestie, które znajdują się w bardzo wczesnej fazie rozwoju na świecie i jednocześnie, w których Tajwan posiada silną pozycję (ang. *comparative advantage*) w porównaniu z innymi krajami (Lee et al., 2012, 822). Do ograniczeń metody zaliczono fakt, że wskaźnik RCA ignoruje czynniki inne niż sama tylko liczba publikacji na dany temat, np. jej siłę wpływu lub jakość naukową. Metoda autorów nie uwzględniała ponadto czynników takich, jak np. działania rządu w zakresie subsydiowania różnych gałęzi przemysłu i rolnictwa, kwestii podatkowych lub ryzyka wynikającego z kwestii dokładności szacowania RCA (Lee et al., 2012, 828).

Xianwen Wang i in. (2013) wystąpili z oryginalnym pomysłem polegającym na badaniu najnowszych trendów w nauce poprzez skupienie uwagi na środowisku publikacji elektronicznych. Według Wang'a i in. miałyby to być sposób na zaradzenie trudnościom, jakie napotyka tradycyjna analiza cytowań stosowana w perspektywie prognostycznej. W tym kontekście wspomniano m.in. o fakcie istnienia opóźnienia w publikowaniu artykułów naukowych i opóźnienia związanego z pojawianiem się kolejnych cytowań albo o problemie z zakresu socjologii nauki, którym są konstruktywistyczne zarzuty stawiane przed paradygmatem normatywnym<sup>12</sup> (Wang et al., 2013, 718–719). Z uwagi na tego rodzaju problemy, autorzy podjęli próbę podsumowania aktualnych tendencji i najnowszych osiągnięć (ang. *hot topics*) w naukometrii w oparciu o dane na temat pobrań elektronicznych wersji artykułów zamieszczonych na platformie *SpringerLink*. Metodą śledzenia nowo powstających obszarów badawczych było w ujęciu autorów zidentyfikowanie takich przypadków pobrań publikacji elektronicznych, w których publikacje te opatrzone były słowami kluczowymi cechującymi się nowością w ostatnich latach (nieczęstym występowaniem), i w których – po drugie – liczba pobrań artykułów z nowymi słowami kluczowymi była wyższa niż 50. Po trzecie, przypadki te obejmowały sytuacje, w których stosunek pobrań prac z danym słowem kluczowym do liczby publikacji opisywanych tym słowem był wyższy niż 20 (tj. akcentowano tutaj wysoką liczbę pobrań i zarazem niewielką liczbę publikacji, którym przypisano dane słowa kluczowe) (Wang et al., 2013, 725, 727). Jako mankament metody wskazano możliwość, że pewna liczba publikacji mogła być pobierana z powodów innych niż czysto naukowe, w związku z czym nie można było całkowicie wykluczyć czynników arbitralności i pewnej przypadkowości obecnych w przynoszonych przez metodę wynikach (Wang et al., 2013, 728).

Xiaoling Sun i in. (2016) podzielali opinię Uphama i Smalla (2010), zgodnie z którą interdyscyplinarność badań naukowych jest wyznacznikiem zdolności do wzbudzania i przyspieszania innowacyjnych kierunków rozwoju nauki. Z drugiej strony, Sun i in. zgadzali się z tezą, według której konferencje naukowe są źródłem danych szczególnie podatnych na monitorowanie dyscyplin naukowych pod kątem kierunku i tempa ich rozwoju. W związku z tym obrali oni za przedmiot badania sieć współpracy między autorami, którzy w latach 1980–2010 brali udział w cyklicznych konferencjach naukowych poświęconych informatyce

<sup>12</sup> Chodzi tutaj o kwestionowanie normatywnej tezy, według której cytowania publikacji świadczą bezpośrednio o sile jej wpływu i naukowej wartości lub o podnoszone często kwestie związane ze specyfiką zachowań indywidualnych naukowców w zakresie cytowań (ang. *citation behaviour*), ich złożonymi motywacjami i procesami decyzyjnymi kryjącymi się za każdym aktem zamieszczenia przypisu itd. Szczegółowy przegląd tych zagadnień przedstawili np. Lutz Bornmann i Hans-Dieter Daniel (2008) oraz Mengxiong Liu (1993).

wraz z jej wielod dziedzinowymi zastosowaniami. Dane bibliograficzne takich konferencji indeksowane są m.in. w bazie *Digital Bibliography & Library Project* (DBLP) (Sun et al., 2016, 751–752). Przyjmując powyższe założenia utworzono złożoną z klastrów sieć (ang. *network of communities*), która obejmowała wszystkie poszczególne lata z przedziału 1980–2010. Konferencje spełniały tu funkcję wierzchołków sieci, a rolę połączeń (krawędzi) między wierzchołkami odgrywała relacja podobieństwa zdefiniowana jako współwystępowanie autorów referatów, które ogłoszono podczas różnych konferencji i opublikowano w materiałach pokonferencyjnych (Sun et al., 2016, 751–752). Algorytm grupujący wyjściowy zbiór konferencji na mniejsze podzbiory stanowił wariant algorytmu Newmana (2004) i został przez autorów zapożyczony z pracy Vincenta Blondela i in. (2008). Ponieważ autorzy referatów publikowali na ogół w różnych specjalnościach informatyki, a z czasem angażowali się w coraz większą liczbę specjalności i obszarów problemowych<sup>13</sup>, skonstruowana sieć odzwierciedlała ewolucję informatyki, rozprzestrzenianie się specjalności i przepływ informacji pomiędzy danymi specjalnościami, wyłanianie się nowych obszarów i zanikanie tych starszych, czy dokonywanie się przełomów i wzrost popularności pewnych nowo postawionych zagadnień (Sun et al., 2016, 752–753, 757–758, 760). Za słabą stronę swojej metody Sun i in. uznali fakt, że nie radziła sobie ona z niejednoznacznością nazwisk badaczy, tj. z występowaniem nazwisk błędnie zapisanych i nazwisk homonimicznych (istnienie wielu autorów o tym samym nazwisku). Po drugie wskazano, że możliwa jest sytuacja, w której nawet jeżeli na dwóch różnych konferencjach nie było ani jednego tego samego autora, nie oznacza to jeszcze, iż podczas takich konferencji nie poruszano tych samych zagadnień. Zasugerowano ponadto włączenie w zakres zademonstrowanej metody aspektu leksykalnego, np. w postaci uzupełnienia jej o mapę współwystąpień terminów (Sun et al., 2016, 759).

Ostatnią pracą, o której należy wspomnieć jest artykuł autorstwa Olega Eny i in. (2016), w którym podjęto tematykę monitorowania trendów w technice i technologiach (TTM – ang. *technology trend monitoring*) jako narzędzia pomocnego w regulowaniu strategii badawczo-rozwojowych. Środkiem wiodącym do tego celu było kompleksowe rozpatrzenie nie tylko artykułów zamieszczanych w czasopismach naukowych i patentów, ale też szerokiego spektrum raportów i projektów badawczych, prezentacji biznesowych i konferencyjnych, rozpraw doktorskich, prasy i mediów społecznościowych (ang. *social networks*) z zamiarem uchwycenia całego cyklu życiowego i ogółu ścieżek rozwoju nowych technologii (Ena et al., 2016, 1015, 1021, 1025–1027). Autorzy zarzucili ponadto dotychczas wykorzystywanym metodom identyfikacji nowo powstałych specjalności naukowo-technologicznych, a w szczególności funkcjonującym w ich ramach narzędziom służącym terminologicznemu kodowaniu („etykietowaniu”) najnowszych prądów, generowanie terminów zbyt szerokich i zbyt ogólnych, które były w stanie opisać jedynie szerokie i mało konkretne tendencje, zamiast ścisłego wyszczególnienia merytorycznie adekwatnych do aktualnej sytuacji na „froncie badań” kierunków rozwoju i bardziej wyrafinowanych wzorców tego rozwoju (Ena et al., 2016, 1020). Zdaniem autorów zamiar ten można było osiągnąć poprzez uwzględnienie kontekstu terminów/słów kluczowych oraz udoskonalenie istniejących algorytmów grupujących dane wyjściowe. Należy ponadto zapewnić sobie pomoc ekspertów

<sup>13</sup> Sun i in. skupili się przede wszystkim na zbiorze takich właśnie autorów, którzy wykazywali się aktywnością w wielu różnych specjalnościach. Określono ich jako „autorów publikujących w krzyżujących się obszarach” (ang. *cross-fields authors*) (Sun et al., 2016, 751, 757).



w zakresie budowania zbioru terminów/słów kluczowych i interpretowania wszelkiego rodzaju wskaźników ilościowych oraz zwrócić baczniejszą uwagę na aspekty i etapy cyklu życiowego produktów lub technologii takie, jak czołowi badacze i instytucje, kraje wiodące prym w opracowywaniu i stosowaniu danych technologii czy dynamika wzajemnych zależności pomiędzy wszystkimi wymienionymi czynnikami (Ena et al., 2016, 1021–1022, 1024). Odnośnie do udoskonalenia dotychczasowych algorytmów grupujących dane, zdaniem Eny i in., należy wyeliminować z nich elementy, które powodują m.in. konieczność podziału zbioru wyjściowego na węższe podokresy czasowe celem uchwycenia dynamicznego aspektu rozwoju badanych dyscyplin lub niezdolność algorytmu do automatycznego odrzucenia zbędnych (pobocznych) klastrów, powstających jako efekt uboczny słabych i mało istotnych połączeń między dokumentami. Problemem wymagającym uwagi jest również brak uniwersalnej miary podobieństwa dokumentów jako kryterium, w oparciu o które odbywa się podział na podzbiory reprezentujące specjalizacje i obszary problemowe (ang. *core clustering metric*) (Ena et al., 2016, 1023). W miejsce technik grupujących obarczonych powyższymi trudnościami zaproponowano więc własny algorytm, stworzony w *National Research University Higher School of Economics* (HSE) i zaprojektowany z myślą o m.in. możliwości przetwarzania pełnych tekstów dokumentów, inteligentnym stosowaniu list „słów stopujących” (ang. *stop-lists*) oraz funkcjonowaniu na zasadzie tzw. grupowania rozmytego (ang. *fuzzy clustering*). Polega ono na zapewnieniu algorytmowi możliwości przydzielenia tego samego dokumentu do więcej niż jednego klastra. Motywacją wprowadzenia grupowania rozmytego było założenie, że słownictwo używane w publikacjach naukowych jest pochodną wielu obecnych w publikacji koncepcji, idei czy teorii, które zachodzą na siebie i manifestują się poprzez warstwę terminologiczną tekstu. Koncepcje te powtarzają się, współwystępują w kolekcji dokumentów dyscyplinarnych, ale proporcje, w jakich poszczególne dokumenty odnoszą się do poszczególnych koncepcji są niejednakowe, do pewnego stopnia przypadkowe, a ich modelowanie wymaga specjalnego podejścia. Z tego powodu kwalifikacja dokumentów do odpowiednich klastrów powinna dopuszczać przypisywanie ich do wszystkich klastrów, z którymi wykazują one semantyczne pokrewieństwo w mniejszym lub większym stopniu. Dzięki temu ilościowe i terminologiczne charakterystyki klastrów stają się bardziej miarodajne i wiarygodne (Ena et al., 2016, 1023, 1031, 1033; nt. grupowania rozmytego w informatologii zob. też: Blei & Lafferty, 2007, 18; Hood & Wilson, 2002; Miyamoto, 1990, 125–188; Pindlowa, 1994, 102–103).

## 8. Wnioski

Na wstępie należy podkreślić, że obszerność powyższego przeglądu literatury wynika przede wszystkim z braku dostępnych w języku polskim, kompleksowych i systematycznych opracowań poświęconych tematyce metodologii prognozowania i oceny rozwoju dyscyplin naukowych. Autor kierował się w tym względzie chęcią przynajmniej częściowego wypełnienia tej luki i zademonstrowania tego, na jak zaawansowanym poziomie prowadzone są badania nad rozwojem dyscyplin na świecie. W polskiej społeczności specjalistów z zakresu informatologii (w tym bibliometrii) zainteresowanie metodami matematycznymi jest raczej niewielkie, na co wskazała m. in. Wanda Pindlowa (1994, 109). W szczególności, większość istniejących opracowań tego rodzaju, wydanych w języku polskim, skupiała się przede



wszystkim na praktycznym zastosowaniu „mapowania” nauki, metody współcytowań lub metody powiązań bibliograficznych, do analizy konkretnych (fragmentarycznych) korpusów literatury naukowej (zob. np.: Grygiel i in., 2009; Kolasa, 2013; Kowalska & Radzicka, 2012; Osińska, 2010; Osińska, 2012; Potkański, 2010; Ważna & Bartkowski, 2002).

Podsumowując powyższy przegląd, warto zwrócić uwagę po pierwsze na bardzo dużą liczbę metodologicznych podejść do problemu prognozowania rozwoju dyscyplin naukowych oraz na fakt, że różne, konkurencyjne metody są często używane przez naukowców bez wzajemnego nawiązywania do siebie, w izolacji (Tseng et al., 2009, 75). Do tej pory, w obrębie samego tylko nurtu „mapowania” i wizualizowania rozwoju dyscyplin zaproponowano już ponad 20 różnych metod tworzenia map nauki (Yan & Yu, 2016, 1953). Niemal tak samo zróżnicowane i liczne są kryteria podziału zbioru publikacji zasadzające się na ich zakładanym podobieństwie oraz możliwe poziomy agregacji danych doświadczalnych, o których wspomniano już w zakończeniu części pierwszej niniejszego artykułu w ramach podsumowania głównych zarzutów w stosunku do omówionych w niej bibliometrycznych metod prognostycznych. Wobec tak dużego wyboru możliwych technik realizacji badań empirycznych, bardzo trudne (jeżeli nie niemożliwe) jest stwierdzenie, która z dostępnych metod lub ich kombinacji przynosi najlepsze rezultaty z uwagi na zdolność odzwierciedlenia rzeczywistych wzorców rozwoju nauki i która z metod zapewnia największą precyzję wyodrębniania nowo powstałych specjalności. W tym kontekście należy też zwrócić uwagę na niewielką liczbę badań porównawczych i na znaczną rozbieżność ich wyników (zob. podrozdział drugi). W efekcie, w dalszym ciągu pozostaje otwarte pytanie o to, która z metod mapowania nauki jest najlepsza pod względem zdolności wczesnej detekcji frontu badawczego (zob. np.: Huang, 2008, 49; Huang & Chang, 2015, 2043). W związku z tym zachodzi potrzeba dalszego udoskonalania istniejących na gruncie naukometrii metod prognostycznych.

Metody stosowane w omówionym obszarze cechuje ponadto wysoka interdyscyplinarność. Metody te modyfikuje się i zapożycza głównie z dyscyplin ścisłych takich, jak matematyka, informatyka, statystyka matematyczna czy teoria decyzji oraz (rzadziej) z dyscyplin społeczno-ekonomicznych (np. demografii, ekonometrii). Zgadzając się z twierdzeniami niektórych autorów o interdyscyplinarności jako „sile napędowej” innowacyjności, można zająć stanowisko, według którego naukometryczne „mapowanie” i prognozowanie rozwoju dyscyplin jest samo w sobie perspektywiczną specjalnością przeżywającą aktualnie okres swojego rozkwitu. Najprawdopodobniej będzie ona w dalszym ciągu rozwijać się w szybkim tempie za sprawą jej praktycznej przydatności oraz angażowania się w nią autorów pracujących w różnych dyscyplinach nauki, tak ścisłych, jak i społecznych, humanistycznych i ekonomicznych.

Kolejnym wnioskiem płynącym z powyższego przeglądu literatury jest spostrzeżenie, które niejako uzupełnia ogólną krytykę metod prognostycznych wypracowanych na gruncie bibliometrii. Spostrzeżenie to zawiera się w przeświadczeniu, że w metodologii zaczerpniętej ze statystyki matematycznej zawarty jest znaczny potencjał, chociaż wydaje się on być dotąd nie w pełni wykorzystany. Świadczy o tym niewielka liczba publikacji, w których autorzy posługiwali się *stricte* analizą szeregów czasowych. W szczególności zauważalne jest, że w przeciwieństwie do metody tworzenia klastrów publikacji i analizy głównych ścieżek rozwoju dyscyplin, statystyka matematyczna jako jedyna stwarza możliwość szacowania niepewności prognoz, chociaż żaden z autorów prac streszczonych w podrozdziale 6 z niej

nie skorzystał. Ponadto wydaje się, że wyodrębnienie klastrów nie jest mimo wszystko dokładnie tym samym, co prognozowanie w ścisłym sensie. W kategoriach analiz klastrów publikacji nie podaje się bowiem konkretnych, ilościowych wartości jakichkolwiek wskaźników, które można byłoby uznać za np. przewidywaną liczbę cytowań, jakie otrzymają kluczowe dokumenty, patenty czy czasopisma. Nie określa się też precyzyjnie (liczbowo) tempa rozwoju dyscyplin w postaci np. funkcji matematycznej lub prawdopodobieństwa, że specjalność albo obszar badawczy osiągną maksimum swojej siły wpływu np. w ciągu pięciu lat. Wymiar prognostyczny metod grupowania zbioru publikacji (wydzielania klastrów) polega raczej na przyjmowanym *implicite* założeniu, że klastry najaktywniejsze w chwili badania, wykazujące się najwyższą cytawalnością lub największą liczbą publikacji, utrzymają swoją wiodącą pozycję także i w (bliżej nieokreślonej) przyszłości oraz, że to na nich będą koncentrować się dalsze badania, rozwijając je, modyfikując, testując, poszukując dla nich zastosowań praktycznych i wykorzystując osiągnięte już dotychczas wyniki w nowatorski sposób. W przypadku statystycznej analizy szeregów czasowych możliwe jest natomiast nie tylko dokonanie oceny niepewności prognozy, ale też podanie konkretnych liczb, które sygnalizują np. to, jak będzie kształtować się przyszła cytawalność zbioru publikacji dyscyplinarnych czy, alternatywnie, ich liczba, jako tempo produkcji piśmienniczej. W związku z tym wydaje się, że można postawić tezę, według której metody wyróżniania klastrów publikacji oraz identyfikacji głównych ścieżek rozwoju dyscyplin, jako reprezentantów frontów badawczych, służą w większym stopniu ocenie dynamiki rozwoju dyscyplin, podczas gdy ich aspekt prognostyczny jest drugoplanowy. W przypadku metod statystyki matematycznej mamy natomiast do czynienia z sytuacją przeciwną – aspekt prognostyczny dominuje tutaj nad oceną aktualnego stanu rozwoju dyscyplin.

Ostatnim wnioskiem, jaki można wysnuć w oparciu o zaprezentowany przegląd literatury, jest widoczny w niej podział na badania prognostyczne ukierunkowane na analizę cytowań oraz badania prognostyczne ukierunkowane na analizę „surowych” liczb publikacji naukowych przyporządkowanych do określonych specjalności naukowych. Wspólnym mianownikiem wszystkich badań polegających na analizie cytowań jest konieczność akceptacji też tzw. paradygmatu normatywnego funkcjonującego w ramach socjologii wiedzy naukowej. Główne założenia tego paradygmatu sformułowała m.in. Linda Smith (1981). Brzmiały one następująco:

- (1) Cytowanie dokumentu implikuje jego faktyczne wykorzystanie przez naukowca. To znaczy, że przywołuje on wszystkie (lub przynajmniej większość) prace przezeń wykorzystane oraz wszystkie prace przywołane rzeczywiście były wykorzystane. Inaczej mówiąc, naukowiec ani nie cytuje prac nieużywanych, ani nie pomija tych wykorzystanych.
- (2) Cytowanie odzwierciedla zalety dokumentu cytowanego, jego jakość, znaczenie, siłę wpływu.
- (3) Cytuje się prace najlepsze z możliwych.
- (4) Istnieje związek treściowy pracy cytującej i cytowanej.
- (5) Wszystkie cytowania są równorzędne (o równej wartości) (Smith, 1981, 87–89).

Założenia te były przedmiotem krytyki ze strony tzw. paradygmatu konstruktywistycznego, w którym akcentuje się konceptualizację cytowań jako aktów posiadających rozliczne uwarunkowania socjologiczne i psychologiczne, niezwiązane tylko z bezinteresownym dążeniem do spłaty intelektualnego długu wobec poprzedników. Jako takie, cytowania są

w świetle konstruktywizmu tworem podporządkowanym bieżącym potrzebom cytującego. Kiedy używa on pracy cudzej do swoich celów, może dostosowywać, zmniejszać, nadinterpretować znaczenie cytowanych ustępów, naginać ich sens do swoich potrzeb, konstruując go na nowo w postaci odpowiadającej jego intencjom i w oderwaniu od pierwotnego zamysłu naukowca cytowanego. Zrozumienie natury, celu i funkcji cytowań wymaga więc uwzględnienia złożonych sytuacji, w jakich powstają publikacje. Obejmują one codzienną rutynę i okoliczności pracy naukowców, osobiste motywacje, interakcje ze współpracownikami, przełożonymi i instytucjami itd. (zob. np.: Law & French, 1974, 589; Moed, 2005, 196). Kwestia zasadności i wiarygodności metod opartych na analizie cytowań, takich jak grupowanie publikacji w klastry, analizowanie ścieżek rozwoju dyscyplin i statystyczne przewidywanie przyszłej cytawalności, spoczywa więc ostatecznie na indywidualnych zachowaniach badaczy w zakresie cytowań (ang. *citation behaviour*). W tym sensie epistemologiczną podstawą całego nurtu prognostycznego w bibliometrii jest socjologia wiedzy naukowej, w ramach której toczy się dyskusja nad tym, który ze wspomnianych paradygmatów trafniej oddaje rzeczywistą rolę, jaką w nauce odgrywają cytowania piśmiennictwa.

Z drugiej strony, niektóre zarzuty wobec polegania na „surowej” liczbie publikacji w dyscyplinie, jako na zbyt mało wyrafinowanym wskaźniku tempa rozwoju nauki, częściowo przedstawiono już w podrozdziale 6. W tym miejscu można natomiast dodać, że jeszcze jedną trudnością związaną z tym podejściem jest problem w precyzyjnym wytyczeniu granic poszczególnych dyscyplin. Problem ten wiąże się bezpośrednio z dyskutowanym w informatologii zagadnieniem relewancji tematycznej publikacji naukowych (zob. np.: Hjørland, 2000; Saracevic, 1975). Oba podejścia do prognozowania rozwoju dyscyplin okazują się zatem nie tylko bardzo złożone, ale też możliwe do zakwestionowania w perspektywie teoretycznej i zaangażowane w debatę pozornie tylko odległą od swojego przedmiotu badań. Stanowi to dodatkowy, głębszy wymiar metodologii przewidywania ewolucji dyscyplin naukowych.

## Bibliografia

- Aczel, A.D. (2007). *Statystyka w zarządzaniu*. Warszawa: PWN.
- Bingham, N.H., Fry, J.M. (2010). *Regression. Linear Models in Statistics*. London: Springer-Verlag.
- Blei, D.M., Lafferty, J.D. (2007). A Correlated Topic Model of Science. *The Annals of Applied Statistics*, 1(1), 17–35.
- Blondel, V.D., Guillaume, J-L., Lambiotte, R., Lefebvre, E. (2008). Fast Unfolding of Communities in Large Networks [online]. *Journal of Statistical Mechanics: Theory and Experiment*, 2008, 1–12, [30.08.2017], <http://iopscience.iop.org/article/10.1088/1742-5468/2008/10/P10008/meta>
- Bornmann, L., Daniel, H-D. (2008). What Do Citation Counts Measure? A Review of Studies on Citing Behavior. *Journal of Documentation*, 64(1), 45–80.
- Boyack, K.W., Klavans, R. (2010). Co-Citation Analysis, Bibliographic Coupling and Direct Citation: Which Citation Approach Represents the Research Front Most Accurately? *Journal of the American Society for Information Science and Technology*, 61(12), 2389–2404.
- Boyack, K.W., Klavans, R. (2014). Creation of Highly Detailed, Dynamic, Global Model and Map of Science. *Journal of the Association for Information Science and Technology*, 65(4), 670–685.
- Chen, S-H., Huang, M-H., Chen, D-Z. (2012). Identifying and Visualizing Technology Evolution: A Case Study of Smart Grid Technology. *Technological Forecasting and Social Change*, 79(6), 1099–1110.
- Chi, Y., Tseng, B.L., Tatemura, J. (2006). Eigen-Trend: Trend Analysis in the Blogosphere Based on Singular Value Decompositions [online]. *CIKM, 06, Proceedings of the 15th ACM International*

- Conference on Information and Knowledge Management*, Arlington, Virginia, USA, November 06–11, 2006, 68–77, [30.08.2017], <http://dl.acm.org/citation.cfm?id=1183628>
- Cipra, T., Hanzák, T. (2008). Exponential smoothing for irregular time series. *Kybernetika*, 44(3), 385–399.
- Choudhury, N., Uddin, S. (2016). Time-Aware Link Prediction to Explore Network Effects on Temporal Knowledge Evolution. *Scientometrics*, 108(2), 745–776.
- Daim, T., Rueda, G., Martin, H., Gerdtsri, P. (2006). Forecasting Emerging Technologies: Use of Bibliometrics and Patent Analysis. *Technological Forecasting and Social Change*, 73(8), 981–1012.
- Debackere, K., Verbeek, A., Luwel, M., Zimmermann, E. (2002). Measuring Progress and Evolution in Science and Technology – II: The Multiple Uses of Technometric Indicators. *International Journal of Management Reviews*, 4(3), 213–231.
- Ena, O., Mikova, N., Saritas, O., Sokolova, A. (2016). A Methodology for Technology Trend Monitoring: The Case of Semantic Technologies. *Scientometrics*, 108(3), 1013–1041.
- Érdi, P., Makovi, K., Somogyvári, Z., Strandburg, K., Tobochnik, J., Volf, P., Zalányi, L. (2013). Prediction of Emerging Technologies Based on Analysis of the US Patent Citation Network. *Scientometrics*, 95(1), 225–242.
- Glänzel, W. (2012). Bibliometric Methods for Detecting and Visualizing Emerging Research Topics. *El Profesional de la Información*, 21(1), 194–201.
- Glänzel, W., Thijs, B. (2012). Using ‘Core Documents’ for Detecting and Labelling New Emerging Topics. *Scientometrics*, 91(2), 399–416.
- Grygiel, P., Rębisz, S., Humenny, G. (2009). Analiza bibliometryczna jako narzędzie badania efektywności nauczycieli akademickich na przykładzie Uniwersytetu Rzeszowskiego. *Zarządzanie Publiczne. Zeszyty Naukowe Instytutu Spraw Publicznych Uniwersytetu Jagiellońskiego*, 3(7), 65–84.
- Guo, H., Weingart, S., Börner, K. (2011). Mixed-Indicators Model for Identifying Emerging Research Areas. *Scientometrics*, 89(1), 421–435.
- Haegeman, K., Marinelli, E., Scapolo, F., Ricci, A., Sokolov, A. (2013). Quantitative and Qualitative Approaches in Future-Oriented Technology Analysis (FTA): From Combination to Integration? *Technological Forecasting and Social Change*, 80(3), 386–397.
- Hjørland, B. (2000). Relevance Research: The Missing Perspective(s): ‘Non-Relevance’ and ‘Epistemological Relevance’. *Journal of the American Society for Information Science*, 51(2), 209–211.
- Hood, W.W., Wilson, C.S. (2002). Solving Problems in Library and Information Science Using Fuzzy Set Theory. *Library Trends*, 50(3), 393–405.
- Huang, A. (2008). Similarity Measures for Text Document Clustering. In: J. Holland, A. Nicholas, D. Brignoli (eds.). *Proceedings of the Sixth New Zealand Computer Science Research Student Conference (NZCSRSC 2008)*. Christchurch, New Zealand, April 14–18, 2008, 49–56 [online]. University of Canterbury, [30.08.2017], [https://nzcsrsc08.canterbury.ac.nz/site/proceedings/NZCSRSC\\_2008\\_Proceedings.pdf](https://nzcsrsc08.canterbury.ac.nz/site/proceedings/NZCSRSC_2008_Proceedings.pdf)
- Huang, M-H., Chang, C-P. (2015). A Comparative Study on Detecting Research Fronts in the Organic Light-Emitting Diode (OLED) Field Using Bibliographic Coupling and Co-Citation. *Scientometrics*, 102(3), 2041–2057.
- Hummon, N.P., Doreian, P. (1989). Connectivity in a Citation Network: The Development of DNA Theory. *Social Networks*, 11(1), 39–63.
- Jarneving, B. (2005). A Comparison of Two Bibliometric Methods for Mapping of the Research Front. *Scientometrics*, 65(2), 245–263.
- Jun, S., Park, S.S., Jang, D.S. (2012). Technology Forecasting Using Matrix Map and Patent Clustering. *Industrial Management and Data Systems*, 112(5), 786–807.
- Klafter, J., Sokolov, I.M. (2015). *First Steps in Random Walks*. Oxford: Oxford University Press.
- Kolasa, M.W. (2013). *Historiografia prasy polskiej (do 1918 roku)*. *Naukometryczna analiza dyscypliny 1945–2009*. Kraków: Wydaw. Naukowe Uniwersytetu Pedagogicznego.

- Kowalska, A., Radzicka, J. (2012). Programy BibExcel i Pajek w analizach bibliometrycznych. *Biuletyn EBIB* [online], 3(130), 1–7, [30.08.2017], [http://www.ebib.pl/images/stories/numery/130/130\\_kowalska\\_radzicka.pdf](http://www.ebib.pl/images/stories/numery/130/130_kowalska_radzicka.pdf)
- Krampen, G., Von Eye, A., Schui, G. (2011). Forecasting Trends of Development of Psychology from a Bibliometric Perspective. *Scientometrics*, 87(3), 687–694.
- Krawiec, S. (2014). *Adaptacyjne modele wygładzania wykładniczego jako instrumenty prognozowania krótkoterminowego zjawisk ilościowych*. Gliwice: Wydaw. Politechniki Śląskiej.
- Krzywda, M. (2010). *Analiza techniczna w praktyce*. Gliwice: Wydaw. Złote Myśli.
- Kucharavy, D., De Guio, R. (2011). Application of S-Shaped Curves. *Procedia Engineering*, 9, 559–572.
- Kuhn, T.S. (2001). *Struktura rewolucji naukowych*. Warszawa: Fundacja Aletheia.
- Kuusi, O., Meyer, M. (2007). Anticipating Technological Breakthroughs: Using Bibliographic Coupling to Explore the Nanotubes Paradigm. *Scientometrics*, 70(3), 759–777.
- Law, J., French, D. (1974). Normative and Interpretive Sociologies of Science. *The Sociological Review*, 22(4), 581–595.
- Lee, L.-C., Lee, Y.-Y., Liaw, Y.-C. (2012). Bibliometric Analysis for Development of Research Strategies in Agricultural Technology: The Case of Taiwan. *Scientometrics*, 93(3), 813–830.
- Leydesdorff, L. (2007). Betweenness Centrality as an Indicator of the Interdisciplinarity of Scientific Journals. *Journal of the American Society for Information Science and Technology*, 58(9), 1303–1319.
- Liu, M. (1993). The Complexities of Citation Practice: A Review of Citation Studies. *Journal of Documentation*, 49(4), 370–408.
- Liu, P., Chen, B.-L., Liu, K., Xie, H. (2016). Magnetic Nanoparticles Research: A Scientometric analysis of Development Trends and Research Fronts. *Scientometrics*, 108(3), 1591–1602.
- Liu, C., Gui, Q. (2016). Mapping Intellectual Structures and Dynamics of Transport Geography Research: A Scientometric Overview from 1982 to 2014. *Scientometrics*, 109(1), 159–184.
- Liu, J.S., Lu, L.Y.Y. (2012). An Integrated Approach for Main Path Analysis: Development of the Hirsh Index as an Example. *Journal of the American Society for Information Science and Technology*, 63(3), 528–542.
- Łapkowska-Baster, B. (2009). *Miary współzależności i dynamiki zjawisk w statystyce opisowej. Przykłady i zadania*. Kraków: Wydaw. UJ.
- Ma, V.C., Liu, J.S. (2016). Exploring the Research Fronts and Main Paths of Literature: A Case Study of Shareholder Activism Research. *Scientometrics*, 109(1), 33–52.
- Malina, A. (1994). Prognozowanie zjawisk ekonomicznych w oparciu o metody wykładniczego wygładzania szeregów czasowych. *Zeszyty Naukowe/Akademia Ekonomiczna w Krakowie*, 440, 15–29.
- Miyamoto, S. (1990). *Fuzzy Sets in Information Retrieval and Cluster Analysis*. Dordrecht: Springer Science+Business Media B.V.
- Moed, H. (2005). *Citation Analysis in Research Evaluation*. Dordrecht: Springer.
- Montgomery, D.C., Jennings, C.L., Kulahci, M. (2008). *Introduction to Time Series Analysis and Forecasting*. Hoboken: John Wiley & Sons, Inc.
- Morris, S.A. (2005). Manifestation of Emerging Specialties in Journal Literature: A Growth, Model of Papers, References, Exemplars, Bibliographic Coupling, Cocitation, and Clustering Coefficient Distribution. *Journal of the American Society for Information Science and Technology*, 56(12), 1250–1273.
- Newman, M.E.J. (2004). Fast Algorithm for Detecting Community Structure in Networks. *Physical Review E*, 69(6), art. no. 066133, 1–5.
- Osińska, V. (2010). Rozwój metod mapowania domen naukowych i potencjał analityczny w nim zawarty. *Zagadnienia Informatyki Naukowej*, 2(96), 41–51.
- Osińska, V. (2012). Wizualizacja paradygmatów badawczych. *Zagadnienia Naukoznawstwa*, 3(193), 205–220.
- Persson, O. (1994). The Intellectual Base and Research Fronts of JASIS 1986–1990. *Journal of the American Society for Information Science*, 45(1), 31–38.



- Pindlowa, W. (1994). *Informetria w nauce o informacji. Metody i problemy*. Kraków: Towarzystwo Autorów i Wydawców Prac Naukowych „Universitas”.
- Potkański, J. (2010). Samoobserwacja współczesnej nauki: spojrzenie liczb. *Przegląd Humanistyczny*, 05/06, 137–150.
- Price, D. de Solla (1967). *Mała Nauka – Wielka Nauka*. Warszawa: PWN.
- Saracevic, T. (1975). Relevance: A Review of and a Framework for the Thinking on the Notion in Information Science. *Journal of the American Society for Information Science*, 26(6), 321–343.
- Shibata, N., Kajikawa, T., Matsushima, K. (2007). Topological Analysis of Citation Networks to Discover the Future Core Articles. *Journal of the American Society for Information Science and Technology*, 58(6), 872–882.
- Shibata, N., Kajikawa, Y., Takeda, Y., Matsushima, K. (2008). Detecting Emerging Research Fronts Based on Topological Measures in Citation Networks of Scientific Publications. *Technovation*, 28(11), 758–775.
- Shibata, N., Kajikawa, Y., Takeda, Y., Matsushima, K. (2009). Comparative Study on Methods of Detecting Research Fronts Using Different Types of Citation. *Journal of the American Society for Information Science and Technology*, 60(3), 571–580.
- Shumway, R.H., Stoffer, D.S. (2011). *Time Series Analysis and Its Applications: With R Examples*. New York: Springer.
- Sitarz, R., Heneczkowski, M., Jabłońska-Sabuka, M., Krasławski, A. (2012). *Trend Analysis Methods as an Indicator of the Dynamic of Knowledge Flow in Research on Polymeric Materials: Composites, Nanocomposites and Blends* [online]. Extended version of a paper presented during 11th International Symposium on Process Systems Engineering, 15–19 July 2012, Singapore, [30.08.2017], [https://www.researchgate.net/profile/Matylda\\_Jablonska-Sabuka/publication/266142586\\_Clustering\\_Method\\_for\\_Analysis\\_of\\_Research\\_Fields\\_Examples\\_of\\_Composites\\_Nanocomposites\\_and\\_Blends/links/555c3f2d08aec5ac2232aec1.pdf](https://www.researchgate.net/profile/Matylda_Jablonska-Sabuka/publication/266142586_Clustering_Method_for_Analysis_of_Research_Fields_Examples_of_Composites_Nanocomposites_and_Blends/links/555c3f2d08aec5ac2232aec1.pdf)
- Small, H., Boyack, K.W., Klavans, R. (2014). Identifying Emerging Topics in Science and Technology. *Research Policy*, 43(8), 1450–1467.
- Smith, L.C. (1981). Citation Analysis. *Library Quarterly*, 30(1), 83–106.
- Snarska, A. (2011). *Statystyka. Ekonometria. Prognozowanie. Ćwiczenia z Excelem 2007*. Warszawa: Placet.
- Sobczyk, M. (2008). *Prognozowanie. Teoria, przykłady, zadania*. Warszawa: Placet.
- Sobczyk, M. (2015). *Statystyka*. Warszawa: PWN.
- Sun, X., Ding, K., Lin, Y. (2016). Mapping the Evolution of Scientific Fields Based on Cross-Field Authors. *Journal of Informetrics*, 10(3), 750–761.
- Taylor, J.R. (2012). *Wstęp do analizy błędu pomiarowego*. Warszawa: PWN.
- Tseng, Y-H., Lin, Y-I., Lee, Y-Y., Hung, W-C., Lee, C-H. (2009). A Comparison of Methods for Detecting Hot Topics. *Scientometrics*, 81(1), 73–90.
- Tu, Y-N., Hsu, S-L. (2016). Constructing Conceptual Trajectory Maps to Trace the Development of Research Fields. *Journal of the Association for Information Science and Technology*, 67(8), 2016–2031.
- Upham, S.P., Small, H. (2010). Emerging Research Fronts in Science and Technology: Patterns of New Knowledge Development. *Scientometrics*, 83(1), 15–38.
- Van Den Besselaar, P., Heimeriks, G. (2006). Mapping Research Topics Using Word-Reference Co-Occurrences: A Method and an Exploratory Case Study. *Scientometrics*, 68(3), 377–393.
- Verbeek, A., Debackere, K., Luwel, M., Zimmermann, E. (2002). Measuring Progress and Evolution in Science and Technology – I: The Multiple Uses of Bibliometric Indicators. *International Journal of Management Reviews*, 4(2), 179–211.
- Wang, C-Ch., Ho, Y-S. (2016). Research Trend of Metal-Organic Frameworks: A Bibliometric Analysis. *Scientometrics*, 109(1), 481–513.



- Wang, X., Wang, Z., Xu, S. (2013). Tracing Scientists' Research Trends Realtimely. *Scientometrics*, 95(2), 717–729.
- Ważna, L., Bartkowski, A. (2002). Patenty i publikacje naukowe jako wskaźniki technologiczne. Bibliometria strukturalna. *Management*, 6(2), 171–182.
- Yan, E., Yu, Q. (2016). Using Path-Based Approaches to Examine the Dynamic Structure of Discipline-Level Citation Networks: 1997–2011. *Journal of the Association for Information Science and Technology*, 67(8), 1943–1955.
- Yoon, B., Park, Y. (2007). Development of New Technology Forecasting Algorithm: Hybrid Approach for Morphology Analysis and Conjoint Analysis of Patent Information. *IEEE Transactions on Engineering Management*, 54(3), 588–599.
- Zeliaś, A., Pawełek, B., Wanat, S. (2013). *Prognozowanie ekonomiczne. Teoria, przykłady, zadania*. Warszawa: PWN.
- Zheng, T., Wang, J., Wang, Q., Nie, C., Shi, Z., Wang, X., Gao, Z. (2016). A Bibliometric Analysis of Micro/Nano-Bubble Related Research: Current Trends, Present Application, and Future Prospects. *Scientometrics*, 109(1), 53–71.
- 

## **Bibliometric Methods to Foresee and Assess the Development of Scientific Disciplines. Literature Analysis. Part 2. Comparisons, Hybrid and Statistical Methods, Analysis of Patents and Main Paths of Literature Development and Other Original Approaches in Terms of Predictive Methodology**

### **Abstract**

**Purpose/Thesis:** The aim of the paper is to provide an evaluative review of literature on the quantitative methods of bibliometric forecasting and assessing the current state of scientific discipline development. By “evaluative” review it is understood that the author discussed and pinpointed the drawbacks of the literature-proposed ways to assess and predict the development of scientific disciplines.

**Approach/Methods:** The method used by the author is the analysis of the literature concerning the problems under consideration. The author also proposed a typology of methods applied in the bibliometric predictions area and emphasized the flaws of some previously used methods. Moreover, the possibilities of practical application of foresight exercises were shown. Finally, the conclusions were drawn, with regard to the methods oriented towards the assessment of the rate and direction of scientific disciplines development.

**Results and conclusions:** First conclusion was that there was a great diversity within the sphere of quantitative methods concerning the subject of evolution forecasting for scientific disciplines. The methods in question often are applied without referring to each other. In other words, the majority of predictive methods are used in isolation. The flaws, occurring in many methods, usually have a negative influence on the results obtained. Next conclusion was that there was a significant potential in the statistical analysis of the time series. However, it seems that this potential has not been fully exploited yet. Furthermore, there is a considerable degree of interdisciplinarity in the scope of bibliometric predictive methods. Often, they are modified and adopted from other disciplines such as mathematics, computer science, statistics and econometrics. The last conclusion was that

the predictive process was based either on purposefully designed citation analysis or solely on the number of disciplinary publications. In the first case the methodology must implicitly accept the theses of so-called normative paradigm, existing within the sociology of scientific knowledge and constituting an epistemological background for the methodology. In the latter case the methodology is engaged in the debate over topical relevance, which is observed within the frames of library and information science.

**Originality/Value:** The originality and value of the paper lies in an effort to demonstrate the scope and a high degree of progress of the studies on the possibilities of predicting and assessing the development of scientific disciplines worldwide. In Polish scientific environment such studies are relatively rare, and it seems that the review of the predictive bibliometric methodology could contribute to the stimulation of interest in the research directed towards the forecasting of future scientific trends. It is also important to stress the disadvantages of selected methods of bibliometric forecasting as in this context they create a substantial amount of space destined for possible improvements and modifications of existing methods. In the field of bibliometrics there is a considerable need for such refinements due to the fact that currently there is no perfect method which could be described as "the last word" on the technique of anticipating future research directions.

### Keywords

Bibliometrics. Development of scientific disciplines. Predictions. Quantitative methods in information science.

---

*ŁUKASZ OPALIŃSKI* uzyskał tytuł magistra w zakresie nauki o języku i komunikacji (jako specjalności wyodrębnionej z kierunku filozofia) na Uniwersytecie Marii Curie-Skłodowskiej w Lublinie w 2005 r. Ukończył studium podyplomowe *Technologie Informacyjne i Bibliotekoznawstwo* w Wyższej Szkole Informatyki i Zarządzania z siedzibą w Rzeszowie w 2009 r. oraz studium podyplomowe *Bibliotekarz Dziedziny* w Instytucie Informacji Naukowej i Bibliotekoznawstwa Uniwersytetu Jagiellońskiego w 2012 r. Pracuje w Oddziale Informacji Naukowej Biblioteki Politechniki Rzeszowskiej na stanowisku bibliotekarza. Najważniejsze publikacje: Ł. Opaliński (2012). *Niealfabetyczne systemy piśmiennicze a komunikacja językowa w Internecie* [online]. *Podkarpackie Studia Biblioteczne*, 1/2012, 1–19, [25. 02. 2017], <http://repozytorium.ur.edu.pl/handle/item/180>; Ł. Opaliński (2013). *Wybrane aspekty metodologii badań cyklu życiowego publikacji naukowych*. *Przegląd Biblioteczny*, 81(2), 152–171; Ł. Opaliński, M. Jaromin, J. Wikiera (2015). *Problem stabilności zachowań naukowców w zakresie cytowań w kontekście metodologii badań starzenia się publikacji naukowych i możliwość jego ujęcia ilościowego*. *Zagadnienia Informacji Naukowej*, 53(2/106), 65–83.

Kontakt z autorem:

[lopa@prz.edu.pl](mailto:lopa@prz.edu.pl)

Oddział Informacji Naukowej Biblioteki Politechniki Rzeszowskiej

al. Powstańców Warszawy 12

35–959 Rzeszów

# Zastosowanie statystycznej analizy szeregów czasowych do krótkoterminowego prognozowania rozwoju dyscyplin naukowych

Łukasz Opaliński

*Biblioteka Politechniki Rzeszowskiej im. I. Łukasiewicza*

Marcin Jaromin

*Zakład Biotechnologii i Bioinformatyki Wydziału Chemicznego Politechniki Rzeszowskiej*

## Abstrakt

**Cel/Teza:** Celem pracy jest opis podstaw statystycznej metody analizy szeregów czasowych, zbadanie możliwości jej zastosowania do krótkoterminowego prognozowania rozwoju dyscyplin naukowych w oparciu o dane na temat cytawalności publikacji dyscyplinarnych, wykorzystanie zapewnianej przez metodę możliwości oceny niepewności prognozy, a także interpretacja uzyskanych wyników w kategoriach naukoznawczych.

**Koncepcja/Metody badań:** W ramach niniejszego opracowania zastosowano metodę analitycznej dekompozycji szeregu czasowego jako metodę wyodrębniania trendu rozwojowego w oparciu o dane na temat dynamiki zjawisk masowych. Zjawiskiem masowym było w tym przypadku cytowanie literatury naukowej w polskojęzycznej sferze nauk o Ziemi. Metoda polega w głównej mierze na sformułowaniu równania opisującego ilościowo przebieg zjawiska z rozróżnieniem trendu głównego oraz składowej odpowiadającej wahaniom przypadkowym. Równanie stanowi podstawę przewidywania przyszłych wartości szeregu czasowego.

**Wyniki i wnioski:** Uzyskane wyniki przemawiają za możliwością efektywnego zastosowania statystycznej analizy szeregów czasowych do krótkoterminowego prognozowania tempa rozwoju dyscyplin naukowych. W przypadku zrealizowanego badania prognozy okazały się akceptowalne dla dwuletniego horyzontu prognozy, chociaż istnieją podstawy, aby przypuszczać, że skonstruowanie prognoz w postaci przedziałów ufności lub nieznaczne zmodyfikowanie metody mogłoby wydłużyć ten okres. Wyniki zinterpretowano w kategoriach naukoznawczych jako uproszczony indeks natychmiastowości dyscyplin, który jest szczególnie predysponowany do praktycznych zastosowań komparatywnych.

**Oryginalność/Wartość poznawcza:** W ramach analizy literatury przedmiotu nie natrafiono na przykłady badań, które byłyby prowadzone w sposób zaproponowany w ramach niniejszej publikacji. Analiza szeregów czasowych była wcześniej stosowana jedynie do wyników podziału zbioru publikacji dyscyplinarnych na, nazywane „klastrami”, podzbiory prac w założeniu reprezentatywnych dla określonych subdyscyplin. Operacja podziału jest zawsze dokonywana w oparciu o jedno z wielu dostępnych kryteriów podobieństwa publikacji, a w szczególności kryterium oparte na współwystępowaniu słów i terminów kluczowych oraz w oparciu o bardzo liczne i dające różne efekty algorytmy grupujące. Wydaje się w związku z tym, że podejście oparte na bezpośredniej cytawalności publikacji pozwala wyeliminować ze zbioru wyników element pewnej niejednoznaczności (względności) uwarunkowany znaczącym zróżnicowaniem i brakiem uniwersalności w zakresie sposobu działania algorytmów grupujących oraz kryteriów kognitywnego podobieństwa publikacji naukowych.

## Słowa kluczowe

Analiza szeregów czasowych. Bibliometria. Czasopisma naukowe. Metody statystyczne w informatologii. Rozwój dyscyplin naukowych.

*Otrzymany: 1 marca 2017. Zrecenzowany: 19 lipca 2017. Zaakceptowany: 22 grudnia 2017.*

## 1. Wprowadzenie

Prognozowanie tempa i kierunków rozwoju dyscyplin naukowych jest jednym z tradycyjnych zadań stojących przed bibliometrią. Wspominał o nim już Alan Pritchard w 1969 r. pisząc, że celem bibliometrii jest m.in. „naświetlenie (...) natury i kierunków rozwoju dyscyplin [naukowych] (...) za pomocą zliczania i analizowania różnorodnych aspektów komunikacji piśmienniczej”<sup>1</sup> (Pritchard, 1969, 348). Z oceną dynamiki i prognozowaniem dalszego rozwoju dyscyplin naukowych ściśle wiąże się wyodrębnianie nowo powstających i nie w pełni jeszcze ukształtowanych (ang. *emerging*) subdyscyplin (obszarów problemowych czy, inaczej mówiąc, wąskich specjalności dyscyplinarnych) (Skalska-Zlat, 1993, 20; Skalska-Zlat, 1999, 60). Ta ostatnia funkcja działalności bibliometrycznej polega najczęściej na tzw. mapowaniu (ang. *mapping*) nauki, czyli odwzorowywaniu i wizualizowaniu poznawczej (teoretycznej, logicznej czy heurystycznej) struktury różnych dyscyplin naukowych i tworzeniu tzw. topologicznych map nauki (zob. np. Chen, 2006; Garfield, 2004, 119). Niektórzy badacze twierdzą przy tym, że okres najintensywniejszych prac prowadzonych w ramach tego odgałęzienia bibliometrii przypada na ostatnie 15 lat (Ena et al., 2016, 1015; Wang & Ho, 2016, 482), chociaż początków realizacji takich zadań można dopatrywać się już w latach 50. XX w., zaraz po założeniu filadelfijskiego Institute for Scientific Information (Boyack & Klavans, 2014, 671). Obecnie można nawet spotkać się z opinią, że zadanie „mapowania” nauki i rozpoznawanie – mówiąc słowami Dereka de Solli Price’a – „frontów badawczych” (ang. *research fronts*) jest pierwszoplanowym komponentem bibliometrii (Huang & Chang, 2015, 2042; Price, 1967, 114; Wang & Ho, 2016, 482).

Przewidywanie dalszego rozwoju dyscyplin naukowych niesie ze sobą rozliczne korzyści praktyczne. Jest ono na tyle istotne, że istnieją nawet towarzystwa naukowe i cykliczne konferencje, które poświęca się tym zagadnieniom (Small et al., 2014, 1450). Po pierwsze, jest ono cenne z punktu widzenia samych naukowców. Śledzenie najnowszych tendencji dyscyplinarnych i identyfikacja najbardziej perspektywicznej tematyki i pytań badawczych (ang. *hot topics*) pozwoliłyby naukowcom nadążyć za szybkim i wielopłaszczyznowym rozwojem dyscyplin, a także racjonalnie ukierunkować swoje wysiłki badawcze lub ułatwić planowanie przyszłej współpracy naukowej (Takeda & Kajikawa, 2009, 544; Tu & Hsu, 2016, 2016–2017). Drugim obszarem praktycznych zastosowań bibliometrycznego prognozowania rozwoju dyscyplin naukowych, jest ogólnie rozumiana polityka naukowa, w skład której wchodzi administrowanie i zarządzanie działalnością naukową oraz kwestie finansowania badań naukowych z budżetu państwa (zob. np. Braun et al., 2000, 24). Ponieważ badania te są „zaczynem” (ang. *seeds*) innowacyjności przemysłowej, znaczenie prac badawczo-rozwojowych (ang. *R&D – Research and Development*) jest uznawane za fundamentalne w procesie stymulowania innowacyjności i konkurencyjności narodowego przemysłu i technologii (Shibata et al., 2008, 758–759). W kontekście pragmatycznego znaczenia prognoz rozwoju dyscyplin wydaje się też, że mogą one okazać się wartościowe także dla bibliotekarzy, a zwłaszcza dla bibliotekarzy dziedzinowych. Hanning Guo i in. (2011) zauważyli bowiem, że ta grupa zawodowa może odnieść korzyść wynikającą z dostępu do narzędzi ułatwiających tworzenie nowych kategorii tematycznych lub haseł opisu

<sup>1</sup> Oryg.: „(...) to shed light on (...) the nature and course of development of a discipline (...), by means of counting and analysing the various facets of written communication” (Pritchard, 1969, 348).

rzeczowego w trakcie opracowywania zbiorów, uzupełnianie słowników istniejących języków informacyjno-wyszukiwawczych (np. tezaurusów) o nowe terminy, albo wyodrębnianie specjalnych kolekcji zbiorów bibliotecznych, ukierunkowanych na potrzeby informacyjne badaczy prowadzących studia o profilu zgodnym z nowo powstającymi subdyscyplinami (Guo et al., 2011, 422).

## 2. Literatura przedmiotu

Literatura przedmiotu dotycząca wyodrębniania nowo powstających specjalności w ramach dobrze ugruntowanych teoretycznie dyscyplin oraz prognozowania ich rozwoju została szczegółowo omówiona w dwóch odrębnych opracowaniach pierwszego z autorów niniejszego artykułu (Opaliński, 2017a; Opaliński, 2017b). W związku z tym wydaje się, że w tym miejscu wystarczy przypomnieć jeden z głównych wniosków płynących z analizy literatury przedmiotu, według którego w metodologii zaczerpniętej ze statystyki matematycznej zawarty jest znaczny potencjał, chociaż wydaje się on dotąd nie w pełni wykorzystany. Świadczy o tym niewielka liczba publikacji, w których autorzy posługiwali się *stricte* analizą szeregów czasowych (zob. Opaliński, 2017b). W szczególności zauważalne jest, że w przeciwieństwie do m.in. metody tworzenia klastrów publikacji i analizy głównych ścieżek rozwoju dyscyplin, statystyka matematyczna jako jedyna stwarza możliwość szacowania niepewności prognoz. Ponadto wydaje się, że wyodrębnienie klastrów nie jest dokładnie tym samym co prognozowanie w ścisłym sensie. W kategoriach analiz klastrów publikacji nie podaje się bowiem konkretnych, ilościowych wartości jakichkolwiek wskaźników, które można byłoby uznać za np. przewidywaną liczbę cytowań, jakie otrzymają kluczowe dokumenty, patenty czy czasopisma. Nie określa się też liczbowo tempa rozwoju dyscyplin w postaci np. funkcji matematycznej lub prawdopodobieństwa, że specjalność albo obszar badawczy osiągnie maksimum swojej siły wpływu np. w ciągu pięciu lat. Wymiar prognostyczny metod grupowania zbioru publikacji polega raczej na przyjmowanym *implicite* założeniu, że klastry najaktywniejsze w chwili badania, wykazujące się najwyższą cytawalnością lub największą liczbą publikacji, utrzymają swoją wiodącą pozycję także w (bliżej nieokreślonej) przyszłości oraz że to na poruszanej w nich problematyce będą koncentrować się dalsze badania, rozwijając ją, modyfikując, testując i poszukując dla niej zastosowań praktycznych. W przypadku statystycznej analizy szeregów czasowych możliwe jest natomiast nie tylko dokonanie oceny niepewności prognozy, ale też podanie konkretnych wartości, które sygnalizują np. to, jak będzie kształtować się przyszła cytawalność zbioru publikacji dyscyplinarnych czy, alternatywnie, ich liczba, jako tempo produkcji piśmienniczej. W związku z tym można postawić tezę, że metody wyróżniania klastrów publikacji oraz identyfikacji głównych ścieżek rozwoju dyscyplin, jako reprezentantów frontów badawczych, służą w większym stopniu ocenie dynamiki rozwoju dyscyplin, podczas gdy ich aspekt prognostyczny jest drugoplanowy. W przypadku metod statystyki matematycznej mamy natomiast do czynienia z sytuacją przeciwną – aspekt prognostyczny dominuje tutaj nad oceną aktualnego stanu rozwoju dyscyplin.

### 3. Cel pracy

Celem pracy jest opis podstaw statystycznej metody analizy (dekompozycji) szeregów czasowych, zbadanie możliwości jej zastosowania do krótkoterminowego prognozowania tempa rozwoju dyscyplin naukowych w oparciu o dane na temat cytawalności publikacji dyscyplinarnych, wykorzystanie zapewnianej przez metodę możliwości oceny niepewności prognozy, a także interpretacja uzyskanych wyników w kategoriach naukoznawczych.

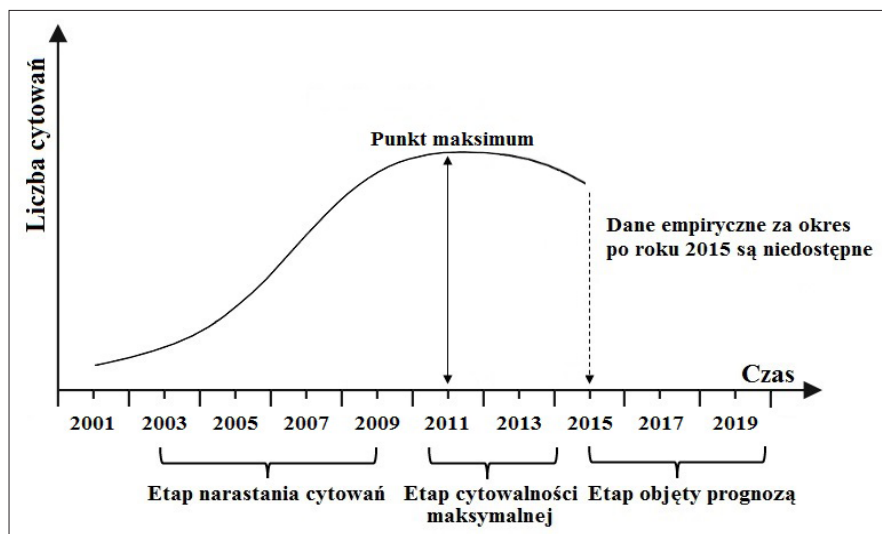
### 4. Metoda badań

Metoda analizy szeregów czasowych jest bliższa naukom społecznym i ekonomicznym (np. socjologii, ekonometrii, demografii) niż naukom ścisłym, w przeciwieństwie do większości streszczonych powyżej algorytmicznych metod tworzenia klastrów publikacji wzajemnie powiązanych. Pod nazwą „analiza szeregów czasowych” kryje się cały zespół różnych metod analizy dynamiki zjawisk masowych, których wspólnym celem jest przewidywanie przyszłego zachowania się zjawisk zachodzących w czasie przy pomocy identyfikacji głównych trendów rozwojowych rządzących zdarzeniami, ich sezonowych lub okresowych odchyłeń od trendu i oceny niepewności towarzyszącej prognozie. Są to przede wszystkim tzw. metody indeksowe, metody wygładzania wykładniczego oraz metody mechanicznego i analitycznego wyodrębniania trendu. Dekompozycja szeregu czasowego, która należy do analitycznych metod wyodrębniania trendu i zostanie zastosowana w ramach niniejszego opracowania, polega na sformułowaniu równania opisującego ilościowo przebieg zjawiska z rozróżnieniem trendu głównego, składowej wahań cyklicznych lub sezonowych oraz składowej odpowiadającej wahaniom przypadkowym (zob. np. Aczel, 2007, 621–667; Kildiszew & Frenkel, 1976; Krawiec, 2014; Łapkowska-Baster, 2009, 72–92; Sobczyk, 2015, 307–356; Sobczyk, 2008, 35–87).

W niniejszym opracowaniu pominięto wyznaczanie składowej wahań cyklicznych oraz składowej wahań okresowych. Wahania okresowe wywołują regularny wzrost lub spadek wartości badanego zjawiska i występują w miarowych odstępach czasu, np. dobowych, tygodniowych lub miesięcznych. Przykładem może być tu dobowa zmienność w zużyciu energii elektrycznej. Wahania sezonowe charakteryzują się natomiast zawsze cyklem rocznym i wywołują równie regularną sezonowość zjawiska. Przykładem wahań sezonowych jest np. kwartalny wzrost sprzedaży owoców (Aczel, 2007, 629–632; Krzysztofiak & Luszniwicz, 1976, 311; Łapkowska-Baster, 2009, 72). Wydaje się w związku z tym, że w przypadku, gdy rozpatrywanym zjawiskiem jest cykl życia publikacji naukowych, nie może być mowy o tak uregulowanych czynnikach wpływających na poziom cytawalności. Za przyjęciem tego założenia przemawia fakt, że autorzy niniejszej pracy nie natrafili dotąd na żadną publikację, w której pojawiałyby się stwierdzenia czy wnioski sygnalizujące istnienie jakichkolwiek zjawisk wywołujących okresowość lub sezonowość w obrębie cykli życiowych dokumentów naukowych. Z tego powodu w niniejszej pracy uwzględniono jedynie wahania przypadkowe, które są spowodowane działaniem czynników losowych, występują nieregularnie z różnym nasileniem i z reguły nie są znane ich przyczyny (Sobczyk, 2015, 355; Zeliaś et al., 2013, 50–51). Uznaje się też, że uwidaczniają się one we wszystkich rodzajach szeregów czasowych (Sobczyk, 2008, 40).



W ramach niniejszego artykułu przyjęto definicję szeregu czasowego, zgodnie z którą w kontekście cykli życiowych publikacji naukowych jest nim ciąg liczb cytowań publikacji o różnym wieku, które były cytowane w jednym, wybranym roku. W związku z tym podstawą proponowanej w ramach niniejszej pracy metody prognozowania jest obserwacja tzw. synchronicznego (retrospektywnego) cyklu życiowego publikacji naukowych (nt. definicji bibliometrycznej metody synchronicznej i alternatywnej wobec niej metody diachronicznej zob. np. Burrell, 2001, 3; Diodato, 1994, 122–123; Diodato & Smith, 1993, 101; Glänzel, 2004, 512–513; Stinson & Lancaster, 1987, 65–67). Typowy diachroniczny cykl życiowy publikacji składa się z etapu narastania jej cytawalności, osiągnięcia momentu cytawalności maksymalnej oraz następującego po punkcie maksimum spadku cytawalności. Na przedstawionym poniżej rysunku 1 zobrazowano hipotetyczny (i nieco wyidealizowany) przebieg synchronicznego odpowiednika diachronicznego cyklu życiowego pewnego zbioru publikacji, z których wszystkie były cytowane – w różnej liczbie – w 2015 r. Publikacje wydane w latach 2011–2015 były cytowane najczęściej i uformowały one ten fragment krzywej cytowań, który w modelu synchronicznym odpowiada diachronicznej fazie narastania cytawalności. Publikacje wydane w 2011 r. były cytowane z największym nasileniem, w związku z czym można stwierdzić, że znajdują się one w centrum synchronicznego punktu maksimum krzywej. Publikacje wydane przed rokiem 2011 (a więc w latach 2010–2001 i wcześniej) są cytowane (w 2015 r.) coraz rzadziej i w coraz mniejszej liczbie w miarę przesuwania się na osi czasu ku jej przecięciu z pionową osią współrzędnych. Etap objęty prognozą rozpoczyna się natomiast po 2015 r. – w praktyce jego początek to punkt osi czasu oznaczający 2016 r. – ponieważ źródłowym materiałem badawczym autorów niniejszego artykułu były prace wydane w 2015 r. Jakiegokolwiek źródła bibliograficzne ogłoszone po tym czasie z naturalnych przyczyn nie mogły więc już figurować w bibliografiach załącznikowych badanych publikacji (zob. Rys. 1; zob. też Aneks 3).



Rys.1. Synchroniczny cykl życiowy zbioru hipotetycznych publikacji naukowych

Prognozy liczb cytowań, które odnoszą się do okresu po 2015 r. wymagają w związku z tym odpowiedniej interpretacji, którą autorzy postarali się przedstawić w szóstej sekcji niniejszego artykułu.

Wyjątki od tego wzorca, który określono wyżej jako typowy dla przeciętnych i występujących najliczniej publikacji, mogą stanowić np. publikacje klasyczne (tzw. *citation classics*), które z biegiem czasu są cytowane coraz wyżej i w ogóle nie osiągają punktu maksimum, publikacje cytowane bardzo rzadko lub wcale, albo tzw. publikacje efemeryczne (ang. *flash in the pan*). Ten ostatni rodzaj to prace, które stają się wysoko cytowane natychmiast po opublikowaniu, ale bardzo szybko wytracają swój impet, a ich cytowalność spada niemal do zera (zob. np. Costas et al., 2010, 331; Garfield, 1985, 193–194; Levitt & Thelwall, 2008, 42; Van Dalen & Henkens, 2005, 211). Są to jednak raczej przypadki odosobnione, a zdecydowana większość publikacji zachowuje się pod względem swojej cytowalności tak, jak przykładowy zbiór prac naukowych z rysunku 1. Inaczej mówiąc, w dziedzinie lub w dyscyplinie naukowej, która znajduje się na etapie tzw. nauki normalnej w sensie teorii Thomasa Kuhna (2001), ma się na ogół do czynienia z publikacjami cytowanymi na przeciętnym poziomie, które tworzą gros dziedzinowej/dyscyplinarnej literatury (zob. np. Redner, 1998, 134; Seglen, 1992, 629–630, 635; Shibata et al., 2008, 762–763). Przez „naukę normalną” rozumie się tutaj taki etap jej rozwoju, na którym nie dokonują się akurat żadne „przewroty kopernikańskie”, nie przewiduje się żadnych nadchodzących przełomów, zmian paradygmatu ani naukowych rewolucji. Założenie, że wybrana do badania polskojęzyczna sfera nauk o Ziemi podlega aktualnie prawidłowości opisywanej terminem „nauka normalna” jest konieczne w kontekście przyjęcia synchronicznej metodologii badań cyklu życiowego losowej próby publikacji. Ponieważ w ramach niniejszej pracy autorzy dysponują wyłącznie danymi o roczniku 2015 wyselekcjonowanych czasopism (zob. Aneks 1), obraz synchronicznego cyklu życiowego publikacji, zbudowany na podstawie tych danych, jest w pewnym sensie statyczny. To znaczy, że obraz ten odzwierciedla stan dyscypliny taki, z jakim mamy do czynienia w tym właśnie (2015) roku. Prognoza przyszłego rozwoju dyscypliny będzie w związku z tym opisywać przebieg tego rozwoju w warunkach, jakie panowały w zbadanej gałęzi nauki w roku 2015, będzie przez warunki te podyktowana i nimi uwarunkowana.

Ogólnie rzecz biorąc trendy rozwojowe zjawisk można podzielić na liniowe (tzw. trendy addytywne) i nieliniowe (tzw. trendy multiplikatywne). Wśród trendów nieliniowych wyróżnia się trendy wykładnicze, potęgowe, wielomianowe, hiperboliczne i logistyczne (De Gooijer & Hyndman, 2006, 444; Krzysztofiak & Luszniwicz, 1976, 364–366; Łapkowska-Baster, 2009, 73; Sobczyk, 2008, 57–58; Sobczyk, 2015, 345–349). W przypadku cyklu życia zbioru dyscyplinarnych publikacji, w którym występuje tendencja rosnąco-malejąca (zob. Rys. 1; Aneks 2; Aneks 3), zastosowanie znajdują funkcje wielomianowe, a zwłaszcza wielomiany drugiego stopnia, czyli parabole (Krzysztofiak & Luszniwicz, 1976, 365). Należy zaznaczyć, że w literaturze przedmiotu daje się zauważyć polaryzację poglądów na temat oceny przydatności stosowania trendu wielomianowego w zagadnieniach predykcyjnych. Ma ona zarówno swoich zwolenników, jak i przeciwników (szerzej na ten temat zob. np. Purczyński, 2010; Purczyński, 2014, 120–121; zob. też: Żurowska, 2005).

W ramach niniejszej pracy zastosowano analizę szeregów czasowych do publikacji cytowanych w obrębie polskojęzycznej literatury czasopiśmienniczej z pogranicza nauk o Ziemi i nauk biologicznych (ekologii, ochrony środowiska, botaniki, mikrobiologii). Do badań

wybrano czasopisma figurujące w części „B” wykazu MNiSW za 2015 r., które otrzymały pięć lub więcej punktów MNiSW (zob. Aneks 1). Zastosowano ponadto ograniczenie dla zebranych danych o cytowaniach literatury przywoływanej na łamach czasopism wyszczególnionych w Aneksie 1. Dane empiryczne zawężono mianowicie do wyłącznie tych, które obejmują lata 2004–2014 z uwagi na fakt, że wykorzystana w ramach prezentowanego modelu parabola jest funkcją symetryczną w stosunku do swojego wierzchołka (tzw. ogniska paraboli). Inaczej mówiąc, modelowanie krzywej cytowań za pomocą paraboli nie powinno uwzględniać całego etapu narastania cytowalności publikacji (zob. Rys. 1; Aneks 3), a w szczególności jego wczesnego stadium, w którym cytowania są nieliczne. Wydaje się, że aby modelowanie funkcją paraboliczną odpowiadało rzeczywistemu przebiegowi zjawiska kumulowania się cytowań, należy ograniczyć zasięg danych wejściowych tylko do tego etapu cyklu, który zachowuje się jak typowa parabola, tj. cechuje się najwyraźniej przebiegiem rosnąco-malejącym (zob. Aneks 3).

## 5. Wyniki badań

Ogólna postać paraboli w opisie tendencji rozwojowej zjawisk ilościowych wyraża się następującym równaniem (Sobczyk, 2008, 57):

$$y_t = a_0 + (a_1 \times t) + (a_2 \times t^2) + \varepsilon_t \quad (a_2 \neq 0, t = 1, 2, 3, \dots, n),$$

gdzie „ $a_0$ ”, „ $a_1$ ” i „ $a_2$ ” to nieznane (szacowane na podstawie danych empirycznych) parametry strukturalne paraboli, „ $n$ ” jest długością szeregu czasowego, „ $\varepsilon_t$ ” jest tzw. składnikiem losowym, zaś „ $y_t$ ” to wartość zjawiska (w tym przypadku liczba cytowań zbioru publikacji) w momencie „ $t$ ”. Zmienną „ $y_t$ ” można też określić jako analityczną postać trendu. Wartość „ $n$ ” w rozpatrywanym przypadku wynosi 11 (szereg liczb cytowań publikacji z lat 2004–2014 zawiera 11 wyrazów). Prognozę punktową uzyskuje się podstawiając do wzoru, z oszacowanymi wartościami „ $a_0$ ”, „ $a_1$ ” i „ $a_2$ ”, wartości „ $t$ ” równe „ $n+1$ ”, „ $n+2$ ” itd., czyli wartości momentów czasu, dla których nie dysponuje się już danymi empirycznymi. Parametry strukturalne paraboli wyznacza się na podstawie następującego równania macierzowego<sup>2</sup> oraz przy przyjęciu takiej numeracji jednostek czasu „ $t$ ”, która spełnia warunek  $\sum t = 0$  (Sobczyk, 2008, 57–60):

$$a = (T^T T)^{-1} \times T^T \times y, \text{ gdzie: } T = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2^2 \\ \vdots & \vdots & \vdots \\ 1 & n & n^2 \end{bmatrix}, y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, a = \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix}$$

<sup>2</sup> Podstawy rachunku macierzy, takie jak mnożenie macierzy przez liczbę i przez macierz, transpozycję macierzy czy macierze odwrotne omówiono np. w podręczniku Tadeusza Trajdosa (2013) lub Włodzimierza Szkutnika i Agnieszki Przybylskiej-Mazur (2008). W Internecie dostępne są ponadto kalkulatory wykonujące działania na macierzach. Jeden z nich zamieszczony jest np. pod adresem: <http://obliczone.pl/start/kalkulatory/924-mnozenie-macierzy-kalkulator>

Korzystając z danych zamieszczonych w Aneksie 2 uzyskujemy następujące zależności:

$$y = \begin{bmatrix} 150 \\ 143 \\ 152 \\ 166 \\ 169 \\ 175 \\ 204 \\ 179 \\ 170 \\ 168 \\ 152 \end{bmatrix}, T = \begin{bmatrix} 1 & -5 & 25 \\ 1 & -4 & 16 \\ 1 & -3 & 9 \\ 1 & -2 & 4 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \\ 1 & 5 & 25 \end{bmatrix}$$

$$T^T T = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -5 & -4 & -3 & -2 & -1 & 0 & 1 & 2 & 3 & 4 & 5 \\ 25 & 16 & 9 & 4 & 1 & 0 & 1 & 4 & 9 & 16 & 25 \end{bmatrix} \times \begin{bmatrix} 1 & -5 & 25 \\ 1 & -4 & 16 \\ 1 & -3 & 9 \\ 1 & -2 & 4 \\ 1 & -1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 4 & 16 \\ 1 & 5 & 25 \end{bmatrix} \times \begin{bmatrix} 11 & 0 & 110 \\ 0 & 110 & 0 \\ 110 & 0 & 1958 \end{bmatrix}$$

Wyznacznik macierzy „ $T^T T$ ”, oznaczany symbolem „ $\det(T^T T)$ ”, jest w tym przypadku równy 1038180. Macierz odwrotna do macierzy „ $T^T T$ ”, oznaczana symbolem „ $(T^T T)^{-1}$ ” jest iloczynem odwrotności wyznacznika macierzy „ $T^T T$ ” oraz transponowanej macierzy dopełnień algebraicznych dla macierzy „ $T^T T$ ”. Macierz odwrotna do „ $(T^T T)$ ” ma więc postać:

$$(T^T T)^{-1} = \frac{1}{1038180} \times \begin{bmatrix} 215380 & 0 & -12100 \\ 0 & 9438 & 0 \\ -12100 & 0 & 1210 \end{bmatrix} = \begin{bmatrix} 0.20746 & 0 & -0.01166 \\ 0 & 0.00909 & 0 \\ -0.01166 & 0 & 0.001166 \end{bmatrix}$$

Iloczyn macierzy  $T^T \times y$  wynosi:

$$T^T y = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -5 & -4 & -3 & -2 & -1 & 0 & 1 & 2 & 3 & 4 & 5 \\ 25 & 16 & 9 & 4 & 1 & 0 & 1 & 4 & 9 & 16 & 25 \end{bmatrix} \times \begin{bmatrix} 150 \\ 143 \\ 152 \\ 166 \\ 169 \\ 175 \\ 204 \\ 179 \\ 170 \\ 168 \\ 152 \end{bmatrix} = \begin{bmatrix} 1828 \\ 225 \\ 17177 \end{bmatrix}$$

Wektor ocen parametrów strukturalnych paraboli ma wobec tego postać:

$$a = (T^T T)^{-1} \times T^T y = \begin{bmatrix} 0.20746 & 0 & -0.01166 \\ 0 & 0.00909 & 0 \\ -0.01166 & 0 & 0.001166 \end{bmatrix} \times \begin{bmatrix} 1828 \\ 225 \\ 17177 \end{bmatrix} = \begin{bmatrix} 178.95 \\ 2.045 \\ -1.286 \end{bmatrix}$$

Parabola wyraża się zatem następującym wzorem:

$$y_t = 178.95 + (2.045 \times t) - (1.286 \times t^2) + \varepsilon_t, \text{ przy zachowaniu warunku } \sum t = 0.$$

Precyzję przybliżenia tendencji rozwojowej badanego zjawiska (w tym przypadku procesu zmian cytowalności publikacji) za pomocą otrzymanej funkcji, szacuje się za pomocą tzw. współczynnika zmienności resztowej ( $V_e$ ), współczynnika zbieżności ( $\varphi^2$ ), współczynnika determinacji ( $R^2$ ) oraz średniego błędu oceny parametrów strukturalnych ( $D(a_0)$ ,  $D(a_1)$  i  $D(a_2)$ ) (Krzysztofiak & Luszniewicz, 1976, 362–369; Sobczyk, 2008, 45–47, 61–64; Sobczyk, 2015, 339–342; Zeliaś et al., 2013, 78–81). Współczynnik  $V_e$  informuje, jaką część średniego poziomu analizowanego zjawiska są przypadkowe odchylenia danych od teoretycznej funkcji trendu (paraboli). Parabola tym bliżej odpowiada danym doświadczalnym, im wartość  $V_e$  jest bliższa zeru. Współczynnik  $\varphi^2$  określa, jaka część ogólnej zmienności wartości „y” (w tym przypadku zaobserwowanych liczb cytowań) nie została wyjaśniona przez uzyskaną funkcję trendu. Współczynnik  $R^2$  – przeciwnie, wskazuje jaka część zmienności y została wyjaśniona przez funkcję trendu (wartość  $R^2 = 1 - \varphi^2$ ). Średnie błędy szacunku parametrów strukturalnych mówią z kolei o tym, o ile średnio mylimy się szacując te parametry w oparciu o wyniki próby losowej (czyli zebrane dane empiryczne). Im mniejsze błędy szacunku, tym lepiej funkcja opisuje zebrane dane. Przyjmuje się, że błędy przekraczające 50% przekreślają wartość poznawczą otrzymanych ocen. Wartości liczbowe tych błędów zapisuje się w nawiasach pod odpowiednimi parametrami strukturalnymi równania, które jest w tym przypadku równaniem paraboli (Sobczyk, 2008, 46–47).

Tabela 1. Wyznaczone wartości współczynników określających stopień precyzji dopasowania paraboli do danych doświadczalnych

| L.p. | Nazwa współczynnika  | Wartość współczynnika |
|------|--|-----------------------|
| 1    | Współczynnik zmienności resztowej ( $V_e$ )                            | 7.34%                 |
| 2    | Współczynnik zbieżności ( $\varphi^2$ )                                | 0.41 (41%)            |
| 3    | Współczynnik determinacji ( $R^2$ )                                    | 0.59 (59%)            |
| 4    | Średni błąd oceny parametru strukturalnego $a_0$ paraboli ( $D(a_0)$ ) | 5.55                  |
| 5    | Średni błąd oceny parametru strukturalnego $a_1$ paraboli ( $D(a_1)$ ) | 1.16                  |
| 6    | Średni błąd oceny parametru strukturalnego $a_2$ paraboli ( $D(a_2)$ ) | 0.42                  |
| 7    | Odchylenie standardowe składnika resztowego paraboli ( $S_e$ )         | 12.19                 |

Aby uniknąć przedstawiania wszystkich obliczeń, których również dokonuje się w zapisie macierzowym i które zajęłyby dużo miejsca, poniżej podano jedynie wartości wskaźników

wyliczone w oparciu o dane empiryczne<sup>3</sup> przedstawione w Aneksach 2 i 3. Przejrzystym sposobem prezentacji wydaje się zebranie ich w tabeli (zob. Tab. 1).

Krótkoterminowa prognoza punktowa dla lat  $T=2015$ ,  $T=2016$ ,  $T=2017$ ,  $T=2018$  i  $T=2019$  jest skonstruowana przy użyciu wskazanego powyżej równania paraboli. Np. dla roku 2015, do równania podstawia się w miejsce zmiennej „ $t$ ” wartość „6”, ponieważ rok 2015 jest rokiem szóstym z kolei, jeżeli zachowany jest warunek „ $\sum t = 0$ ”. Bezwzględny błąd *ex ante* prognozy to z kolei oczekiwana różnica między wartością prognozowaną a rzeczywistą realizacją zmiennej prognozowanej (wartością empiryczną zjawiska lub pewnej jego cechy, którą się prognozuje, w okresie nadchodzącym, tj. takim, za który nie dysponuje się już danymi doświadczalnymi). O błędzie „bezwzględnym” mówi się kiedy jednostka, w jakiej wyraża się błąd, jest tą samą jednostką, w której podaje się wartości zjawiska (w tym przypadku jest nią liczba cytowań). O błędzie „względnym” – kiedy jego wartość jest niemianowana (procentowa). Wartość błędu względnego pozwala ocenić, czy prognoza jest dopuszczalna. Jeżeli błąd względny przekracza 10% prognozę kwalifikuje się na ogół jako niedopuszczalną, chociaż przedział ten (0% – 10%) może być uznany za relatywny. Duże znaczenie ma tutaj bowiem rodzaj, znaczenie i interpretacja zmiennej prognozowanej (Krawiec, 2014, 10; Sobczyk, 2008, 27–28; Zeliaś, 1997, 71–72; Zeliaś et al., 2013, 49–50). Na podstawie odpowiednich wzorów (Sobczyk, 2008, 48–49, 58, 63) uzyskano następujące wartości prognozowane oraz wartości błędu *ex ante*, jakim są one obarczone (zob. Tab. 2).

Tab. 2. Przewidywane liczby cytowań dla lat z przedziału 2015–2019 oraz szacunek ich niepewności

| Rok, dla którego obowiązuje prognoza | Wartość „ $t$ ” | Prognozowana wartość liczby cytowań publikacji | Bezwzględny błąd <i>ex ante</i> prognozy punktowej | Względny błąd prognozy punktowej | Empirycznie stwierdzona wartość liczby cytowań publikacji (tylko rok 2015) |
|--------------------------------------|-----------------|--|--|----------------------------------|--|
| 2015                                 | 6               | 145 (+/ – 12)                                  | 18.1   | 12.49%                           | 109  |
| 2016                                 | 7               | 134 (+/ – 12)                                  | 22.18  | 16.55%                           |  |
| 2017                                 | 8               | 113 (+/ – 12)                                  | 27.46  | 24.3%                            |  |
| 2018                                 | 9               | 93 (+/ – 12)                                   | 33.84  | 36.38%                           |  |
| 2019                                 | 10              | 71 (+/ – 12)                                   | 41.24  | 58.08%                           |  |

Jak wynika z tabeli 2, tylko prognozy na lata 2015–2016 wydają się akceptowalne przy założeniu, że można tu zastosować nieco wyższy (od dziesięcioprocentowego) próg dopuszczalności. Przyczyną takiego posunięcia jest przede wszystkim fakt istnienia pewnej płynności granicy dopuszczalności prognoz, tj. uwarunkowanie jej konkretnym kontekstem praktycznym i zależnością wymaganej precyzji prognozy od indywidualnych potrzeb badacza i charakteru objaśnianych zmiennych (Sobczyk, 2008, 27–28; Zeliaś, 1997,

<sup>3</sup> Na życzenie osób zainteresowanych szczegółowe wyliczenia mogą zostać przesłane na wskazany adres mailowy po skontaktowaniu się z pierwszym z autorów niniejszego artykułu. Ten sam komentarz odnosi się również do przedstawionych poniżej wartości wskaźników oceny dokładności prognoz.



71–72; Zeliaś et al., 2013, 49–50). W ramach teorii zachowań w zakresie cytowań jako determinantów procesów decyzyjnych leżących u podstaw samego zjawiska wstawiania przypisów bibliograficznych często podnosi się kwestie z pogranicza psychologii i nauk społecznych (zob. np. Bornmann, Daniel, 2008), co przybliża niektóre aspekty bibliometrycznego prognozowania cytowalności do ogólnego obszaru nauk społecznych, w których osiągalny poziom dokładności szacunków i statystyk wydaje się być z konieczności niższy niż w naukach ekonometrycznych, ścisłych i przyrodniczych.

Ponadto warto podkreślić, że, ponieważ autorzy dysponują empiryczną wartością liczby cytowań za rok 2015, można stwierdzić, że prognoza jest tylko nieznacznie niedoszacowana. To znaczy, prognozowane 145 cytowań pomniejszone o wartość względnego błędu prognozy ( $145 - (12.5\% \times 145) = 145 - 18.125 = 126.875$ ) i ponownie pomniejszone o szacowaną wartość składnika losowego ( $126.875 - 12.19 = 114.685$ ) daje w wyniku liczbę 115 cytowań, co przekracza wartość empiryczną tylko o sześć cytowań. Należy tu dodatkowo zaznaczyć, że w tym przykładzie niedoszacowanie może wynikać z faktu, że liczba publikacji wydanych i zarazem cytowanych w 2015 r. jest przypuszczalnie nieco zaniżona ze względu na to, że nie wszystkie prace wydane w tym roku były dostępne dla autorów cytujących w trakcie przygotowywania artykułów. Jeżeli np. artykuł powstawał w pierwszych miesiącach 2015 r., dokumenty wydane pod koniec tego roku nie miały już szans na wejście w skład jego bibliografii załącznikowej. Dokładnie z tego samego powodu z zakresu danych wejściowych dla konstrukcji modelu wykluczono liczby cytowań zaobserwowane dla prac wydanych w 2015 r.

Dodatkową ewentualnością, istniejącą w ramach metodologii statystycznej, jest wyznaczenie prognoz przedziałowych zamiast prognoz punktowych. Ponieważ jednak w tym przypadku konieczna jest znajomość rozkładu zmiennej prognozowanej, o którym zazwyczaj zakłada się, że jest on tzw. rozkładem normalnym (rozkładem Gaussa – zob. np. Sobczyk, 2008, 64; Zeliaś et al., 2013, 45), możliwość ta nie zostanie wykorzystana w ramach niniejszego opracowania. Rozkład liczby cytowań publikacji o różnym wieku jest bowiem wyraźnie skośny (zob. Aneks 3) i założenie o normalności takiego rozkładu byłoby nieuprawnione. Zagadnienie zbadania charakteru tego rozkładu, chociaż wydaje się istotne z teoretycznego i praktycznego punktu widzenia, wykracza poza ramy niniejszej pracy i zasługuje na osobne opracowanie. Podobnie rzecz ma się ze zjawiskiem tzw. autokorelacji składnika resztowego (zob. np. Krzysztofiak & Luszniwicz, 1976, 390–391), które – jeżeli istnieje – zakłóca model wahań w czasie i utrudnia prognozowanie. Wymaga ono sformułowania odpowiedniej hipotezy zerowej i zbadania jej przy pomocy specjalnej statystyki testowej<sup>4</sup>, co również jest kwestią pozostającą poza nawiasem głównego celu niniejszej pracy.

## 6. Interpretacja wyników i wnioski

Autorzy niniejszej pracy proponują następującą interpretację wartości prognozowanych liczb cytowań, które uzyskano na podstawie analizy synchronicznej. Przede wszystkim z naturalnych przyczyn wartości te nie mogą być interpretowane jako liczby cytowań pojawiające się w tym samym roku, którego dotyczą zebrane dane doświadczalne (tj. w 2015 r.). W 2015 r. nie można bowiem cytować prac, które dopiero zostaną opublikowane. Z drugiej

<sup>4</sup> Jest to tzw. statystyka Durбина-Watsona.

strony, ostatnią wartością szeregu czasowego dostępną podczas badania była liczba cytowań publikacji wydanych i zarazem cytowanych w 2015 r. W związku z tym wydaje się, że znaczenie wartości prognozowanych (tzn. liczb cytowań dla lat 2015–2019) można wyrazić analogicznie, jako liczby cytowań prac zarazem opublikowanych i cytowanych w latach 2015–2019. Przykładowo, liczba 134 cytowań dla 2016 r. (zob. Tabela 2) oznaczałaby więc, że w 2016 r. w dyscyplinie usytuowanej na pograniczu nauk o Ziemi i nauk biologicznych, zostaną zacytowane 134 publikacje, których rokiem wydania jest ten sam 2016 r. Byłby to wobec tego swego rodzaju uproszczony indeks natychmiastowości dyscypliny (ang. *immediacy index*) określający, jak szybko badacze w tej dyscyplinie nauki reagują na najnowsze informacje i publikacje dyscyplinarne, jak szybko dostrzegają i zapoznają się z najaktualniejszymi pracami i włączają je w nurt własnych dociekań. Im szybciej spadałby tak rozumiany indeks (wartości prognozowane dla kolejnych lat), tym wolniej postępowałby rozwój dyscypliny. Konieczna jest tutaj akceptacja wspomnianego już założenia o pewnej stabilności rozwoju dyscypliny. Innymi słowy, koncepcja indeksu natychmiastowości dałaby się utrzymać przy przyjęciu, że tempo rozwoju w dyscyplinie nie zmienia się dowolnie z roku na rok. Tylko wtedy można byłoby domniemywać, że wartości indeksu uzyskane w oparciu o dane np. za 2015 r. będą obowiązywać przynajmniej w dwóch-trzech kolejnych latach, o ile w dyscyplinie nie nastąpi żaden wyraźny przełom, który wywołałby np. dezaktualizację większości treści opublikowanych (i akceptowanych przez społeczność naukowców) przed 2015 r.

Z powyższej analizy wypływają wnioski, które można zawrzeć w następujących stwierdzeniach.

Po pierwsze, statystyczna metodologia prognostyczna wydaje się znajdować zastosowanie w ramach bibliometrycznej analizy tempa rozwoju dyscyplin naukowych. Świadczą o tym akceptowalne wskaźniki zgodności dopasowania trendu teoretycznego do wartości rzeczywistych (zob. Tab. 1), stwierdzonych doświadczalnie w oparciu o stosunkowo dużą próbę losową.

Po drugie, w przypadku przebadanej próby dopasowanie funkcyjne wielomianu drugiego stopnia (paraboli) do danych doświadczalnych cechowało się wprawdzie dopuszczalnym poziomem dokładności, jednak wartości prognozowane w okresie dłuższym niż dwuletni (tj. dla lat 2017–2019), były obciążone znacznymi błędami.

Po trzecie, z uwagi na powyższy fakt należy uznać za wskazane rozpatrzenie możliwości dalszego dopracowania metody prognostycznej. Mogłoby ono przybrać postać analizy i określenia typu rozkładu zmiennej prognozowanej oraz skonstruowania prognoz przedziałowych dla wybranych lat (tzw. przedziały ufności). Alternatywnie, metoda mogłaby zostać zmodyfikowana poprzez zbadanie większej próby losowej, poprzez wybór próby losowej na podstawie cytowań literatury w innej niż rozpatrzona dyscyplinie czy poprzez wybór innych (dłuższych albo krótszych) zakresów danych wejściowych, jako podstawy sformułowania równania opisującego trend rozwojowy. W odniesieniu do tej ostatniej ewentualności należy jednak nadmienić, że autorzy przetestowali dokładność dopasowania funkcji parabolicznej do kilku szerszych przedziałów czasowych obejmujących zakres danych wejściowych służących sformułowaniu równania paraboli. Testy te wykazały, że przyjęty ostatecznie przedział jedenastoletni (2004–2014) przełożył się na najwyższą precyzję dopasowania, która została zmierzona za pomocą współczynników  $V_e$ ,  $\varphi^2$  i  $R^2$ .

Po czwarte, próby udoskonalenia przyjętej metodologii mogłyby przebiegać również w kierunku przyjęcia innej niż parabola, funkcji spełniającej rolę predykcijną. Ponieważ

jednak typowy charakter cyklu życia publikacji naukowych (zob. Rys. 1) posiada swój specyficzny kształt (przebieg rosnąco-malejący) wydaje się, że inne typowe dla statystycznych analiz trendów rozwojowych funkcje (tj. funkcja liniowa, wykładnicza, potęgowa, logistyczna i hiperboliczna) nie spełniłyby się dobrze w roli predyktorów. Należałoby wobec tego poszukiwać innych prawidłowości funkcyjnych, np. za pomocą specjalistycznego oprogramowania w rodzaju pakietu „Origin” lub „MATLAB”.

Po piąte należy zauważyć, że obiecującą ewentualnością wydaje się w tym kontekście możliwość zastosowania innej, niż przyjęta w ramach niniejszej pracy, organizacji danych empirycznych. To znaczy, że dane te można zagregować w sposób skumulowany wzorem Roberta Burtona i Richarda Keblera (1960) i podjąć próbę dopasowania zupełnie innej funkcji do tak ustrukturyzowanych danych. Kumulacja liczb cytowań (również w ujęciu synchronicznym) oznacza tutaj sumowanie liczb cytowań prac wydanych w kolejnych latach. Np. cytowania prac wydanych w 1950 r., które pojawiły się np. w roku 2015 dodaje się do liczby cytowań prac wydanych w 1951 r. (i cytowanych w tym samym roku 2015). Następnie otrzymaną sumę powiększa się o liczbę cytowań prac ogłoszonych w 1952 r. itd. Funkcja prognostyczna mogłaby w tym ujęciu przewidywać przyszłą liczbę cytowań skumulowanych, z której możliwe byłoby wyznaczenie liczby cytowań nieskumulowanych (tj. dotyczących tylko prac wydanych w jednym, wybranym roku). Funkcja Burtona i Keblera (1960) była jednak bardzo specyficzna, wymagała wstępnego wyznaczenia dwóch parametrów (stałych) dla każdej poszczególnej dyscypliny i w niczym nie przypominała typowych funkcji wykorzystywanych w ramach prognozowania statystycznego. Wydaje się w związku z tym, że konstrukcja takiej funkcji byłaby zadaniem znacznie trudniejszym.

Po szóste, zaproponowany indeks natychmiastowości opisujący tempo rozwoju dyscyplin mógłby znaleźć praktyczne zastosowania przy akceptacji założenia o tzw. normalnym, bieżącym rozwoju wybranych dyscyplin w sensie, w jakim o rozwoju tym wypowiadał się Kuhn (2001). Wydaje się, że indeks ten w szczególności mógłby znaleźć zastosowania porównawcze. Jego charakter i jednostka, w jakiej wyrażane są jego wartości sprawiają bowiem, że jest on predysponowany do zastosowań porównawczych w większym stopniu, niż do zastosowań polegających na ocenie tempa rozwoju indywidualnych dyscyplin (lub obszarów problemowych) funkcjonujących w obrębie wybranych dziedzin współczesnej nauki.

Po siódme, w przypadkach, w których do danych doświadczalnych, cechujących się dużą nieregularnością, nie udałoby się zadowalająco dopasować żadnej funkcji matematycznej (takiej jak np. parabola lub funkcja wielomianowa stopnia wyższego niż drugi), możliwe byłoby wykorzystanie prognostycznych metod indeksowych lub tzw. mechanicznych metod dekompozycji szeregów czasowych. Polegają one na badaniu średniego tempa zmian poziomu zjawiska w czasie przy użyciu m. in. średniej geometrycznej, średniej ruchomej („kroczącej”) prostej lub scentrowanej, wag harmonicznych albo prostego bądź adaptacyjnego wygładzania (wyrównywania) wykładniczego. Modele wygładzania adaptacyjnego to przede wszystkim wygładzanie Browna, Holta, Wintersa oraz model tzw. trendu pełzającego, zwanego też segmentowym (zob. np. Aczel, 2007, 645, 647; De Gooijer & Hyndman, 2006; Gelper et al., 2007, 285–286; Hirschey, 2009, 219–222; Krawiec, 2014; Snarska, 2011, 212–244; Sobczyk, 2008, 97–98, 138; Yaffee & McGee, 2000, 23–44; Zeliaś et al., 2013, 140–153).

## Aneks 1

Wykaz objętych badaniami czasopism wydawanych w języku polskim, należących do zakresu z pogranicza nauk o Ziemi i nauk biologicznych (ekologii, ochrony środowiska, botaniki, mikrobiologii), występujących w części „B” wykazu czasopism naukowych MNiSW z 2015 roku. Online: Figshare repository, [21.12.2017], <https://figshare.com/s/5a746333fb21f9e9a4e3>

| L.p. | Tytuł   | ISSN      | Punkty | Tematyka*  | Dziedzina                         | Język publikacji            |
|------|---|-----------|--------|--|-----------------------------------|-----------------------------|
| 1.   | Soil Science Annual   | 2300–4967 | 13     | Gleboznawstwo  | Nauki o Ziemi / nauki biologiczne | Angielski, polski           |
| 2.   | Acta Scientiarum Polonorum Formatio Circumietus   | 1644–0765 | 10     | Kształtowanie środowiska, ekologia, geologia, hydrologia | Nauki o Ziemi / nauki biologiczne | Polski, angielski           |
| 3.   | Gospodarka Wodna  | 0017–2448 | 9      | Hydrologia, hydrogeologia, ochrona środowiska            | Nauki o Ziemi / nauki biologiczne | Polski                      |
| 4.   | Landform Analysis   | 1429–799X | 8      | Geomorfologia  | Nauki o Ziemi / nauki biologiczne | Polski, angielski           |
| 5.   | Problemy Ekologii Krajobrazu  | 1899–3850 | 8      | Geoekologia, geografia fizyczna                          | Nauki o Ziemi / nauki biologiczne | Polski, angielski           |
| 6.   | Przegląd Przyrodniczy   | 1230–509X | 6      | Ochrona przyrody, fizjografia                            | Nauki o Ziemi / nauki biologiczne | Polski                      |
| 7.   | Zeszyty Naukowe Południowo-Wschodniego Oddziału Polskiego Towarzystwa Inżynierii Ekologicznej z siedzibą w Rzeszowie i Polskiego Towarzystwa Gleboznawczego Oddział w Rzeszowie | 1642–3828 | 6      | Gleboznawstwo, ekologia                                  | Nauki o Ziemi / nauki biologiczne | Polski                      |
| 8.   | Acta Geographica Silesiana  | 1897–5100 | 5      | Geografia, ekologia                                      | Nauki o Ziemi / nauki biologiczne | Polski, rosyjski, angielski |
| 9.   | Rocznik Świętokrzyski seria B – Nauki Przyrodnicze  | 1427–5929 | 5      | Geografia, biologia, ekologia                            | Nauki o Ziemi / nauki biologiczne | Polski                      |
| 10.  | Studia Limnologica et Telmatologica   | 1897–645X | 5      | Ekologia, geografia, geologia                            | Nauki o Ziemi / nauki biologiczne | Polski, angielski           |

\* Tematyka poruszana na łamach poszczególnych czasopism została tu wskazana na podstawie oceny autora niniejszej pracy dokonanej w oparciu o deklaracje redakcji czasopism, treści publikowanych w nich artykułów i afiliacje autorów artykułów, kategorie tematyczne wyszczególnione dla każdego z tytułów w Polskiej Bibliografii Naukowej (<https://pbn.nauka.gov.pl/>), jak również w odniesieniu do Rozporządzenia MNiSW z dnia 8 sierpnia 2011 roku w sprawie obszarów nauki, dziedzin nauki i sztuki oraz dyscyplin naukowych i artystycznych.

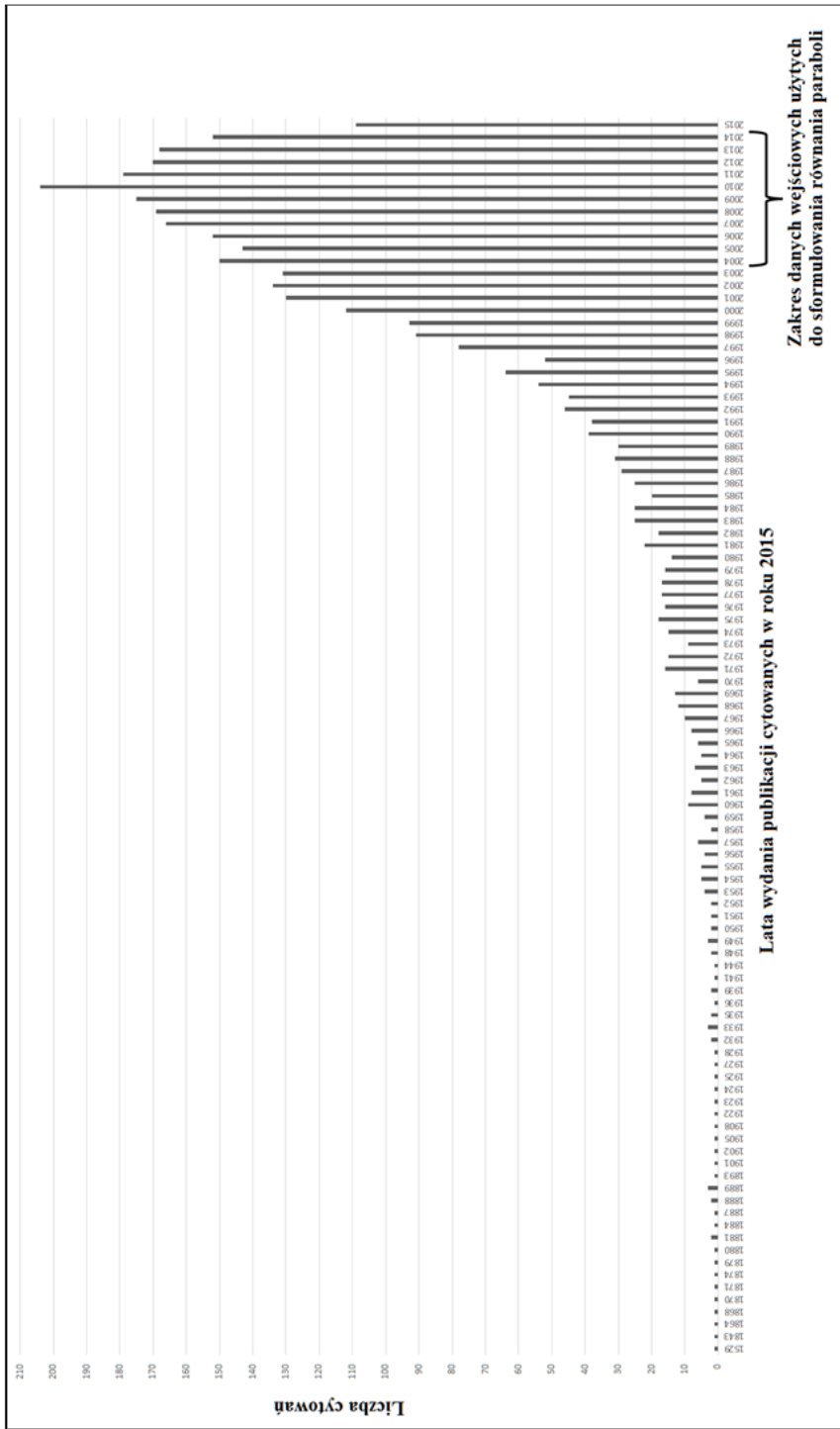
## Aneks 2

Cytowania publikacji naukowych, o wszelkich formach wydawniczych, opublikowanych w 2015 r. i latach uprzednich, zaobserwowane w badaniu empirycznym rocznika 2015 czasopism wymienionych w Aneksie 1. Online: Figshare repository, [21.12.2017], <https://figshare.com/s/ad0defa403c7dd9300bc>

| Rok wydania | Liczba cytowań | Rok wydania | Liczba cytowań | Rok wydania | Liczba cytowań | Rok wydania | Liczba cytowań |
|-------------|----------------|-------------|----------------|-------------|----------------|-------------|----------------|
| 2015        | 109            | 1987        | 29             | 1959        | 4              | 1901        | 1              |
| 2014        | 152            | 1986        | 25             | 1958        | 2              | 1893        | 1              |
| 2013        | 168            | 1985        | 20             | 1957        | 6              | 1889        | 3              |
| 2012        | 170            | 1984        | 25             | 1956        | 4              | 1888        | 2              |
| 2011        | 179            | 1983        | 25             | 1955        | 5              | 1887        | 1              |
| 2010        | 204            | 1982        | 18             | 1954        | 5              | 1884        | 1              |
| 2009        | 175            | 1981        | 22             | 1953        | 4              | 1881        | 2              |
| 2008        | 169            | 1980        | 14             | 1952        | 2              | 1880        | 1              |
| 2007        | 166            | 1979        | 16             | 1951        | 2              | 1879        | 1              |
| 2006        | 152            | 1978        | 17             | 1950        | 2              | 1874        | 1              |
| 2005        | 143            | 1977        | 17             | 1949        | 3              | 1871        | 1              |
| 2004        | 150            | 1976        | 16             | 1948        | 2              | 1870        | 1              |
| 2003        | 131            | 1975        | 18             | 1944        | 1              | 1868        | 1              |
| 2002        | 134            | 1974        | 15             | 1941        | 1              | 1864        | 1              |
| 2001        | 130            | 1973        | 9              | 1939        | 2              | 1843        | 1              |
| 2000        | 112            | 1972        | 15             | 1936        | 1              | 1529        | 1              |
| 1999        | 93             | 1971        | 16             | 1935        | 2              |             |                |
| 1998        | 91             | 1970        | 6              | 1933        | 3              |             |                |
| 1997        | 78             | 1969        | 13             | 1932        | 2              |             |                |
| 1996        | 52             | 1968        | 12             | 1928        | 1              |             |                |
| 1995        | 64             | 1967        | 10             | 1927        | 1              |             |                |
| 1994        | 54             | 1966        | 8              | 1925        | 1              |             |                |
| 1993        | 45             | 1965        | 6              | 1924        | 1              |             |                |
| 1992        | 46             | 1964        | 5              | 1923        | 1              |             |                |
| 1991        | 38             | 1963        | 7              | 1922        | 1              |             |                |
| 1990        | 39             | 1962        | 5              | 1908        | 1              |             |                |
| 1989        | 30             | 1961        | 8              | 1905        | 1              |             |                |
| 1988        | 31             | 1960        | 9              | 1902        | 1              |             |                |

**Aneks 3**

Graficzna prezentacja rozkładu cytowań publikacji zaobserwowanego w badaniu empirycznym rocznika 2015 czasopism wymienionych w Aneksie 1 (dane z Aneksu 2). Online: Figshare repository, [21.12.2017], <https://figshare.com/s/0a620b9e3d37f480623b>





## Bibliografia

- Aczel, A.D. (2007). *Statystyka w zarządzaniu*. Warszawa: PWN.
- Bornmann, L., Daniel, H.-D. (2008). What Do Citation Counts Measure? A Review of Studies on Citing Behavior. *Journal of Documentation*, 64(1), 45–80.
- Boyack, K.W., Klavans, R. (2014). Creation of Highly Detailed, Dynamic, Global Model and Map of Science. *Journal of the Association for Information Science and Technology*, 65(4), 670–685.
- Braun, T., Schubert, A., Kostoff, R. (2000). Growth and Trends of Fullerene Research as Reflected in Its Journal Literature. *Chemical Reviews*, 100(1), 23–37.
- Burrell, Q. (2001). Stochastic Modelling of the First-Citation Distribution. *Scientometrics*, 52(1), 3–12.
- Burton, R.E., Kebler, R.W. (1960). The 'Half-Life' of Some Scientific and Technical Literatures. *American Documentation*, 11(1), 18–22.
- Chen, Ch. (2006). CiteSpace II: Detecting and Visualizing Emerging Trends and Transient Patterns in Scientific Literature. *Journal of the American Society for Information Science and Technology*, 57(3), 359–377.
- Costas, R., van Leeuwen, T.N., van Raan, A.F.J. (2010). Is Scientific Literature Subject to a 'Sell-By-Date'? A General Methodology to Analyze the 'Durability' of Scientific Documents. *Journal of the American Society for Information Science and Technology*, 61(2), 329–339.
- De Gooijer, J.G., Hyndman, R.J. (2006). 25 Years of Time Series Forecasting. *International Journal of Forecasting*, 22(3), 443–473.
- Diodato, V. (1994). *Dictionary of Bibliometrics*. New York: London: Norwood: The Haworth Press.
- Diodato, V., Smith, F. (1993). Obsolescence of Music Literature. *Journal of the American Society for Information Science*, 44(2), 101–112.
- Ena, O., Mikova, N., Saritas, O., Sokolova, A. (2016). A Methodology for Technology Trend Monitoring: The Case of Semantic Technologies. *Scientometrics*, 108(3), 1013–1041.
- Garfield, E. (1985). The Articles Most Cited in the SCI from 1961 to 1982. Another 100 Citation Classics: The Watson-Crick Double Helix Has Its Turn [online]. *Essays of an Information Scientist*, 8, 187–196, [21.12.2017], <http://garfield.library.upenn.edu/essays/v8p187y1985.pdf>
- Garfield, E. (2004). Historiographic Mapping of Knowledge Domains Literature. *Journal of Information Science*, 30(2), 119–145.
- Gelper, S., Fried, R., Croux, Ch. (2007). Robust Forecasting with Exponential and Holt–Winters Smoothing. *Journal of Forecasting*, 29(3), 285–300.
- Glänzel, W. (2004). Towards a Model for Diachronous and Synchronous Citation Analyses. *Scientometrics*, 60(3), 511–522.
- Guo, H., Weingart, S., Börner, K. (2011). Mixed-Indicators Model for Identifying Emerging Research Areas. *Scientometrics*, 89(1), 421–435.
- Hirsch, M. (2009). *Fundamentals of Managerial Economics*. Mason, OH: South-Western Cengage Learning.
- Huang, M.-H., Chang, C.-P. (2015). A Comparative Study on Detecting Research Fronts in the Organic Light-Emitting Diode (OLED) Field Using Bibliographic Coupling and Co-Citation. *Scientometrics*, 102(3), 2041–2057.
- Kildiszew, G., Frenkel, A. (1976). *Analiza szeregów czasowych i prognozowanie*. Warszawa: PWE.
- Krawiec, S. (2014). *Adaptacyjne modele wygładzania wykładniczego jako instrumenty prognozowania krótkoterminowego zjawisk ilościowych*. Gliwice: Wydaw. Politechniki Śląskiej.
- Krzysztofak, M., Luszniwicz, A. (1976). *Statystyka*. Warszawa: PWE.
- Kuhn, T.S. (2001). *Struktura rewolucji naukowych*. Warszawa: Fundacja Aletheia.
- Levitt, J.M., Thelwall, M. (2008). Patterns of Annual Citation of Highly Cited Articles and the Prediction of Their Citation Ranking: A Comparison Across Subjects. *Scientometrics*, 77(1), 41–60.

- Lapkowska-Baster, B. (2009). *Miary współzależności i dynamiki zjawisk w statystyce opisowej. Przykłady i zadania*. Kraków: Wydaw. Uniwersytetu Jagiellońskiego.
- Opaliński, Ł. (2017a). Bibliometryczna metodologia prognozowania i oceny rozwoju dyscyplin naukowych. Analiza piśmiennictwa. Część 1: Publikacje pionierskie, metoda powiązań bibliograficznych, metoda współcytowań i metoda współwystępowania specjalistycznej terminologii naukowej. *Zagadnienia Informatyki Naukowej. Studia Informatyczne*, 109(1), 34–65.
- Opaliński, Ł. (2017b). Bibliometryczna metodologia prognozowania i oceny rozwoju dyscyplin naukowych. Analiza piśmiennictwa. Część 2: badania porównawcze, hybrydowe, statystyczne, analizy dokumentów patentowych, ścieżek rozwoju dyscyplin oraz pozostałe oryginalne podejścia metodologiczne. *Zagadnienia Informatyki Naukowej. Studia Informatyczne*, 110(2), 74–106.
- Price, D. de Solla (1967). *Mała Nauka – Wielka Nauka*. Warszawa: PWN.
- Pritchard, A. (1969). Statistical Bibliography or Bibliometrics? *Journal of Documentation*, 25(4), 348–349.
- Purczyński, J. (2010). Wybrane problemy stosowania trendu wielomianowego w prognozowaniu gospodarczym. *Zeszyty Naukowe Uniwersytetu Szczecińskiego. Ekonomiczne Problemy Usług*, 60, 324–339.
- Purczyński, J. (2014). Wybrane aspekty prognozowania z wykorzystaniem klasycznych modeli trendu. *Zeszyty Naukowe Uniwersytetu Szczecińskiego. Studia i Prace Wydziału Nauk Ekonomicznych i Zarządzania*, 36(2), 119–130.
- Redner, S. (1998). How Popular is Your Paper? An Empirical Study of the Citation Distribution. *The European Physical Journal B-Condensed Matter and Complex Systems*, 4(2), 131–134.
- Seglen, P.O. (1992). The Skewness of Science. *Journal of the American Society for Information Science*, 43(9), 628–638.
- Shibata, N., Kajikawa, Y., Takeda, Y., Matsushima, K. (2008). Detecting Emerging Research Fronts Based on Topological Measures in Citation Networks of Scientific Publications. *Technovation*, 28(11), 758–775.
- Skalska-Zlat, M. (1993). *Bibliometryczne badania rozwoju dyscypliny naukowej*. Seria: Bibliotekoznawstwo XVII. Wrocław: Wydaw. UWr.
- Skalska-Zlat, M. (1999). Bibliografia jako odbicie rozwoju nauki. *Zagadnienia Naukoznawstwa*, 1(139), 57–63.
- Small, H., Boyack, K.W., Klavans, R. (2014). Identifying Emerging Topics in Science and Technology. *Research Policy*, 43(8), 1450–1467.
- Snarska, A. (2011). *Statystyka. Ekonometria. Prognozowanie. Ćwiczenia z Excelem 2007*. Warszawa: Placet.
- Sobczyk, M. (2008). *Prognozowanie. Teoria, przykłady, zadania*. Warszawa: Placet.
- Sobczyk, M. (2015). *Statystyka*. Warszawa: PWN.
- Stinson, E.R., Lancaster, F.W. (1987). Synchronous Versus Diachronous Methods in the Measurement of Obsolescence by Citation Studies. *Journal of Information Science*, 13(2), 65–74.
- Szkutnik, W., Przybylska-Mazur, A. (2008). *Algebra liniowa dla studiów ekonomicznych: metody i zadania, algebra macierzy*. Katowice: Wydaw. Uczelniane Akademii Ekonomicznej im. Karola Adamieckiego.
- Takeda, Y., Kajikawa, Y. (2009). Optics: A Bibliometric Approach to Detect Emerging Research Domains and Intellectual Bases. *Scientometrics*, 78(3), 543–558.
- Trajdos, T. (2013). *Matematyka. Cz. 3, Liczby zespolone, wektory, macierze, wyznaczniki, geometria analityczna i różniczkowa*. Warszawa: WNT.
- Tu, Y-N., Hsu, S-L. (2016). Constructing Conceptual Trajectory Maps to Trace the Development of Research Fields. *Journal of the Association for Information Science and Technology*, 67(8), 2016–2031.
- Van Dalen, H.P., Henkens, K. (2005). Signals in Science – On the Importance of Signaling in Gaining Attention in Science. *Scientometrics*, 64(2), 209–233.

- Wang, Ch-Ch., Ho, Y-S. (2016). Research Trend of Metal-Organic Frameworks: A Bibliometric Analysis. *Scientometrics*, 109(1), 481–513.
- Yaffee, R.A., McGee, M. (2000). *Introduction to Time Series Analysis and Forecasting: With Applications of SAS and SPSS*. San Diego: Academic Press.
- Zeliaś, A. (1997). *Teoria prognozy*. Warszawa: PWE.
- Zeliaś, A., Pawełek, B., Wanat, S. (2013). *Prognozowanie ekonomiczne. Teoria, przykłady, zadania*. Warszawa: PWN.
- Żurowska, J. (2005). Zastosowanie modelu trendu wielomianowego do opisu kształtowania się wskaźnika motoryzacji w Polsce w latach 1980–2004. *Transport Miejski i Regionalny*, 12, 15–21.
- 

## Application of Statistical Time Series Analysis in Short-Term Forecasting of the Development of Scientific Disciplines

### Abstract

**Purpose/Thesis:** The aim of this paper is to: describe the basics of statistical time series analysis, investigate the possibility of using this approach in short-term forecasting of the development of scientific disciplines based on data showing the citedness of research papers, assess the uncertainty of the forecasts (feature available with this method) and read the results of the research in terms of science studies.

**Approach/Methods:** The analytical method of time series decomposition was used to isolate so-called development trends based on the observations of the dynamics of mass phenomena. In this case, the mass phenomenon was the citedness of scientific literature in the field of Polish earth science. The method consists of the formulation of an equation which quantitatively describes the course of the phenomenon and distinguishes the main trend from the accidental variability (fluctuations). The equation serves as a keystone in the process of predicting future values of the time series.

**Results and conclusions:** The results confirm the possibility of an effective application of the approach in short-term forecasting of the scientific discipline development rate. In the case of the research in question, the forecasts proved acceptable as regards two-year horizon of the forecast. However, it is assumed that building the forecast in the form of confidence intervals or slightly modifying the approach could contribute to the lengthening of the horizon. The results were read as a simplified immediacy index, available in particular for comparative applications.

**Originality/Value:** The authors were unable to find any examples of research which was conducted in a manner presented in this paper. Time series analysis was earlier applied to so-called clusters only, which are a result of dividing disciplinary publications into narrow groups. The groups are usually considered to be representative for certain scientific subdisciplines. Clustering is always done on the basis of one out of many available similarity criteria, in particular the one called a co-word similarity. Furthermore, there is a wide variety of clustering algorithms, which typically yield considerably differing results. As a result, it seems that the approach based on direct citedness of scientific publications helps to eliminate the redundant vagueness and the inevitable lack of universality which appear when one analyzes the results of an algorithmic clustering.

### Keywords

Bibliometrics. Development of scientific disciplines. Scientific periodicals. Statistical methods in information science. Time series analysis.

---

*EUKASZ OPALIŃSKI* uzyskał tytuł magistra w zakresie nauki o języki i komunikacji (jako specjalności wyodrębnionej z kierunku filozofia) na Uniwersytecie Marii Curie-Skłodowskiej w Lublinie w 2005 r. Ukończył studium podyplomowe „Technologie informacyjne i bibliotekoznawstwo” w Wyższej Szkole Informatyki i Zarządzania z siedzibą w Rzeszowie w 2009 r. oraz studium podyplomowe „Bibliotekarz dziedzinowy” w Instytucie Informacji Naukowej i Bibliotekoznawstwa Uniwersytetu Jagiellońskiego w 2012 r. Pracuje w Oddziale Informacji Naukowej Biblioteki Politechniki Rzeszowskiej na stanowisku bibliotekarza. Najważniejsze publikacje: *Ł. Opaliński* (2012). Niealfabetyczne systemy piśmiennicze a komunikacja językowa w Internecie [online]. *Podkarpackie Studia Biblioteczne*, 1/2012, 1–19, [25. 02. 2017], <http://repozytorium.ur.edu.pl/handle/item/180>; *Ł. Opaliński* (2013). Wybrane aspekty metodologii badań cyklu życiowego publikacji naukowych. *Przegląd Biblioteczny*, 81(2), 152–171; *Ł. Opaliński, M. Jaromin, J. Wikiera* (2015). Problem stabilności zachowań naukowców w zakresie cytowań w kontekście metodologii badań starzenia się publikacji naukowych i możliwość jego ujęcia ilościowego. *Zagadnienia Informacji Naukowej*, 53(2/106), 65–83.

*Rola w przygotowaniu artykułu:* Opracowanie części teoretycznej, analiza literatury przedmiotu, opracowanie wykresów, tabel i aneksów, zebranie danych empirycznych i interpretacja wyników badania.

*Udział:* 60%.

*Kontakt z autorem:*

*lopa@prz.edu.pl*

*Oddział Informacji Naukowej Biblioteki Politechniki Rzeszowskiej*

*al. Powstańców Warszawy 12*

*35–959 Rzeszów*

*MARCIN JAROMIN* pracuje na stanowisku asystenta w Zakładzie Biotechnologii i Bioinformatyki Wydziału Chemicznego Politechniki Rzeszowskiej. Tytuł magistra inżyniera uzyskał w 2004 r. na Wydziale Chemicznym Politechniki Rzeszowskiej oraz, równolegle, w 2005 r. na Wydziale Elektrotechniki i Informatyki Politechniki Rzeszowskiej. Specjalizuje się w dziedzinie bioinformatyki i statystyki matematycznej. W dniu 11 stycznia 2012 r. Rada Wydziału Chemicznego podjęła decyzję o otwarciu jego przewodu doktorskiego. Najważniejsze publikacje: *B. Dębska, M. Jaromin* (2006). Komputerowo wspomagany dobór materiałów inżynierskich w procesie projektowania. W: *B. Dębska, G. Fic* (red.), *Systemy Informacyjne w Chemii. Tom 3*. Rzeszów: Oficyna Wydaw. Politechniki Rzeszowskiej, 37–46.; *B. Dębska, M. Jaromin, P. Peszko* (2006). Nowe metody kształcenia studentów na kierunku Technologia Chemiczna. *Przemysł Chemiczny*, 85(8–9), 1171–1174; *A. Bocian, K. Hus, M. Jaromin, M. Tyrka, A. Łyskowski* (2017). Identification of Proteins Differentially Accumulated in *Enterococcus Faecalis* Under Acrylamide Exposure. *Turkish Journal of Biology*, 41(1), 166–177.

*Rola w przygotowaniu artykułu:* analiza statystyczna.

*Udział:* 40%.

*Kontakt z autorem:*

*mjaromin@prz.edu.pl*

*Wydział Chemiczny Politechniki Rzeszowskiej*

*al. Powstańców Warszawy 6*

*budynek H, pokój H-242*

*35–959 Rzeszów*

---

# Quantified Self. Self-tracking a problem tożsamości

Łukasz Iwasiński

*Katedra Informatologii, Wydział Dziennikarstwa, Informacji i Bibliologii  
Uniwersytet Warszawski*

---

## Abstrakt

**Cel/Teza:** Self-tracking to praktyka polegająca na pozyskiwaniu przez jednostki najróżniejszych danych dotyczących ich samych, przede wszystkim za pomocą mobilnych urządzeń i powiązanych z nimi aplikacji. Funkcją self-trackingu jest poprawa zdrowia, wydajności, organizacji różnych aktywności itp. W artykule postawiono tezę, że jego wpływ na życie współczesnych ludzi jest większy – self-tracking staje się ważnym instrumentem kształtowania ich tożsamości.

**Koncepcja/Metody badań:** Analiza opiera się na literaturze teoretycznej i empirycznej dotyczącej problematyki self-trackingu. Podjęte w artykule rozważania zilustrowano przykładami zaczerpniętymi ze źródeł zastanych oraz własnych obserwacji.

**Wyniki i wnioski:** Self-tracking przynosi obietnicę dogłębnego poznania swego ciała i umysłu – a więc odkrycia i, na ile to możliwe, przekształcenia swej tożsamości. Praktyka ta daje wrażenie lepszej, skuteczniejszej kontroli swego życia (przede wszystkim swej cielesności), a dzięki temu – poszerzenia możliwości kreowania tożsamości. I zapewne w jakimś zakresie faktycznie pomaga realizować te cele. Technologie Quantified Self stwarzają jednak pozory dostępu do wiedzy „czystej”, to jest niezapośredniczonej przez żaden arbitralny system ekspercki, a opartej na surowych danych empirycznych i tym samym pewnej. Ta ich właściwość prowadzić może do nadmiernej ufności w ich wskazania.

**Oryginalność/Wartość poznawcza:** Temat self-trackingu / Quantified Self jest aktualny, w polskim piśmiennictwie słabo rozpoznany, a istotny poznawczo, choćby ze względu na powszechność i rozwojowość tego zjawiska.

## Keywords

Danetyzacja. Quantified Self. Self-tracking. Tożsamość.

*Otrzymano: 15 października 2017. Zrecenzowano: 26 października 2017. Zaakceptowano: 28 grudnia 2017.*

---

## 1. Wprowadzenie. Self-tracking i Quantified Self

Self-tracking<sup>1</sup> to praktyka polegająca na pozyskiwaniu przez jednostki najróżniejszych danych dotyczących ich samych, przede wszystkim za pomocą mobilnych urządzeń i powiązanych z nimi aplikacji. Jej inne, synonimiczne lub bliskoznaczne nazwy, obecne

---

<sup>1</sup> W polskim piśmiennictwie zwykle nie stosuje się tłumaczenie tego terminu, dlatego użyto jego angielskiej formy. Niektóre inne obecne w niniejszym artykule pojęcia występują w rodzimym piśmiennictwie zarówno w wersji angielskiej, jak i w polskiej. W takich przypadkach posłużono się najpowszechniej przyjętym przekładem polskim, podając jednocześnie w nawiasie termin oryginalny.

w piśmiennictwie specjalistycznym i publicystyce, to: lifelogging, auto-analytics, personal analytics, self-quantifying, self-surveillance, personal informatics, body hacking (Lupton, 2016, 102; Hill, 2011; Wilson, 2012; The Economist, 2012). Zjawisko to stało się w ostatnich latach dość powszechne. Jego entuzjaści powołali nieformalny ruch pod nazwą Quantified Self (QS)<sup>2</sup>.

Self-tracking (jako praktyka) i Quantified Self (jako będący jej efektem społeczny fenomen) są owocem szeregu, częściowo sprzężonych ze sobą czynników społeczno-kulturowych, ekonomicznych i technologicznych:

- wzrastającej racjonalizacji życia, co znajduje wyraz w dążności do coraz większej jego kontroli i coraz efektywniejszego nim zarządzania;
- postępującej indywidualizacji i oddziaływaniu kultury terapeutycznej na jednostkę (Jacyno, 2007);
- rozwoju biopolityki technocentrycznej i jej demokratyzacji (Lemke, 2010, 34–40, 115) oraz medykalizacji życia (Jacyno, 2007, 91–95);
- rozwoju technik informacyjnych i komunikacyjnych, w szczególności technik pomiaru, rejestracji i przetwarzania danych oraz cyfrowej komunikacji bezprzewodowej;
- postępującej kwantyfikacji, a tym samym danetyzacji najróżniejszych wymiarów rzeczywistości, w tym człowieka.

Kwantyfikacji, monitoringowi i zapisowi podlega coraz więcej sfer ludzkiego funkcjonowania. Współczesne mobilne aplikacje umożliwiają ujmowanie w liczby i rejestrowanie fizjologicznych i emocjonalnych parametrów ich organizmów (np. czynności serca, poziomu hormonów, EEG, ECG), różnych form aktywności (np. wykonane kroki, dystans przejechany na rowerze), interakcji ze środowiskiem społecznym (np. różne cechy i wzory komunikacji) i fizycznym (np. poziom zanieczyszczenia powietrza otaczającego użytkownika). Entuzjaści self-trackingu (zwani self-trackerami lub QSerami) wychodzą z założenia, że: po pierwsze – każde działanie, a nawet każdy stan generuje określone dane; po drugie – teoretycznie policzyć da się w zasadzie wszystko, każdą informację jakościową można ująć w formułę liczbową. Stosowane przez nich praktyki nie ograniczają się do gromadzenia łatwo uchwytanych danych, ale obejmują także te pozyskiwane za pomocą złożonych wskaźników. Często korzystają oni również ze zautomatyzowanej analizy, interpretacji danych i różnych form ich prezentacji, a w konsekwencji otrzymują stosowne rekomendacje, za co odpowiadają specjalnie konstruowane algorytmy. Technologie wykorzystywane przez zwolenników self-trackingu mogą, na przykład, na podstawie dostarczonych przez poszczególnych użytkowników informacji, obliczyć liczbę spożytych i spalonych w wybranej jednostce czasu kalorii, ale i wskazać odpowiednią dla każdego ilość energii, czy zaplanować dietę.

A zatem przyjąć można, że idea Quantified Self odnosi się nie tylko do pozyskiwania, ale też do organizacji, analizy, interpretacji oraz wizualizacji danych i w rezultacie – tworzenia wytycznych dotyczących zachowania, podejmowanych aktywności, stylu życia. Ogół środków umożliwiających wszystkie te działania nazwę technologiami Quantified Self (technologiami QS). Technologie takie często pobierają dane (a także je przetwarzają,

<sup>2</sup> Pojęcie „Quantified Self” ukuli w 2007 r. redaktorzy magazynu *Wired*, Gary Wolf i Kevin Kelly, autorzy strony [quantifiedself.com](http://quantifiedself.com). Chcieli oni skonsolidować nieformalny ruch entuzjastów cyfrowego samonadzoru. W 2011 r. odbyła się pierwsza sygnowana tym mianem konferencja w Mountain View w Kalifornii. Z czasem pojęcie „Quantified Self” upowszechniło się i przyjęło w piśmiennictwie – zarówno publicystycznym, jak i naukowym.



realizując wymienione wyżej funkcje) za pomocą tzw. urządzeń ubieralnych (ang. *wearable technologies*)<sup>3</sup>, to jest akcesoriów oraz elementów odzieży, wyposażonych w czujniki rejestrujące określone parametry oraz w inne rozwiązania z zakresu technologii informacyjnych. Niektóre z takich obiektów określa się również poprzedzając ich pierwotne nazwy przedrostkiem „smart-”<sup>4</sup>. Przyjąć można, że historia urządzeń ubieralnych sięga lat 70. XX w., kiedy to zaczęły upowszechniać się mikroprocesory. W kolejnych dekadach techniczne możliwości ich projektowania sukcesywnie poszerzały się dzięki postępującej miniaturyzacji układów scalonych oraz rozwojowi cyfrowej komunikacji bezprzewodowej, a w końcu coraz śmielszym eksperymentom z inteligentnymi tkaninami (ang. *smart fabrics*). Dziś urządzenia ubieralne funkcjonują w powiązaniu (a raczej: jako element) Internetu Rzeczy (ang. *Internet of Things*), ponieważ mają zwykle zdolność łączenia się z ogólnosiatową siecią i wymieniać między sobą dane.

Technologie QS zwykle związane są z dążeniem do lepszego zarządzania wybranymi aspektami życia, dzięki ich ściślejszej kontroli. Najczęściej znajdują zastosowanie w dziedzinie zdrowia oraz rekreacji, służąc profilaktyce chorób, poprawie sprawności, wyglądu i innych elementów jakości życia. Wykorzystywane są także w celu zwiększania produktywności, lepszej organizacji codziennych aktywności oraz budowania relacji z innymi osobami. Sukcesywnie zawłaszczają kolejne obszary funkcjonowania ludzi. Deborah Lupton pisze, że praktyki self-trackingowe na pierwszy rzut oka zdają się być zarezerwowane dla konkretnych subkultur złożonych z jednostek chronicznie chorych, czy też obsesyjnych, narcystycznych, wykazujących ponadprzeciętne zainteresowanie ochroną zdrowia, zwiększeniem sprawności fizycznej oraz efektywności w różnych obszarach funkcjonowania – i taki wizerunek Quantified Self utrwaliły przekazy publicystyczne oraz popkultura (Lupton, 2013). W rzeczywistości praktyki te nie są ograniczone do żadnej zamkniętej populacji, stają się coraz powszechniejsze, wkraczając w dziedzinę pracy, edukacji, ubezpieczeń, marketingu i reklamy, ekonomii, energetyki, wojska, urbanistyki i innych (Lupton, 2016). Przewodnik *Quantified Self Guide to Self-Tracking Tools* wylicza ponad 500 technologii QS używanych w najróżniejszych kontekstach (Quantified Self Guide, b.d.).

Stawiam zatem tezę, że self-tracking staje się ważnym instrumentem kształtowania tożsamości współczesnych ludzi.

## 2. Tożsamość w późnej nowoczesności

Socjologia ujmuje tożsamość jako „układ autodefinicji aktora społecznego” czy też „zbiór wyobrażeń, sądów i przekonań, które konstruuje on wobec samego siebie” (Bokszański, 1989, 12). W społeczeństwach tradycyjnych była ona w dużej mierze przypisana i stała, nadawana niemal automatycznie przez zewnętrzne wobec jednostki kulturowe mechanizmy.

<sup>3</sup> Co w piśmiennictwie określane bywa także jako „Smart Cloths”, „Responsive Cloths”, „Computational Garments”, „Fashionable Technologies”, „Smart Textiles” (Ryan, 2014, 6), a oparte jest na „inteligentnych tkaninach” (*Smart Fabrics*).

<sup>4</sup> Przedrostek „smart” zwykle tłumaczony jest na język polski jako „inteligentny”, ale w przypadku najpowszechniejszych przedmiotów mamy do czynienia ze spolszczonym terminem oryginalnym – stąd „smartfon”, „smartwatch” bądź „inteligentny zegarek”, „smartband” bądź „inteligentna opaska” (funkcjonują obie wersje).

W wyniku rozwoju nowoczesności, zwłaszcza w jej późnej fazie<sup>5</sup>, charakteryzującej się postępującą indywidualizacją społeczeństw, następowało sukcesywne autonomizowanie się tożsamości od zewnętrznych instancji. Towarzyszy temu wzrastające poczucie, że człowiek jest bytem – jak pisze Chris Shilling (2010, 194) – nieukończonym. Późnonowoczesna tożsamość jest zatem w znacznie większym stopniu osiągnana, tworzona w wyniku refleksyjnie konstruowanego projektu życia; trzeba ją samodzielnie „wyrzeźbić”, mając jednak do dyspozycji całą mnogość ofert. Jest ona wyrażana poprzez styl życia. Tożsamość podlega potwierdzeniu oraz negocjacji podczas wszelkiego typu społecznych interakcji. Podmioty owej interakcji stają się dla jednostki źródłem samowiedzy; swoistymi lustrami, w których się ona przegląda, a obrazy poddaje interpretacji.

Zygmunt Bauman<sup>6</sup> (2004, 220–221) u progu lat 90. XX w. twierdził, że życie we współczesnym świecie wymaga zarządzania sobą niczym przedsiębiorstwem. Ludzie traktują egzystencję jako zbiór wyzwań i problemów, które można wyizolować. Na każdy z nich istnieje przygotowane przez ekspertów rozwiązanie. Tę ideę rozwija Anthony Giddens. Jego zdaniem jednostka funkcjonująca w późnonowoczesnej rzeczywistości, w trudnym zadaniu konstruowania własnej tożsamości musi szukać wsparcia w różnorodnych systemach eksperckich<sup>7</sup>. Stanowią one owoc kluczowej dla nowoczesności dążności do racjonalizacji świata (w tym ludzkiego życia), by móc go skuteczniej kontrolować. Jednak w charakteryzującej się rosnącą złożonością, niestabilnością i małą przejrzystością współczesnej rzeczywistości trudno, a czasem w ogóle nie sposób, zweryfikować faktycznej racjonalności proponowanego przez dany system ekspercki rozwiązania; pozostaje ono często kwestią wiary, przekonania czy mody. Stąd oferowane przez systemy eksperckie koncepcje konkurują ze sobą – co pluralizuje „rynek tożsamości”. Wiedza ekspercka przekładana jest na różnego typu usługi doradcze oraz terapeutyczne. Niemal każda dziedzina życia (od mody, przez dietę, po kwestie ideologiczne czy duchowe), a tym samym każdy aspekt tożsamości jest przez owe systemy mediowany. Tomasz Leszniewski zauważa, że

w codziennych sytuacjach życiowych posilkujemy się zaleceniami naukowców, np. żywiąc się masłem albo margaryną, stosując medycynę konwencjonalną bądź naturalną, czy też wychowując dzieci bezstresowo lub ucząc je dyscypliny itp. Coraz częściej sięgamy po poradniki podpowiadające np. jak osiągnąć sukces, jak żyć we dwoje, jak mówić ‘nie’ itp. (Leszniewski 2007, 50).

<sup>5</sup> Nie można precyzyjnie wyznaczyć daty oddzielającej poszczególne fazy nowoczesności. Socjologowie zwykle ujmują tę epokę dwudzielnie (Bokszański, 2007, 36–38), bądź też, chcąc silniej zaznaczyć ową cezurę, posługują się pojęciami nowoczesności i ponowoczesności. Jako umowny moment rozdzielający fazy nowoczesności można przyjąć drugą połowę lat 60. XX wieku. Decyzja ta wynika z faktu, że wtedy to właśnie mieliśmy do czynienia z intensyfikacją przemian kulturowych i obyczajowych, które istotnie wpłynęły na poczucie tożsamości mieszkańców zachodnich społeczeństw. Warto przywołać tu chronologię rozwoju zachodniej tożsamości zaproponowaną przez Hermanna Bausingera. Autor ten wyszczególnił tożsamość stabilną (obecną w społeczeństwach przednowoczesnych, tradycyjnych), tożsamość ideologiczną (która ufundowana była przede wszystkim na wielkich narracjach i narzuconych przez nie systemach normatywnych i symbolach, zwłaszcza ideologii narodowej – obecną w pierwszej fazie nowoczesności), tożsamość bezsensowną (będącą wynikiem upadku owych narracji – obecną w późnej nowoczesności czy też ponowoczesności) (Bokszański, 1989, 39–40).

<sup>6</sup> Korzystam z wydania książki *Socjologia* z 2004 r., ale jej pierwsza edycja pochodzi z 1990 r.

<sup>7</sup> Giddens stwierdza: „Systemy eksperckie wiążą czas i przestrzeń rozwijając wiedzę techniczną (...). Takie systemy przeniknęły właściwie wszystko: dotyczą tego, co jemy, lekarstw, które zażywamy, domów, w których mieszkamy, środków transportu, z których korzystamy, i całego mnóstwa innych aspektów życia społecznego w warunkach nowoczesności” (Giddens, 2010, 34).

Dokonując w każdej z tych dziedzin wyboru, samookreślamy się. Anthony Giddens pisze:

Nowoczesność jest porządkiem posttradycyjnym, który codziennie, wraz z każdą decyzją, jak się zachować, w co się ubrać, co zjeść, i mnóstwem innych, wymusza na każdym pytanie „jak żyć” i każe interpretować odpowiedzi w kategoriach własnej, dynamicznie zmieniającej się tożsamości (Giddens, 2010, 28).

Późnonowoczesny świat coraz bardziej problematyzuje tożsamość, co chwila zmusza ludzi do postawienia sobie nowych, związanych z nią pytań. Stąd rozwój coraz to nowych zawodów, których rolą jest dostarczyć wiedzy i instrumentów do rozwiązywania owych problemów – wszelkiego typu doradców, terapeutów, trenerów, couchów, dietetyków, stylistów, projektantów itd.

Chris Shilling (2010, 14, 179, 194) argumentuje, że we współczesnym świecie wzrasta rola ciała dla poczucia jednostkowej tożsamości, choć oczywiście ciało nie determinuje w pełni poczucia tożsamości. Ontologiczne pewniki będące na zewnątrz jednostki ulegają erozji, ciało staje się jedynym pewnikiem. Shilling uważa, że

Tym, którzy utracili wiarę w religijne autorytety i wielkie polityczne narracje i nie są już przez te transpersonalne struktury znaczeniowe wyposażeni w jasną wizję świata czy siebie, przynajmniej ciało zdaje się (...) dostarczać silnych fundamentów, na których można w nowoczesnym świecie budować na nowo stabilne poczucie tożsamości (Shilling, 2010, 17).

Dodatkowo, neuronauka głosi ideę materialności umysłu, a więc redukuje umysł również do cielesności (Waszewski & Zybortowicz, 2015, 59–104; Dennett, 2016, 50–61). Odwołując się raz jeszcze do Shillinga:

współczesna epoka jest coraz bardziej „poza kontrolą”. W tym kontekście ciało zaczyna być traktowane jako ostatni „surowy materiał”, na który jednostki mogą wywierać wpływ. Ciało, poddane niespotykanemu dotąd poziomowi racjonalizacji, postrzega się teraz jako przedmiot, który daje się modyfikować w kategoriach genów, ciśnienia krwi, cholesterolu, rozmiaru, kształtu, wyglądu, wydolności oddechowej, koloru, wagi oraz wielu innych czynników<sup>8</sup>. Jako takie wydaje się, że dostarcza mocnych podstaw, na których można zbudować stabilne i sensowne poczucie siebie we współczesnym świecie (Shilling, 2010, 227).

Zatem, jeśli zgodzić się z przytoczonymi wyżej argumentami, ciało jest jak nigdy wcześniej w historii kluczowe dla samookreślenia jednostki, jest być może podstawowym nośnikiem tożsamości.

Zarazem „osiągnięcia techniczne i wiedza ekspercka dokonały inwazji na ciało i udostępniły możliwość pracowania nad nim i jego rekonstruowania” (Shilling, 2010, 194). Oczywiście może to odbywać się w pewnych ramach, wyznaczonych zarówno przez biologię (zwłaszcza geny, które postrzega się jako czynniki determinujące określone aspekty życia człowieka, a więc nie dające się przewartościować (Domaradzki, 2017), jak i kulturę (posiadany kapitał ekonomiczny oraz kulturowy).

Zgodzić się jednak należy, że współczesna epoka dostarcza bezprecedensowych narzędzi monitoringu i kontroli ciała, a także, że pozwala na nieznanie wcześniej sposoby je kształtować, czyniąc ciało przedmiotem opcji i wyboru (Shilling, 2010, 17). Szczególnego znaczenia nabierają dziś „reżimy ciała”, czyli praktyki służące kultywowaniu wybranych cech ciała. Już w latach 80. XX w. pisano, że

<sup>8</sup> Ta lista oczywiście nie wyczerpuje wszystkich czynników, to tylko kilka przykładów, wskazanych przez Shillinga w publikacji z 2003 r. Dziś, w realiach rozpowszechniania się technologii QS, ich liczba i możliwości kontroli oraz modyfikacji rośnie.

w Stanach Zjednoczonych podporządkowanie się reżimom zdrowia i fitness osiągnęło wśród niektórych specjalistów z klasy średniej taki poziom, że ci, którzy je odrzucają, traktowani są jak dewianci (Shilling, 2010, 194).

Radykalne formy transformacji ludzkiej cielesności określa Shilling (2010, 228) mianem „opcji ciała”. Mają one swe, wspomniane wyżej, ograniczenia. Jednak ich rozwój postępuje – już dziś obejmować mogą cyborgizację, co pozwala na nowe, bardziej elastyczne formy restrukturyzacji cielesności i przekraczanie kolejnych ograniczeń, jakie ciało narzuca biologia. Opcje ciała wpisują się w promowaną przez transhumanistów zasadę „wolności morfologicznej”, która mówi o

prawie każdego człowieka do takich zmian własnego ciała i umysłu, które zwiększą fizyczne i intelektualne możliwości człowieka, poprawią zdrowie, zwiększą długość życia i zagwarantują satysfakcję z niego płynącą (Bobryk, 2014, 10–11).

### 3. Self-tracking jako narzędzie budowania tożsamości

Różny bywa stopień zaangażowania w self-tracking – od sporadycznego monitoringu wybranych parametrów, po obsesyjną kontrolę jakiegoś aspektu życia, dążenie do jak najdokładniejszego odzwierciedlenia go w strumieniu danych, śledzenia setek wskaźników i bezkrytycznego podporządkowania się wytycznym, opracowanym przez algorytmy technologii Quantified Self. Różne też bywają motywy i stopień intencjonalności poddania się cyfrowej analizie. Przyjmuję tu, że praktyki self-trackingowe są dobrowolne (jak wynika z etymologii określenia self-tracking)<sup>9</sup>. Dla najbardziej zagorzałych QSerów, self-tracking stanowi nowy, ważny czynnik regulujący proces kształtowania się tożsamości. Tezę tę uzasadniam poniżej.

W epoce danetyzacji poradniki, kursy, terapie i inne usługi doradcze, przynajmniej w części, wypierane są przez wskazania technologii QS. Zewnętrznych ekspertów coraz częściej zastępuje bazująca na self-trackingu samoekspertyza, stając się środkiem służącym zarówno poznaniu, jak i konstruowaniu siebie. Technologie QS przynoszą więc obietnicę dogłębnego zdiagnozowania, a także przekształcania, ulepszenia własnego ja. Można by rzec – stworzenia optymalnej wersji samego siebie.

Czy zatem danetyzacja tożsamości prowadzi do jej postępującej emancypacji, innymi słowy technologie QS stają się dla współczesnych ludzi narzędziem stwarzającym większe możliwości uwalniania „ja” od zewnętrznych determinacji, a także demonstrowania swej indywidualności? Technologie QS monitorują i analizują coraz to nowe parametry organizmu człowieka, by móc go doskonalić. Sugerują wykonanie określonej pracy na ciele i umyśle (ucieleśnionym, tzn. kształtowanym przede wszystkim przez ciało albo wręcz zredukowanym do wymiaru somatycznego) oraz pozwalają na kontrolowanie jej efektów za pomocą kolejnych pomiarów. Poprzez taką pracę (diety, treningi, manipulacje mentalne czy emocjonalne) realizowane są reżimy oraz opcje ciała. Należy zgodzić się, że self-tracking sprzyja aktualizowaniu samowiedzy oraz traktowaniu tożsamości jako przedmiotu refleksyjnej samokontroli. Na swój sposób intensyfikuje mechanizm podwójnej

<sup>9</sup> Jakkolwiek Lupton (2016) rozumie ów terminy szerzej (nie tylko jako kwantyfikację „ja” przez dobrowolny autonadzór, ale także poprzez nadzór narzucony).

hermeneutyki – z samowiedzy płyną impulsy do pracy nad sobą i potencjalnego przededefiniowania swego ja. Ten mechanizm doskonale ilustrują technologie oparte na biofeedbacku.

Jak już powiedziano, przebudowa tożsamości odbywać może się w pewnych ramach, dla których krytyczną granicę wyznaczają geny. Rozwój usług z obszaru genetyki spersonalizowanej zmierza do znalezienia „unikatowego genetycznego profilu”, który odpowiadać ma za najróżniejsze fizyczne i psychologiczne cechy jednostki. Dąży się tu do wyrażenia w esencjalistycznych, biologicznych kategoriach niemal każdego aspektu tożsamości jednostki<sup>10</sup>. Jak zauważa Jan Domaradzki (2017, 307) – „W każdym przypadku odwoływanie się do retoryki tożsamości jest podstawową strategią marketingową firm świadczących tego typu usługi”. Usługi owe obejmują jednak nie tylko orzekanie o niezmiennym genetycznym statusie człowieka, ale także wskazanie kierunków możliwego rozwoju. Jedna z firm oferujących testy z zakresu genetyki spersonalizowanej reklamuje się hasłem „Zostań Genoświadomym!” i deklaruje, że pomoże „optymalnie wykorzystywać swoje genetyczne predyspozycje, w zgodzie z genotypem swojego » ja « ” (Domaradzki, 2017, 305); ogłasza także:

Dzięki badaniu TALENT poznasz nie tylko skalę swojego potencjału (intelektualnego, fizycznego i psychicznego), ale też sposoby na wzmocnienie słabszych stron. Wiedza ta pozwoli Ci w zgodzie z Twoimi możliwościami wybrać np. kierunek edukacji czy zaplanować lub zmienić swoją ścieżkę zawodową. Dowiesz się także, które grupy swoich zainteresowań powinieneś rozwijać i jaki sport uprawiać, aby odnosić sukcesy. To sposób, by sprawdzić, jakie ukryte talenty znajdują się w Twoim DNA (Domaradzki, 2017, 307)<sup>11</sup>.

Dodajmy, że dotychczas testami z zakresu genetyki spersonalizowanej zajmują się wyspecjalizowane laboratoria, ale niewykluczone, że kiedyś możliwe będzie wykonanie takiej procedury za pomocą mobilnej aplikacji. Dziś istnieją już aplikacje, udostępniane zresztą przez firmy wyspecjalizowane w spersonalizowanej genetyce, wspomagające zarządzanie różnymi aspektami życia w oparciu o uzyskane wyniki badań. Wskazują one najlepszą dietę, optymalny rodzaj treningu fizycznego, czas snu, czy optymalne rodzaje aktywności odpowiadające genetycznemu profilowi użytkownika.

Technologie QS nie koncentrują się wyłącznie na biologicznym aspekcie tożsamości. Analizują także jej parametry społeczno-kulturowe. A więc obok funkcji kontroli i doboru diety czy określania i poprawiania nastroju, pomagają na przykład uświadomić sobie i nazwać swój gust<sup>12</sup>. Przedmiotem zapisu, przetwarzania i wizualizacji często są różne wybory w sferze kultury i rozrywki. Gdy zostaną one unaocznione, mogą stać się instrumentem samopoznania. Łatwo uchwycić ten fenomen, analizując konsumpcję dóbr niematerialnych (np. muzyki). Autorzy opublikowanego w 2010 r. raportu *Młodzi i media*, pisząc o społecznościowym muzycznym serwisie Last.fm zauważyli, że pozwala on użytkownikom „uzewnętrznić i zobaczyć swój gust”, dzięki czemu staje się przedmiotem »refleksji

<sup>10</sup> Warto zwrócić uwagę na nazwy firm oferujących tego typu usługi: deCODEme, Knome czy Mycellf (Domaradzki, 2017, 305).

<sup>11</sup> Abstrahuję tu od oceny wiarygodności tego typu deklaracji, ponieważ nie mam kompetencji do ich oceny. Interesuje mnie jedynie ich społeczne oddziaływanie.

<sup>12</sup> Zgodnie z ujęciem Colina Campbella (2006, 50–51), to gusty „określają nas dokładniej niż wszystko inne [...]. Ów ktoś, za kogo naprawdę się uważamy, nasze prawdziwe ja, stanowi bowiem efekt specyficznej mieszanki czy też kombinacji gustów”. Tradycyjne wyznaczniki tożsamości, jak „płeć, rasa, narodowość, przynależność etniczna i religia” (Campbell 2006, 51) stanowią „ramę” dla tego, co jednostka uznaje za swą autentyczną tożsamość, za to, co przynależy tylko jej, świadczy o jej indywidualności, a więc właśnie gust.



i samowiedzy« oraz »twórczych zabiegów« (Filiciak et al., 2010, 11). Każdy użytkownik serwisu ma możliwość śledzić indywidualne statystyki odsłuchanych utworów wraz z ich metadanymi, a także dokonywać automatycznych porównań w wielu wymiarach (m.in. częstości odsłuchań, preferencji) z innymi uczestnikami. Jak mówi jedna z członkiń społeczności Last.fm – „mogę zobaczyć ewolucję swojego gustu” (Filiciak et al. 2010, 87). Last.fm staje się dla użytkowników narzędziem uzewnętrznienia swego ja i komunikacji z innymi, którzy poprzez reakcje i dostarczenie słownika pozwalają im spojrzeć na siebie niejako z boku i lepiej zdefiniować tożsamość. Im wnikliwszej obserwacji poddają się jednostki, im więcej posiadają na swój temat wiedzy, tym bardziej precyzyjne są te autodefinicje. Dziś serwis Spotify, następcą Last.fm (który w 2014 r. zawiesił streamowanie muzyki, koncentrując się na systemie rekomendacji) oferuje usługę Spotify.me reklamowaną hasłem „you are what you stream” (jesteś tym, co streamujesz). Spotify.me pozwala sprawdzić nie tylko najczęściej odtwarzane przez użytkowników utwory, najchętniej słuchanych artystów i gatunki muzyczne, ale także m.in. „godziny, w których korzystamy najczęściej z platformy” czy „podsumowanie z podziałem na tempo, energię i... taneczność” (Antyweb, 2017).

Informacje pozyskane za pomocą technologii Quantified Self są więc źródłem samopoznania (choć właściwiej byłoby powiedzieć – samopoznawania, jako że jest to samozwrotny proces) i doskonalenia. Są też przedmiotem autoprezentacji oraz rywalizacji (za przykład mogą posłużyć publikowane za pomocą aplikacji do monitoringu osiągnięć sportowych, zdaneyzowane i odpowiednio zwizualizowane rezultaty). Zjawisko to potęguje (czy też jest jednym z jego przejawów) grywalizacja (ang. *gamification*), która polega na przenoszeniu mechanizmów typowych dla gier, zwłaszcza wyzwalania współzawodnictwa i zachęcania do podejmowania wyzwań, do kontekstów pierwotnie niezwiązanych z grami w celu intensyfikacji zaangażowania ludzi w określone działania. Tak oto osoby uprawiające indywidualnie, rekreacyjnie jakąś dyscyplinę sportu (choćby bieganie), korzystające z technologii QS, mogą za ich pomocą wyznaczać sobie cele i monitorować proces ich realizacji, kolekcjonować nagrody w postaci trofeów czy punktów, zdobywać coraz wyższe poziomy oraz demonstrować rezultaty za pośrednictwem publicznych rankingów – ścigając się z innymi użytkownikami. W ten sposób podejmowane w pojedynkę aktywności nabierają wymiaru społecznego (koncepcja Quantified Us). Taki rodzaj gry uprawianej z innymi ludźmi może stać się punktem odniesienia dla budowania tożsamości grupowej, a zarazem areną walki statusowej.

#### 4. Wnioski

Self-tracking przynosi obietnicę dogłębnego poznania swego ciała i umysłu – a więc odkrycia i na ile to możliwe, przekształcenia swej tożsamości. Czy praktyka ta w istocie zwiększa naszą samowiedzę? Czy daje szansę jej amatorom na stanie się najlepszym wariantem samych siebie? Na pewno daje wrażenie lepszej, skuteczniejszej kontroli swego życia (przede wszystkim swej cielesności), a dzięki temu poszerzenia możliwości kreowania tożsamości. I zapewne w jakimś zakresie faktycznie pomaga realizować te cele. Technologie Quantified Self stwarzają jednak pozór dostępu do wiedzy „czystej”, to jest niezapśredniczonej przez żaden arbitralny system ekspercki, a opartej o surowe dane i tym samym pewnej. Ta ich właściwość prowadzić może do nadmiernej ufności w ich wskazania. W istocie zawsze



mamy do czynienia z mniej lub bardziej arbitralną reprezentacją określonych stanów, a nie „czystymi danymi” (w tym sensie Quantified Self jest zawsze w jakimś stopniu Qualified Self; por. Lupton, 2014)<sup>13</sup>. Nie należy zapominać, że technologie QS w istocie stanowią szczególnie przypadek systemów eksperckich. Self-tracking wciąż inspirowany jest przez ekspertów; to oni tworzą technologię, orzekają co, dlaczego i jak ma być mierzone, narzucają określoną interpretację wyników.

W technologiach Quantified Self można też widzieć instrument dyscyplinujący, tłamszący swobodę rozwijania i ekspresji tożsamości. Michał Wróblewski (2016, 7) zauważa, że używanie gadżetów do self-trackingu tworzy określoną podmiotowość – skupioną na sobie, nieustannie samodyscyplinującą się, maksymalizującą kontrolę w codziennym życiu. Rolą tych technologii jest korygować, upominać, zachęcać do doskonalenia. Wtłaczają one jednostkę w reżim – posłużę się sformułowaniem Małgorzaty Jacyno (2007) – „mikroracjonalności”, które łatwo stać się mogą formą samozniewolenia. Potencjalnie mogą one prowadzić także do alienacji „ja”. Dane generowane i dystrybuowane za pomocą technologii QS bywają przechwytywane przez najróżniejsze instytucje i wykorzystywane są do tworzenia cyfrowego wizerunku użytkownika, który potrafi oderwać się od swego właściciela. Fenomen autonomizacji zdanetyzowanej, cyfrowej osoby, opisany został za pomocą kategorii data-double:

Realne podmioty przekształcane są w ciągi dyskretnych informacji, które następnie poddawane są procesowi arbitralnego porównywania, uogólniania, łączenia, w wyniku czego dochodzi do wytworzenia profilu jednostki, jej informacyjnego sobowtóra (data-double) wytworzonego z danych (Ożóg, 2009, 19).

Jednostki tracą kontrolę nad swą tożsamością, ponieważ nie mogą „wpływać na konfigurację danych wyznaczających charakter wirtualnego kłona”. Tożsamość „funkcjonuje już nie tylko jako samoistny wirtualny byt, lecz zyskuje wręcz przewagę nad materialną egzystencją, mogąc realnie wpływać na życie jednostki” (Ożóg, 2009, 20).

## Bibliografia

- Antyweb (2017). *Spotify.me, czyli podsumowanie naszych muzycznych gustów okiem Spotify!* [online]. Antyweb [15.10.2017] <http://antyweb.pl/spotify-me-podsumowanie/>
- Bauman, Z. (2004). *Socjologia*. Poznań: Zysk.
- Bobryk, J. (2014). Transhumanizm, cognitive science i wyzwania dla nauk społecznych. *Studia Socjologiczne*, (3), s. 9–28.
- Bokszański, Z. (1989). *Tożsamość, interakcja, grupa. Tożsamość jednostki w perspektywie teorii socjologicznej*. Łódź: Wydaw. Uniwersytetu Łódzkiego.
- Bokszański, Z. (2007). *Indywidualizm a zmiana społeczna*. Warszawa: PWN.
- Campbell, C. (2006). Kupuję, więc wiem, że jestem. Metafizyczne podstawy nowoczesnego konsumeryzmu. *Kultura Popularna*, 18(4), 47–60.
- Dennett, D. (2016). *Świadomość*. Kraków: Copernicus Center Press.
- Domaradzki, J. (2017). Genetyka, esencjalizm i tożsamość. *Studia Socjologiczne*, 224(1), 297–328.
- Filiciak M., Danielewicz M., Halawa M., Mazurek P., Nowotny A., red. (2010), *Młodzi i media. Nowe media a uczestnictwo w kulturze*. Warszawa: SWPS.

<sup>13</sup> Idea *Qualified Self* zakłada, że nie istnieje jeden, ostateczny sens danych. Nabierają one znaczenia dopiero po umieszczeniu ich w jakiejś ramie, odniesieniu kontekstu własnego życia i konkretnych sytuacji. A zatem te same dane mogą „opowiadać” różne historie.

- Giddens, A. (2010). *Nowoczesność i tożsamość*. Warszawa: PWN.
- Hill, K. (2011). Adventures in Self-Surveillance, aka The Quantified Self, aka Extreme Navel-Gazing [online]. *Forbes* [15.10.2017] <https://www.forbes.com/sites/kashmirhill/2011/04/07/adventures-in-self-surveillance-aka-the-quantified-self-aka-extreme-navel-gazing/#1db77ec46773>
- Jacyno, M. (2007). *Kultura indywidualizmu*. Warszawa: PWN.
- Lemke, T. (2010). *Biopolityka*. Warszawa: Sic!
- Leszniewski, T. (2007). Moda i tożsamość: dylematy współczesnego człowieka w świecie konsumpcji. W: T. Szlendak, K. Pietrowicz (red.). *Rozkoszna zaraza – o rządach mody i kulturze konsumpcji* (49–61). Wrocław: Wydaw. UWŹ.
- Lupton, D. (2013). Understanding the Human Machine. *IEEE Technology & Society Magazine*, 32(4), 25–30.
- Lupton, D. (2014). *Beyond the Quantified Self: The Reflexive Monitoring Self* [online]. This Sociological Life. A blog by sociologist Deborah Lupton [15.10.2017], <https://simplysociology.wordpress.com/2014/07/28/beyond-the-quantified-self-the-reflexive-monitoring-self/>
- Lupton, D. (2016). The Diverse Domains of Quantified Selves: Selftracking Modes and Dataveillance. *Economy and Society*, 45(1), 101–122.
- Ożóg, M. (2009). Transgresje panoptikonu. Nadzór w dobie technologii cyfrowych. *Kultura Współczesna*, 60(2), 14–30.
- Quantified Self Guide (b.d.). *Quantified Self Guide to Self-Tracking Tools* [online]. Quantified Self [21.12.2017], <http://quantifiedself.com/guide/toolsQuantified>
- Ryan, S. E. (2014). *Garments of Paradise: Wearable Discourse in the Digital Age*. Cambridge; London: MIT Press.
- Shilling, C. (2010). *Socjologia ciała*. Warszawa: PWN.
- The Economist (2012). The Quantified Self. Counting Every Moment. *The Economist* [online]. Technology Quarterly, Q1 [15.10.2017] <http://www.economist.com/node/21548493>
- Waszewski, J., Zybortowicz A. (2015). Neuronauka, jej technologie i możliwości. W: A. Zybortowicz (red.). *Samobójstwo Oświecenia? Jak neuronauka i nowe technologie pustoszą ludzki świat* (59–104). Kraków: Wydaw. Kasper.
- Willson, H.J. (2012). You, By the Numbers. *Harvard Business Review* [online], September [15.10.2017] <https://hbr.org/2012/09/you-by-the-numbers>
- Wróblewski, M. (2016). Nowe szaty healthismu. Self-tracking, neoliberalizm i kapitalizm kognitywny. *Folia Sociologica*, 58, 7–25.

---

## Quantified Self. Self-Tracking and Identity

### Abstract

**Purpose/Thesis:** Self-tracking refers to recording various data related to oneself, mainly through mobile devices and applications. It is considered to be a form of improvement of one's health, performance, organization, etc. In this paper, the author argues that the impact of self-tracking on peoples' lives is bigger – it is an important instrument used to shape their identity.

**Approach/Methods:** The analysis is based on theoretical and empirical literature on self-tracking. It is illustrated with the author's own observations or cases obtained from the desk research.

**Results and conclusions:** Self-tracking brings the promise of getting an in-depth knowledge of one's body and mind – thus discovering and, as far as it is possible, transforming the identity. It gives the impression of better, more effective control of one's life (especially one's body) and a sense of bigger possibilities of (re)creating the identity. To some extent it actually helps to achieve these

goals. However, self-tracking technologies create only an illusion of access to "pure" knowledge, i.e., not mediated by any arbitrary expert system, based on raw data and thus certain. This property may lead to excessive trust in their indications.

**Originality/Value:** Self tracking practices are becoming more and more common. The issues of self-tracking and Quantified Self are topical but still poorly recognized in Polish academic literature.

**Keywords**

Datafication. Identity. Quantified Self. Self-tracking.

---

*Dr ŁUKASZ IWASIŃSKI jest absolwentem Wydziału Organizacji i Zarządzania Politechniki Łódzkiej (tytuł magistra inżyniera – 2006 r.) oraz socjologii na Uniwersytecie Łódzkim (tytuł magistra – 2007 r., stopień doktora – 2013 r.). Obecnie jest adiunktem w Katedrze Informatologii na Wydziale Dziennikarstwa, Informacji i Bibliologii Uniwersytetu Warszawskiego. Ważniejsze publikacje: Ł. Iwasiński (2016). Socjologiczne dyskursy o konsumpcji, Gdańsk: Wydaw. Nauk. Katedra; Ł. Iwasiński (2007). Technologia – wszechrynek – konsumpcja. Kultura Współczesna, 4, 166–188; Ł. Iwasiński (2016). Społeczne zagrożenia danetyzacji rzeczywistości. W: B. Sosińska-Kalata et al. (red.), Nauka o informacji w okresie zmian. Informatologia i humanistyka cyfrowa (135–146). Warszawa: Wydaw. SBP.*

*Kontakt z autorem:*

*lukiwas@gmail.com*

*Katedra Informatologii*

*Wydział Dziennikarstwa, Informacji i Bibliologii*

*Uniwersytet Warszawski*

*ul. Nowy Świat 69*

*00–046 Warszawa*

---

# Wpływ cyfryzacji infrastruktury informacyjnej na procesy badawcze w humanistyce. Wybrane aspekty zjawiska

Ewa Głowacka

*Institut Informacji Naukowej i Bibliologii, Uniwersytet Mikołaja Kopernika w Toruniu*

Mirosław Górny

*Institut Językoznawstwa, Uniwersytet im. Adama Mickiewicza w Poznaniu*

Małgorzata Kisilowska

*Katedra Badań nad Bibliotekami i Innymi Instytucjami Kultury,  
Wydział Dziennikarstwa, Informacji i Bibliologii, Uniwersytet Warszawski*

Zbigniew Osiński

*Institut Informacji Naukowej i Bibliotekoznawstwa, Uniwersytet Marii Curie – Skłodowskiej w Lublinie*

---

## Abstrakt

**Cel/Teza:** Zbadanie przekształceń zachodzących w cyfrowej infrastrukturze informacyjnej humanistyki, opracowanie jej modelu oraz ustalenie widocznego już obecnie i przewidywalnego wpływu tych zmian i kształtującego się modelu na praktyki badawcze i publikacyjne badaczy; zaproponowanie metodyki przyszłych badań nad cyfrową infrastrukturą informacyjną.

**Koncepcja/Metody badań:** Pozyskanie i analiza danych dotyczących poszczególnych składników cyfrowej infrastruktury informacyjnej oraz danych o wykorzystywaniu teje przez wybrane grupy humanistów.

**Wyniki i wnioski:** Opracowano model cyfrowej infrastruktury informacyjnej humanistyki; wyróżniono parametry jej badania; wykazano związek pomiędzy wykorzystywaniem składników infrastruktury, a poziomem kompetencji informacyjnych i stanem świadomości informacyjnej; opisano aktualny stan wykorzystywania cyfrowej infrastruktury przez humanistów oraz opracowano mechanizmy jej potencjalnego wpływu na przyszłe praktyki badawcze humanistów.

**Oryginalność/Wartość poznawcza:** Opracowano oryginalny model cyfrowej infrastruktury informacyjnej oraz parametry do jej badania; wykazano potencjalne utrudnienia w takich badaniach oraz możliwy wpływ rozwoju infrastruktury na praktyki badawcze.

## Słowa kluczowe:

Badania humanistyczne. Cyfrowa infrastruktura informacyjna. Kompetencje informacyjne. Świadomość informacyjna.

*Otrzymano: 27 maja 2017. Zrecenzowano: 28 listopada 2017. Poprawiono: 30 listopada 2017.*

*Zaakceptowano: 28 grudnia 2017.*

---

## 1. Wstęp

Pojęcie „infrastruktura informacyjna” weszło do naukowego obiegu w latach 90. XX w. Powstała wtedy koncepcja infrastruktury informacyjnej, ale jako konstrukt politycznego – infrastruktura państwa i administracji (Report, 1994). Termin ten rozpowszechnił się za

sprawą dokumentów programowych OECD, G-7, Unii Europejskiej i rządu Stanów Zjednoczonych, poświęconych zmianom cywilizacyjnym prowadzącym do kształtowania się społeczeństwa informacyjnego. W rzeczywistości naukowej o infrastrukturze informacyjnej zaczęto wspominać w kontekście rozwoju Internetu. Z definicji mających zastosowanie do opisywania tejże rzeczywistości warto przytoczyć tę, którą sformułował John P. Pironti (2006). Do infrastruktury informacyjnej zaliczył on wszystkich ludzi, procesy, procedury, narzędzia, urządzenia i technologie, które uczestniczą w procesie tworzenia, użytkowania, przekazywania, przechowywania i niszczenia informacji. Dla rozważań zaprezentowanych w niniejszej pracy istotna jest także definicja Claudio U. Ciborri i Ole Hansetha (1998, 305–327), z której wynika, że infrastruktura informacyjna może kształtować nie tylko procedury pracy i metody działania, lecz także sposoby postrzegania tych praktyk przez ludzi. Wpływa też na to, że pracownicy i obywatele traktują te praktyki jako „naturalne”, nadaje im walor niezbędności, konieczności. W tej definicji infrastruktura informacyjna jawi się jako czynnik wspierający akceptację i osobistą implementację określonych rozwiązań i praktyk organizacyjnych.

W realiach polskiej nauki można odnieść wrażenie, że definiowanie infrastruktury informacyjnej w początkowym okresie kształtowania się tego pojęcia opanowali głównie informatycy i ekonomiści. Michał Goliński (1997, 6–7) napisał, że jest ona techniczną i technologiczną częścią krajobrazu informacyjnego współczesnego społeczeństwa. Całość tych rozwiązań opiera się na osiągnięciach w obszarze informatyki, telekomunikacji i mediów elektronicznych. W jego koncepcji infrastruktura informacyjna jest pojęciem dosyć wąskim, obejmującym jedynie najnowsze rozwiązania – sferę cyfrową. Szerszą definicję, nieograniczającą się wyłącznie do współczesnych rozwiązań cyfrowych, kilka lat później podał Józef Lubacz (2002, 55). Uznał on, że infrastruktura informacyjna obejmuje oprócz techniki i technologii, także rozwiązania instytucjonalne, regulacje prawne i usługi świadczone z ich wykorzystaniem. Szeroką definicję stworzył także Józef Oleński (2006, 270–271). Ujął w niej kompleks instytucji, jednostek organizacyjnych, systemów, technologii oraz zasobów informacyjnych, z podziałem na infrastrukturę informacyjną społeczną, polityczną, gospodarczą, regionalną, państwową i globalną. Z kolei według Bolesława Szafrąńskiego (2006, 11), infrastruktura informacyjna jest zbiorem procedur, modeli, systemów, zasobów informacyjnych oraz ludzkich, których zadaniem jest zgodne z prawem gromadzenie, przechowywanie, przetwarzanie i udostępnianie informacji w celu realizacji dowolnych zadań.

W początkach XXI w. pojawił się termin „cyberinfrastruktura” (Bowker et al., 2010). Obejmuje on sprzęt i oprogramowanie komputerowe, sieci informatyczne, zbiory danych, a także cyfrowe czujniki, narzędzia i usługi oraz odpowiednio przeszkolony personel. Kilka lat wcześniej Ole Hanseth (2002) słusznie zauważył, że rozwiązania informatyczne, tworzące współczesną infrastrukturę informacyjną, pod wieloma względami znacznie różnią się od tradycyjnych systemów informacyjnych. Co prawda nowa, cyfrowa infrastruktura rozwijana jest poprzez rozszerzanie, modyfikowanie i ulepszanie rozwiązań istniejących, ale ich digitalizacja i cyfryzacja stwarza zupełnie nowe możliwości w zakresie dostępności i potencjalnych możliwości wykorzystania. Na dodatek digitalizacja i cyfryzacja stwarzają samonapędzający się mechanizm generujący kolejne rozwiązania, których kształtu i znaczenia dziś nie możemy nawet przewidzieć, a z którymi niewątpliwie będziemy mieli do czynienia z powodu rosnącego tempa pojawiania się nowości.

W nauce o informacji infrastruktura informacyjna mieści się w trzech innych pojęciach: „środowisko informacyjne”, „infosfera” i „przestrzeń informacyjna”. Przeglądu sposobów ich interpretowania dokonała Małgorzata Kisilowska (2011, 35–52). Uznała, że pojęcie środowisko informacyjne jest używane najczęściej intuicyjnie, ale można wyróżnić dwa sposoby jego definiowania. Pierwszy kładzie nacisk na sferę techniczną i wskazuje na zespół narzędzi i zasobów umożliwiających użytkownikom tworzenie własnych aranżacji przestrzeni informacyjnych. Drugie podejście wskazuje na istnienie zbioru osób, organizacji i systemów, w którym dochodzi do gromadzenia, przetwarzania i rozpowszechniania informacji. Jeżeli chodzi o pojęcie „infosfera”, to Kisilowska uznała, iż jest ono utożsamiane z cyfrowym światem informacji, z techniką umożliwiającą tworzenie rzeczywistości wirtualnej. W przypadku przestrzeni informacyjnej wykazała istnienie cech wspólnych dla różnych definicji. Zaliczyła do nich: wyraźnie zarysowane granice, które mają charakter indywidualny; istnienie centralnego obiektu w postaci aktywnego użytkownika lub elementu nieożywionego – źródła informacji, który decyduje o kształcie i jakości tejże przestrzeni oraz relacje człowieka z przestrzenią informacyjną, które zależą od jego wiedzy i kompetencji informacyjnych. Wskazała na istnienie wąskiej i szerokiej interpretacji tego pojęcia. W wąskim znaczeniu przestrzeń informacyjna jest zbiorem źródeł informacji wykorzystywanych przez poszczególne osoby lub zbiorem informacji generowanych przez jedno źródło. Szersza interpretacja zakłada istnienie środowiska pracy systemów informacyjnych i kształtowania się przestrzeni relacji z użytkownikami informacji. Kisilowska dostrzegła fakt, że trzy omawiane pojęcia używane są synonimicznie, z tym, że za najważniejsze dla nauki o informacji, ze względu na specyfikę badań, uznała pojęcie przestrzeni informacyjnej. Jednakże, bez względu na to, którego z tych pojęć używamy, to zawsze mamy na myśli także infrastrukturę informacyjną. Bowiem każde z nich obejmuje zbiory treści i ich nośniki.

## 2. Cel i metodyka badań

Przedstawiony przegląd definicji wykazuje, że rola infrastruktury informacyjnej w działalności informacyjnej i wiedzytwórczej, także w nauce, jest kluczowa. Co więcej, zdaje się rosnać w związku z procesami cyfryzacji obejmującymi coraz większe obszary ludzkiej aktywności, także badawczej. Stąd też badania infrastruktury informacyjnej, zwłaszcza jej cyfrowej części, wydają się istotne dla zrozumienia procesów naukowych i zachodzących w nich zjawisk oraz pojawiających się tendencji. Wagę tych badań powiększa fakt, iż szybko rozrastające się środowisko informacyjne stwarza dla badaczy nie tylko wiele szans, lecz także zagrożeń. Ustalenie wpływu konkretnych rozwiązań infrastrukturalnych na wspomniane szanse i zagrożenia jest istotne dla sprawnego rozwoju nauki. Dlatego też autorzy niniejszego tekstu postawili sobie za cel zbadanie przekształceń zachodzących w cyfrowej infrastrukturze informacyjnej humanistyki, opracowanie jej modelu oraz ustalenie widocznego już obecnie i przewidywanego wpływu tych zmian i kształtującego się modelu na praktyki badawcze i publikacyjne badaczy. Celem pracy jest także zaproponowanie metodyki przyszłych badań nad cyfrową infrastrukturą informacyjną.

Podstawą badań było pozyskanie danych dotyczących poszczególnych składników cyfrowej infrastruktury informacyjnej oraz danych o wykorzystywaniu tejże infrastruktury przez humanistów. Pierwszą część zadania zrealizowano poprzez eksplorację Internetu.



Druga część sprowadzała się do badań ankietowych przeprowadzonych w 2015 r. wśród historyków, językoznawców oraz biblio- i informatologów z ośrodków uniwersyteckich w Lublinie, w Poznaniu, w Toruniu i w Warszawie (w badaniach wzięły udział 62 osoby). Uznano, że uczelnie te, różniące się wielkością, pozycją w rankingach, a także położeniem geograficzno-kulturowym, utworzyły względnie reprezentatywną próbę. Z kolei wybrane dyscypliny nauk humanistycznych charakteryzuje stosunkowo duże uzależnienie od piśmiennictwa i znaczny potencjał oferowany badaczom przez Internet. Pytania zostały tak sformułowane, by dostarczyć danych na temat wykorzystywania w codziennej praktyce określonych składników infrastruktury, znajomości stanu cyfrowej infrastruktury, kompetencji do posługiwania się jej elementami oraz stanu świadomości w zakresie skali i zasobności tejże infrastruktury oraz możliwości, jakie stwarza (ankieta została opublikowana jako aneks do pracy: Górny i in., 2017). Respondenci odnosili się do gotowych wskaźników (pytania zamknięte) lub musieli podać własne informacje (pytania otwarte krótkiej odpowiedzi). Dodatkowo zebrano dane o stanie kont bibliotecznych humanistów na wybranych uczelniach za 2015 r., wychodząc z założenia, że część pracowników naukowych nie wykorzystuje bibliotecznego zasobu publikacji drukowanych, co oczywiście może wskazywać na wykorzystywanie zasobów cyfrowych. Eksploatowano także dane pochodzące z 2015 r. z serwera Biblioteki Głównej UMCS w Lublinie, pokazujące wykorzystywanie przez tamtejszych humanistów różnych składników infrastruktury, do których biblioteka zapewnia dostęp dzięki wykupionym licencjom. Zgromadzono też dane o praktykach polskich historyków, językoznawców oraz biblio- i informatologów w zakresie cytowań, pochodzące z kilkunastu czasopism (*Annales UMCS Sectio F Historia, Białostockie Teki Historyczne, Bibliotheca Nostra, Czasy Nowożytne, Kwartalnik Historyczny, Pamiętnik Literacki, Przegląd Biblioteczny, Praktyka i Teoria Informatyki i Technicznej, Res Historica, Roczniki Biblioteczne, Toruńskie Studia Bibliologiczne, Zagadnienia Informatyki i Naukowej*). Część danych pochodzi z publikacji z okresu przed gwałtownym rozwojem cyfrowej infrastruktury informacyjnej (1998–2000), a część z ostatnich kilku lat (2013–2015).

### 3. Model cyfrowej infrastruktury informacyjnej humanistyki

Eksploatacja Internetu pozwoliła stworzyć model cyfrowej infrastruktury informacyjnej humanistyki. Według autorów artykułu składa się on z sześciu grup zasobów i narzędzi wyróżnionych ze względu na przydatność do poszczególnych etapów naukowej aktywności humanistów:

- (1) Źródła informacji bibliograficznej:
  - a. międzynarodowe, wielodzinowe bazy bibliograficzne (Central European Journal of Social Sciences and Humanities, EBSCO Host, Scopus, Web of Science);
  - b. międzynarodowe, dziedziczne bazy bibliograficzne (np. Bibliograficzna baza danych do historii Europy Środkowo-Wschodniej Herder Instytut, Slavic Humanities Index, VifaOst – Wirtualna Biblioteka Wschodnia);
  - c. cyfrowe wersje drukowanych bibliografii ogólnych (np. Bibliografia Narodowa Biblioteki Narodowej, Bibliografia Estreichera, Polska Bibliografia Literacka);
  - d. cyfrowe bibliografie dziedziczne i specjalistyczne (np. ARTON – Polska Literatura Humanistyczna, Bibliograficzna Baza Danych do Historii Europy

- Środkowo-Wschodniej LitDok, Bibliografia Etnografii Polskiej, Bibliografia Historii Polskiej, Bibliografia Historii Kościoła, Bibliografia Historii Wielkopolski i Bibliografia Powstania Wielkopolskiego, Bibliografia Polskich Badań nad Antykiem i Tradycją Antyczną, iSybisław – Bibliograficzna Baza Danych Światowego Językoznawstwa Sławistycznego, bibliografie specjalne Biblioteki Narodowej, bibliografie tematyczne Centralnej Biblioteki Wojskowej, Bibliographie de Civilisation Medievale, Bibliographie Internationale de l'Humanisme et de la Renaissance, Bibliography of British and Irish History i inne wydawnictwa Brepols, European Bibliography of Slavic and East European Studies, La Base d'Information Bibliographique en Patristique – tzw. Bibliografia Ojców Kościoła, Polska Bibliografia Literacka, Philosopher's Index, Retrospektywna Bibliografia Literackiej Zawartości Czasopism XIX i XX w. – bibliografia Bara);
- e. bibliografia dorobku poszczególnych badaczy (Polska Bibliografia Naukowa, Nauka Polska – baza Ludzie Nauki, bibliografie dorobku pracowników poszczególnych uczelni, profile pracownicze – w dwóch ostatnich przypadkach dane dostępne są na stronach WWW uczelni);
  - f. międzynarodowe katalogi biblioteczne online (OPAC) obejmujące wiele bibliotek (np. DART – Europe, The European Library, WorldCat, The Karlsruhe Virtual Catalog);
  - g. polskie katalogi biblioteczne online (OPAC) obejmujące wiele bibliotek (Katalog Rozproszony Bibliotek Polskich KaRo, Narodowy Uniwersalny Katalog Centralny NUKAT);
  - h. katalogi online (OPAC) poszczególnych bibliotek (biblioteki tradycyjne – lista: <http://www.ebib.pl/biblioteki>, biblioteki cyfrowe – lista: <http://fbc.pionier.net.pl/owoc/list-libs>).
- (2) Źródła wiedzy naukowej:
- a. książki naukowe w formie e-booka dostępne w e-księgarniach, na stronach WWW wydawców, w repozytoriach (instytucjonalnych lub dziedzinowych) i bibliotekach cyfrowych;
  - b. artykuły naukowe dostępne w czasopismach elektronicznych oraz w czasopismach zdigitalizowanych i udostępnianych w bibliotekach cyfrowych;
  - c. prace naukowe nieformalnie umieszczane przez autorów w ogólnodostępnych repozytoriach typu Academia.edu, BookZZ, Research Gate i Zenodo lub na własnych stronach WWW;
  - d. pełnotekstowe bazy artykułów i książek naukowych (np. Biblioteka Wirtualnej Nauki, BazHum, Czytelnia Czasopism PAN, IBUK Libra, Central and Eastern European Online Library, ISTOR, Academic Search Complete, Cambridge Journals, De Gruyter Open, Oxford Journals, Wiley Online Library, Taylor & Francis Online, ScienceDirect);
  - e. repozytoria prac naukowych: instytucjonalne (np. Repozytorium Uniwersytetu im. Adama Mickiewicza w Poznaniu AMUR, Repozytorium Uniwersytetu Łódzkiego RUŁ, Repozytorium Publikacji Naukowych Politechniki Śląskiej i Repozytorium Dolnośląskiej Szkoły Wyższej oPUB) i dziedzinowe (np. e-Lis, PhilSci Archive).
- (3) Narzędzia do pozyskiwania i gromadzenia danych bibliograficznych oraz pełnych tekstów prac:

- a. internetowe wyszukiwarki ogólne (np. Bing, Google) i specjalistyczne (np. Google Scholar, Google Books, BASE, DRIVER, Microsoft Academic Search);
  - b. rejestry bibliotek cyfrowych i repozytoriów naukowych (np. OpenDOAR, Registry of Open Access Repositories);
  - c. agregatory danych o pracach naukowych (np. OAIster, OpenAIR, Agregator CeON);
  - d. międzynarodowe spisy czasopism (np. Ulrich's Periodical Directory, ERIH Plus);
  - e. międzynarodowe spisy czasopism open access (np. Directory of Open Access Journals, Open Academic Journals Index, JURN, Journals for Free, Electronic Journals Library);
  - f. polskie spisy czasopism (Arianta);
  - g. wewnętrzne mechanizmy wyszukiwawcze repozytoriów i bibliotek cyfrowych;
  - h. menedżery bibliografii (np. Citavi, EndNote, Mendeley, Pepers, RefWorks, Zotero).
- (4) Cyfrowe zasoby źródeł do badań humanistycznych:
- a. projekty digitalizacji dóbr kultury (np.: Brepolis, Europeana, Gallica, London Lives 1690 to 1800, Monumenta Germaniae Historica, Monasterium.net, Patrologia Latina Database, Patristica, Acta Sanctorum, e-Codices, Manuscripta Mediaevalia, Corpus Scriptorum Latinorum, Vatican Library, Znak);
  - b. kolekcje cyfrowe archiwów i bibliotek (np.: kolekcje cyfrowe archiwów państwowych w Poznaniu i Lublinie oraz w Archiwum Państwowym m. st. Warszawy, Narodowe Archiwum Cyfrowe, archiwa Ośrodka Karta, Cyfrowa Biblioteka Narodowa Polona, Cyfrowa Biblioteka Druków Ulotnych, Polonijna Biblioteka Cyfrowa, kolekcje innych bibliotek cyfrowych zrzeszonych w Federacji Bibliotek Cyfrowych, e-kolekcja polskich czasopism Biblioteki Uniwersyteckiej w Warszawie, Kolekcja Starych Druków Biblioteki Instytutu Historycznego Uniwersytetu Warszawskiego, kolekcja cyfrowa Zakładu Narodowego im. Ossolińskich, Wirtualna Biblioteka Literatury Polskiej);
  - c. zbiory źródeł i danych stworzone w trakcie realizacji projektów naukowych (np. Atlas Źródeł i Materiałów do Dziejów Dawnej Polski, Database of Medieval Nubian Texts, The Valley of the Shadow, Railroads and the Making of Modern America, Texas Slavery Project, Photogrammar, Abbreviationes Online).
- (5) Narzędzia do poszukiwania i transkrypcji cyfrowych zasobów źródeł do badań:
- a. wewnętrzne wyszukiwarki archiwów i bibliotek cyfrowych;
  - b. przeszukiwarki zasobów w bibliotekach cyfrowych (np. Carrot2);
  - c. wyszukiwarki zespołów i jednostek archiwalnych (Archives Portal Europe, Szukaj w Archiwach, ELA, IZA, SEZAM, PRADZIAD, ZoSIA);
  - d. specjalistyczne wortale informacyjne (np. European History Primary Sources);
  - e. serwisy do dzielenia się danymi badawczymi (np. FigShare, RepOD);
  - f. narzędzia do transkrypcji (np. Transcription for Paleographical and Editorial Notation, Wirtualne Laboratorium Transkrypcji).
- (6) Narzędzia do nawiązywania kontaktów i wymiany informacji:
- a. naukowe serwisy społecznościowe (np. MethodSpace, H-Net, Index Copernicus Scientist Panel, ResearcherID, Clio-online, ResearchGate, Academici);
  - b. blogi naukowe (np. Warsztat badacza, Pedagog, Wizualizacja Informacji, New Humanist, Humanist Life);

- c. serwisy humanistycznych towarzystw naukowych (np. The Humanist, British Humanist Association, The Alliance of Digital Humanities Organizations, The Association for Computers and the Humanities, The European Association for Digital Humanities, Polskie Towarzystwo Historyczne);
- d. wyszukiwarki konferencji naukowych (np. BazaKonferencji, EduFrog, Konferencja.org, Nature Events Directory, All Conferences).

#### 4. Parametry badania cyfrowej infrastruktury informacyjnej humanistyki

Prowadzenie badań nad cyfrową infrastrukturą informacyjną humanistyki i jej wpływem na praktyki badawcze wymaga przyjęcia zestawu parametrów charakteryzujących poszczególne jej elementy. Ważne jest przy tym przyjęcie określonego punktu widzenia, bowiem informatyk inaczej będzie opisywał infrastrukturę informacyjną, inaczej urzędnik administracji, a zupełnie inaczej naukowiec. Doświadczenia autorów nabyte w trakcie badania tejże infrastruktury pozwalają na zaproponowanie zbioru parametrów przydatnych z punktu widzenia humanistów. Pierwszy zestaw przydatny jest do badania wykorzystywania poszczególnych składników infrastruktury przez badaczy i składa się z trzech par: znany – nieznan; dostępny – niedostępny; przydatny – nieprzydatny.

Przed wszystkim badacz wykorzysta jedynie te składniki infrastruktury, które zna. Stopień znajomości zależy zaś od kompetencji i doświadczeń konkretnego naukowca. Rozległość nieznanego danemu humaniście części cyfrowej infrastruktury może być nawet znaczna i wynika z szybkiego rozwoju internetowych zasobów i narzędzi. Potencjalny wpływ konkretnego składnika infrastruktury na praktyki badawcze w oczywisty sposób zależy od stopnia, w jakim jest znany. Wpływ ten zależy także od faktycznej dostępności składnika infrastruktury dla danego badacza. Dostępność może być ograniczona przez kilka czynników. Część zasobów i narzędzi jest dostępna odpłatnie, a niekiedy nie może być zakupiona przez pojedynczego badacza, lecz jedynie przez organizację (np. licencje na dostęp do Scopus i Web of Science). Realną dostępność ograniczają też kompetencje informatyczne, informacyjne i językowe konkretnej osoby. Pod uwagę należy wziąć także ograniczenia prawne, w tym te wynikające z prawa autorskiego. Powodują one, że do niektórych cyfrowych prac i kolekcji dostęp jest możliwy jedynie z komputerów zlokalizowanych w instytucji przechowującej dane źródła czy wypożyczającej dane prace. Kolejne ograniczenie dostępności jest skutkiem tworzenia plików w formatach (np. DjVu), które utrudniają indeksowanie cyfrowych zasobów przez specjalistyczną wyszukiwarkę tekstów i materiałów dla badaczy – Google Scholar. Częściowo łagodzi ten problem fakt, że wiele materiałów zgromadzonych w archiwach i bibliotekach cyfrowych we wspomnianym formacie indeksuje wyszukiwarka ogólna Google, z tym, że są one ukryte, często na odległych miejscach, w setkach, a nawet w tysiącach pozycji wyników wyszukiwania. Równie istotnym czynnikiem ograniczającym dostępność jest wydawanie cyfrowych czasopism i prac zbiorowych w formie jednego pliku, co skutecznie utrudnia wyszukanie poszczególnych artykułów i rozdziałów przez wyszukiwarki.

Wykorzystywanie danego składnika cyfrowej infrastruktury zależy nie tylko od tego, czy jest on znany i dostępny, lecz także, czy postrzegany jest jako przydatny do wykonania określonego zadania. Poczucie przydatności opierać się musi na znajomości danego

zasobu i narzędzia. Dopiero wtedy możliwe jest skuteczne zastosowanie takich kryteriów przydatności, jak: zasobność i funkcjonalność merytoryczna adekwatna do wykonywanego zadania, funkcjonalność techniczna pozwalająca na sprawne wykorzystanie, a także nakład czasu, środków i sił adekwatny do rezultatu. Przykładem nieprzydatności wynikającej z tego ostatniego kryterium jest La Base d'Information Bibliographique en Patristique, w której brak możliwości wyszukiwania według określonych kryteriów. Dostępne są jedynie pliki PDF do przeglądania. Nieprzydatność może wynikać także z niewielkiej kompletności danych. Na przykład, niewielka reprezentatywność w Scopus i Web of Science środkowo-europejskiej humanistyki powoduje, że dla większości tematów badawczych te są to nieprzydatne bazy. Podobny skutek wywołuje nierozwijanie niektórych zasobów (jak np. bazy ARTON), przez co z czasem stają się coraz bardziej niekompletne.

Inny zestaw parametrów zastosujemy do badania procesów rozwojowych i zmian w cyfrowej infrastrukturze informacyjnej. W takim przypadku pomiary powinny dotyczyć:

- liczby jednostek tworzących dany zasób, jego wielkości i zmian zachodzących w określonym czasie;
- technicznego pochodzenia danego zasobu – powstał poprzez digitalizację wersji drukowanej lub jest „cyfrowo urodzony”;
- sprawstwa powstania i rozwoju zasobu – np. redakcja czasopisma, pracownicy biblioteki cyfrowej, kierownictwo instytucji, aktywność samopublikacyjna badacza;
- formatu danych i materiałów – np. PDF, HTML, DjVu, JPG;
- formy prawnej udostępniania – np. open access, licencje Creative Commons, domena publiczna;
- funkcjonalności – np. baza bibliograficzna, spis czasopism, katalog książek, abstrakty, pełne teksty prac, wyszukiwarka, oprogramowanie, archiwum, biblioteka;
- tematyki – związek z dyscypliną naukową, z problemami badawczymi.

## 5. Infrastruktura informacyjna a świadomość i kompetencje informacyjne

Wykorzystanie cyfrowej infrastruktury informacyjnej zależy od tych samych czynników, od których zależy wykorzystywanie całej infrastruktury informacyjnej nauki, czyli od świadomości i kompetencji informacyjnych konkretnego badacza. Świadomość informacyjna definiowana jest jako orientacja jednostki lub grupy w zakresie funkcjonowania w świecie informacji połączona z ogólną wiedzą na ten temat. To także rozumienie natury mediów i uświadomienie zagrożeń wynikających z bezkrytycznego poruszania się w morzu informacji (Batorowska, 2009, 73). Z kolei kompetencje informacyjne to umiejętności wyszukiwania, oceniania, selekcjonowania, interpretowania i wykorzystywania informacji, a także znajomość ich źródeł oraz sprawne posługiwanie się odpowiednimi narzędziami (Lau, 2011, 13–20).

Współczesne realia uprawiania zawodu naukowca powodują, że na wykorzystywanie elementów cyfrowej infrastruktury informacyjnej, na zaklasyfikowanie ich do przydatnych lub nieprzydatnych wpływają czynniki, które zaliczamy do sfery świadomości informacyjnej:

- prestiż, popularność i ważność w oczach badacza tytułu danego czasopisma, wydawcy, autora, źródła, zasobu;
- opinia, jaką posiada on o poziomie merytorycznym autora, czasopisma i wydawnictwa;

- oczekiwana łatwość dotarcia i skorzystania z danego zasobu i narzędzia;
- „uświadomiona konieczność” wynikająca z zasad oceny okresowej pracownika, opinii przełożonego, a także z działania „pod recenzenta”.

Również kompetencje informacyjne konkretnego badacza w sposób istotny warunkują jego możliwości wykorzystania określonych składników infrastruktury. Możemy wyliczyć kilka związanych z tym zjawisk:

- mnogość rozwiązań, inicjatyw, zasobów i narzędzi i towarzyszący temu niedostatek informacji na ich temat;
- bariery informatyczne – nie każdy format danych czy dokumentu jest równie łatwy w wyszukaniu i wykorzystaniu, specjalistyczne narzędzia i bazy danych wymagają dużej biegłości informatycznej;
- wyniki wyszukiwania dostarczane przez najbardziej popularne wyszukiwarki, Google i Google Scholar, nie są tożsame z całością zasobów internetowych. Prawdopodobnie popularność tych narzędzi wynika z przekonania, że dostarczane przez nie wyniki są satysfakcjonujące. U jego źródła tkwi niewątpliwie brak wiedzy o tym, że wyszukiwarki te nie radzą sobie z przeszukiwaniem sporej części zasobów naukowych, nie tylko płatnych i wymagających logowania się, to także problem formatów udostępniania materiałów i umieszczania wielu tekstów w jednym pliku;
- problemy z tworzeniem indywidualnych strategii informacyjno-wyszukiwawczych, od których zależy skuteczność każdego narzędzia wyszukiwawczego.

## 6. Infrastruktura informacyjna w praktyce badawczej polskich humanistów

Prowadzone przez autorów badania wykazały, że rozwój cyfrowej infrastruktury informacyjnej dopiero zaczyna wpływać na praktykę badawczą polskich humanistów. Co prawda, na poszczególnych wydziałach, które uwzględniono w analizie aktywności na kontach bibliotecznych, od 74 do 13% kont było w 2015 r. nieaktywnych, ale inne dane nie pozwalają na przyjęcie założenia, że aż tak licznej grupie humanistów do działalności dydaktycznej i naukowej wystarczają zasoby cyfrowe dostępne w sieci i nie muszą wypożyczać książek. Jednym z dowodów jest analiza porównawcza cytowań w kilkunastu czasopismach humanistycznych. Przyjęto założenie, że radykalny wzrost liczby i dostępności prac naukowych w wersji cyfrowej powinien przełożyć się na zmiany w cytowaniach. Do analizy wybrano dwa okresy – umownie nazwany „przedcyfrowym”: 1998–2000 i tzw. cyfrowy 2013–2015. Liczba sprawdzonych artykułów w obu próbach wyniosła ok. 600. Uzyskane dane wskazują na to, że strategie cytowań humanistów na razie nie uległy istotnej zmianie, a skutki cyfryzacji zaczynają do nich dopiero docierać. Wyraźny wzrost liczby cytowań na jeden artykuł (u historyków oraz biblio- i informatologów) może być efektem poprawy dostępności informacji o publikacjach i systematycznej digitalizacji samych publikacji. Jednakże u historyków powoływanie się w przypisach na prace dostępne w Internecie nie jest jeszcze popularne – jedynie w 10% badanych artykułów z okresu 2013–2015 występują przypisy internetowe, ale dotyczą głównie materiałów źródłowych, a nie prac naukowych. Nie jest wiadomo czy historycy celowo unikają ujawniania, że powołują się na materiały znalezione w Internecie, czy rzeczywiście rzadko do nich docierają i dlatego rzadko je cytują. To jest jedno z ograniczeń badań opartych na analizach bibliometrycznych i ankietach. Nie da



się tego problemu rozwiązać pytając autora, czy celowo pomijał adresy WWW cytowanych prac, czy może z zasobów sieci nie korzystał. Skoro nie zna zasobów Internetu lub jest dla niego czymś wstydlivym czerpanie z sieci, to raczej nie przyzna się do tego i da wymijającą odpowiedź sugerującą, że w Internecie nie ma nic interesującego lub zasoby sieci nie są wiarygodne.

Przesłanką wskazującą na to, że problem sprowadza się raczej do rzadkiego korzystania przez historyków z zasobów sieciowych jest statystyka dostępu do różnych baz naukowych za pośrednictwem serwera HAN (umożliwia wejście do płatnych baz, na które BG UMCS ma licencje – zarówno przez numery IP komputerów UMCS, jak i zdalnie – po zalogowaniu na stronie głównej BG UMCS). Okazuje się, że w 2015 r. właściwie nie korzystano z takich baz jak: International Directory of Medievalists i International Medieval Bibliography. Co prawda statystyki wejść do baz wielod dziedzinowych (Ebsco, Ebrary, eJournals, IBUK Libra, Science Direct, Scopus, WoS) prezentują się pozornie (bez przeliczenia na liczbę pracowników UMCS) nie najgorzej, to jednak nie ma możliwości określenia, ile z tych wejść przypada na humanistów. Gdyby nawet przyjąć, że humaniści korzystają z tych baz tak, jak przeciętny użytkownik, oznacza to częstotliwość jeden lub dwa razy w roku. Kolejnym dowodem na nikłe wykorzystywanie internetowych zasobów i narzędzi są dane o aktywności polskich humanistów w naukowych portalach społecznościowych i w blogosferze. Od kilku lat autorzy aktywność taką stwierdzają (przy okazji różnych badań) jedynie w przypadku poniżej 1% badaczy zarejestrowanych w bazie Ludzie Nauki. Na tym tle zdecydowanie wyróżnia się aktywność redakcji czasopism naukowych. Polscy historycy wprowadzili do sieci prawie sto czasopism, których przynajmniej jeden kompletny rocznik udostępniają na zasadach open access. Obok czasopism tradycyjnych, które, oprócz wydawania wersji drukowanej, założyły własne strony WWW i zamieszczają na nich pliki PDF z artykułami lub całymi numerami, pojawiły się także czasopisma w pełni cyfrowe, wydawane jako typowy serwis internetowy z artykułami w formacie HTML (np. *Kultura i Historia*, *Prace Naukowe Akademii im. Jana Długosza. Zeszyty Historyczne*). Powstają też specjalne platformy czasopism bazujące na Open Journal Systems (np. Czytelnia Czasopism PAN, Platforma Annales Universitatis Paedagogicae Cracoviensis, Platforma Czasopism Uniwersytetu Papieskiego Jana Pawła II w Krakowie, Akademicka Platforma Czasopism UMK w Toruniu, i poznańska PRESSto), na których aktualnie dostępne są czasopisma humanistyczne.

Badania sondażowe przeprowadzone przez autorów wykazały istnienie jeszcze innych zjawisk związanych z wykorzystywaniem przez humanistów cyfrowej infrastruktury badawczej:

- większość nie zna specjalistycznych narzędzi do wyszukiwania i zarządzania informacją (poza Google Scholar i Google Books, katalogami OPAC bibliotek, serwisu FBC, katalogami i bibliografiami Biblioteki Narodowej oraz NUKAT), słabo orientuje się w stanie cyfrowej infrastruktury informacyjnej;
- większość nie korzysta z możliwości badawczych stwarzanych przez zasoby i narzędzia Internetu (w analizowanej grupie badania oparte na sieci prowadzą pojedyncze osoby);
- duża liczba nie ma pozytywnego stosunku do publikowania wyłącznie cyfrowego, zwłaszcza w formule open access (61%). Wykazuje też nieufność wobec takich publikacji;
- duża liczba nie dzieli się wiedzą w sieci (53%), nie recenzuje, nie nawiązuje tam współpracy;
- znacząca część inicjatyw zmierzających do dzielenia się wiedzą w Internecie jest nie wykorzystywana na szerszą skalę z powodu błędów informatycznych i informacyjnych.

Najprawdopodobniej źródłem tych zjawisk są niskie kompetencje i określona świadomość informacyjna badaczy, zwłaszcza starszej generacji. Potwierdzeniem tej konstatacji mogą być badania dotyczące wykorzystywania Internetu przez naukowców, przeprowadzone w Czechach. Wykazały one, że tamtejsi przedstawiciele nauk społecznych i humanistycznych, co prawda dostrzegają znaczenie Internetu dla nauki, lecz głównie odczuwają związane z tym ograniczenia i trudności typu: ryzyko plagiatów, problemy z krytyką naukową zasobów internetowych, brak szkoleń w zakresie wykorzystywania nowoczesnych technologii oraz brak właściwej polityki naukowej państwa (Stellner & Vokoun, 2014). Podsumowując, należy wyraźnie podkreślić, że aktualnie istnieje bardzo duża rozbieżność pomiędzy naukowymi możliwościami stwarzanymi przez cyfrową infrastrukturę informacyjną, a wykorzystywaniem ich przez polskich humanistów.

## 7. Mechanizmy wpływu infrastruktury informacyjnej na przyszłe praktyki badawcze w humanistyce

Upowszechnienie wśród badaczy wiedzy na temat stanu infrastruktury informacyjnej oraz możliwości i ograniczeń, jakie ona stwarza może (i powinno) prowadzić do zmiany niektórych praktyk badawczych. Reformy polskiej nauki przyczyniły się do wzrostu zainteresowania problematyką obiektywizacji oceniania jakości prac naukowych, a także aktywności naukowej badaczy i jednostek naukowych. W tym celu zaczęto wykorzystywać wskaźniki bibliometryczne oparte na danych o liczbie publikacji oraz cytowaniach. Jednakże, biorąc pod uwagę aktualny stan rozwoju infrastruktury informacyjnej humanistyki, wydaje się, że badania oceniające stan i prognozujące kierunki rozwoju przynajmniej tej dziedziny nauki, oparte na bibliometrii, zarówno te ukierunkowane na analizę cytowań, jak i te oparte na analizie liczby publikacji naukowych przyporządkowanej do określonych badaczy i specjalności naukowych, nie przyniosą wartościowych rezultatów, o ile nie zostaną w nich uwzględnione uwarunkowania wynikające ze stanu tej infrastruktury. Wykorzystywanie jedynie metod obliczeniowych typowych dla matematyki, informatyki, statystyki matematycznej czy ekonometrii nie pozwala bowiem ani na dostrzeżenie wpływu infrastruktury informacyjnej na zachowania badaczy, ani na poprawne zdiagnozowanie ich dorobku. Podkreślić należy, że podstawą metodyki badań opartej na analizie cytowań jest funkcjonujący w socjologii wiedzy naukowej tzw. paradygmat normatywny, będący uzasadnieniem dla badania rozkładów cytowań. Wśród głównych założeń tego paradygmatu Linda Smith (1981) wymieniła następujące:

- naukowiec przywołuje w bibliografii załącznikowej prace, które rzeczywiście wykorzystał;
- cytowanie odzwierciedla znaczenie i wpływ cytowanej pracy;
- cytuje się prace najlepsze z istniejących na dany temat.

Jeżeli weźmiemy pod uwagę motywacje, którymi w czasach presji publikacyjnej i „terroru” wskaźników bibliometrycznych kieruje się wielu badaczy, dobierając bibliografię oraz możliwości w tym zakresie stwarzane przez infrastrukturę informacyjną, to dojdziemy do wniosku, że paradygmat normatywny jest w dużym stopniu oparty na naiwnych założeniach. Dla nikogo znającego realia nauki nie jest tajemnicą, że w bibliografii umieszcza się nie tylko prace wykorzystane, uznane za przydatne i bardzo dobre, lecz także takie, które umieścić

wypada ze względu na jej autora (przełożony, recenzent, kolega) lub chwilową popularność danej pozycji (nie wypada jej nie uwzględnić). Odrębnym problemem są tzw. samocytowania, które nie zawsze mają merytoryczne uzasadnienie. Cyfrowa infrastruktura informacyjna stwarza nowe możliwości w zakresie odchodzenia od założeń paradygmatu normatywnego. Dzięki internetowym wyszukiwarkom stosunkowo łatwo natrafić na dane bibliograficzne pracy, która w konkretnych badaniach powinna zostać wykorzystana. Jednakże nie jest to równoznaczne z dostępem do pełnego tekstu tej pracy. Może on bowiem być udostępniany jedynie za stosunkowo wysoką opłatą lub jedynie w sieci komputerowej konkretnej biblioteki czy uczelni. Może też istnieć wyłącznie w formie drukowanej niedostępnej w danym kraju. Rodzi to pokusę umieszczenia pracy w bibliografii, mimo że jej tekst nie był wykorzystany. Tendencję tę wspiera fakt, iż dzięki licznym otwartym repozytoriom naukowym i czasopismom open access możemy uzyskać dostęp do prac omawiających interesujący nas artykuł czy książkę. Możliwe jest włączenie tego opisu do tworzonoego tekstu w sposób stwarzający wrażenie, że skorzystało się z pracy oryginalnej.

Wspomniany rozwój otwartych repozytoriów naukowych oraz czasopism open access generuje jeszcze jedną pokusę sprzeczną z zasadami paradygmatu normatywnego. Szybki i łatwy dostęp do zgromadzonych tam bardzo już licznych prac naukowych może zniechęcać do korzystania z tekstów trudno dostępnych. W takiej sytuacji bibliografia załącznikowa będzie zbiorem nie tyle prac wartościowych i przydatnych dla danego tematu badawczego, ile prac łatwo dostępnych. Rośnie też ryzyko rozpowszechniania się treści niezwyfikowanych naukowo, gdyż w repozytoriach nieinstytucjonalnych (np. Academia.edu) każdy może umieścić tekst przypominający pracę naukową. Poza tym dużą część upubliczniętych tam prac stanowią wersje robocze, przed recenzjami, a więc mogące zawierać błędy i niedociągnięcia. Nie do końca można być pewnym także naukowej wartości niejednego internetowego czasopisma, stosującego co prawda otwarty dostęp dla czytelników, ale pobierającego opłatę od autorów. Czy właściciele takich czasopism oprą się pokusie zwiększania zysków za cenę liberalizacji zasad recenzowania? Rozwój zasobów open access powoduje, że szczególnie ważne stają się wiedza i kompetencje czytelnika, warunkujące sprawne ocenianie wartości znalezionych materiałów. Słuszne obawy rodzi łatwość plagiowania prac udostępnionych w Internecie, ale równie łatwe jest wykrycie tego nadużycia (np. dzięki możliwościom Dupli Checker – <https://www.duplichecker.com/>). Jednakże wymogi wielu amerykańskich i europejskich uczelni oraz grantodawców, a także polityka Unii Europejskiej wywierają na tyle silną presję na badaczy, że mimo wspomnianych problemów, można być pewnym dalszego, szybkiego rozwoju zasobów prac naukowych dostępnych w Internecie na zasadach open access.

Istnienie licznych baz danych bibliograficznych oraz informatycznych narzędzi wyszukiwawczych i analitycznych skutkuje rozwojem badań ilościowych i tworzeniem na ich podstawie różnorodnych map i diagnoz nauki. Jednakże planując tego typu badania naukowe i bibliometryczne należy uwzględnić fakt, że aktualny stan cyfrowej infrastruktury informacyjnej humanistyki stwarza poważne ryzyko popełnienia błędów na etapie gromadzenia danych do badań ilościowych opartych na metodach matematycznych. Dowodem niech będą badania (Osiński, 2014; 2015), które wykazały, że istniejących bibliograficznych baz danych i indeksów czasopism nie można uznać za w miarę kompletne i wiarygodne źródła informacji o stanie światowej historiografii i dorobku poszczególnych historyków, gdyż zawierają istotne luki i nieścisłości. Powszechnie wykorzystywane wyszukiwarki

Google i Google Scholar, nie są zaś niezawodnymi dostawcami danych o dorobku z zakresu historiografii, co spowodowane jest błędami o charakterze informatycznym, popełnianymi przez redakcje czasopism, pracowników repozytoriów oraz przez samych autorów prac. Tak więc wszelkie diagnozy i mapowania nauki powinny być wolne od przekonania, że dzięki Internetowi i metodom ilościowym można je zrealizować szybko i obiektywnie.

Ważnym aspektem rozwoju cyfrowej infrastruktury informacyjnej humanistyki jest budowa bibliotek i archiwów cyfrowych, a także repozytoriów danych badawczych. Humanisci znajdują w tych miejscach duże zbiory źródeł i danych do badań, często dostępnych online bez żadnych ograniczeń. Skorzystanie z nich jest równie łatwe, jak z cyfrowych prac naukowych. W związku z tym możliwe jest pojawienie się tendencji do wybierania takich tematów badawczych, które można zrealizować dzięki źródłom i danym dostępnym online. W związku z tym może zaznaczyć się wpływ doboru zbiorów danych i materiałów przeznaczonych do digitalizacji na podejmowaną tematykę badań. Wpływ bibliotecznych i archiwalnych decydentów na wybór kolekcji i dokumentów do digitalizacji i udostępniania stanie się czynnikiem warunkującym tematykę, a nawet pośrednio i rezultaty badań oraz powstawania wiedzy. Ten sam problem wystąpi w przypadku aktywności tych badaczy, którzy udostępniają zbiory danych. Naukowiec bowiem może sądzić, że ma do czynienia z kompletem danych i źródeł, a *de facto* będzie miał do czynienia z ich selekcją dokonaną na podstawie sondowania największego zainteresowania potencjalnych użytkowników, chęcią zdigitalizowania i udostępniania danego zbioru. Istnieje też ryzyko dekontekstualizacji materiału, który oderwany od konkretnej kolekcji oraz od specyfiki wytwórcy, może zafałszowywać przekaz informacji, tworzyć realia niezgodne ze stanem faktycznym. Z drugiej strony, wiele kolekcji cyfrowych zawiera filmy, zdjęcia i nagrania audio, dzięki czemu te materiały mają szansę na, częstsze niż do tej pory, wykorzystanie w charakterze źródeł do badań. Na przeszkodzie intensywniejszemu wykorzystywaniu tego elementu infrastruktury informacyjnej stoi Ustawa o prawie autorskim i prawach pokrewnych, zgodnie z zapisami której część materiałów zdigitalizowanych w bibliotekach i archiwach cyfrowych, powstałych po 1939 r., udostępnianych jest jedynie na komputerach znajdujących się w danej instytucji, a więc nie ma do nich łatwej i szybkiej ścieżki dotarcia właściwej dla kolekcji dostępnych online. Rozwój technologiczny stwarza nadzieję na to, że coraz większa część źródeł do badań udostępniana będzie w formie standardowego skanu (pliku graficznego) i w dodatkowej formie tekstowej, opartej na odpowiednim schemacie XML. Każdy element treści może być wtedy opisany właściwym znacznikiem (tagiem), pozwalającym odpowiednim programom rozpoznać znaczenie treści przetwarzanych dokumentów i na tej podstawie tworzyć różnorodne zestawy danych. Dostrzec też można powszechne dążenie do tworzenia kolekcji z funkcją przeszukiwania pełnotekstowego oraz do umożliwiania robotom Google indeksowania zasobów bibliotek cyfrowych. Stworzy to nową jakość, jeżeli chodzi o wykorzystywanie infrastruktury informacyjnej przez humanistów. Należy jednak zwrócić uwagę na fakt, że część cyfrowych zasobów stworzona jest w sposób zniechęcający do korzystania – niewyraźne skany, brak wewnętrznych mechanizmów wyszukiwawczych i przeszukujących (przeszukiwanie pełnotekstowe), konieczność żmudnego przeglądania zdigitalizowanych czasopism, książek i dokumentów na zasadzie – strona po stronie, bardzo wolne ładowanie się stron i plików. Badacze mogą ograniczać swoją aktywność (wybór literatury, tematu badań i źródeł) do zasobów, które takich problemów nie stwarzają.

Zapełnianie Internetu różnymi przejawami kulturotwórczej działalności człowieka (biblioteki, archiwa, muzea i media cyfrowe, media społecznościowe, blogosfera, kanały YouTube, serwisy naukowe i edukacyjne oraz inne podobne składniki tzw. cyberkultury) staje się motorem rozwoju humanistyki cyfrowej. Badacze uzyskują łatwy i tani dostęp do rezultatów umysłowej działalności człowieka, w tym do wielu narzędzi, także bezpłatnych, do wyszukiwania i automatycznego analizowania tychże rezultatów. Możliwe jest pojawienie się tendencji do preferowania tematyki o tyleż nowatorskiej, o ile przyciągającej ze względu na to, że duże ilości nowych oraz łatwo dostępnych źródeł i danych pojawiają się każdego dnia. W takich państwach jak Stany Zjednoczone, a także w Europie Zachodniej już dziś obserwuje się zmiany w aktywności wielu humanistów. Angażują się w tworzenie cyfrowych zasobów źródeł do badań (digitalizacja), różnych baz danych i systemów wyszukiwawczych. Przystąpili do wykorzystywania cyfrowych narzędzi do rozwijania nowej metodologii nauki i nowych zwyczajów publikacyjnych. Te najnowsze tendencje charakteryzują się łączeniem analiz jakościowych z ilościowymi, interdyscyplinarnością i multidyscyplinarnością, współpracą w zespołach, przyjęciem zasad open access i open source, publikowaniem prac multimedialnych, a także zainteresowaniem się cyberkulturą. Dostrzega się istnienie specyficznych, cyfrowych pól badawczych, do eksploracji których naukowcy wykorzystują technologie komputerowe dla znalezienia odpowiedzi na tradycyjne i nowatorskie pytania humanistyki. Cyfrowi humaniści poddają komputerowej eksploracji wielkie, zdigitalizowane korpusy tekstów, filmów, zdjęć itp. w celu poszukiwania nieodkrytych do tej pory regularności, trendów i prawidłowości. Tworzą mapy nauki, bazując na danych z różnorodnych baz bibliograficznych. Analizują słowne i obrazowe opisy rozwoju państw, miast, zabytków, instytucji, podróży, bitew itp., czyli dane jakościowe. Wzbogacają je komputerowo przetworzonymi danymi statystycznymi (ilościowymi). Następnie na tej podstawie generują naukowe syntezы w formie animacji lub cyfrowych opowieści. Badają przejawy internetowej cyberkultury oraz sieciowej aktywności ludzi. Wydaje się jednak, że te nowe metody i pola badawcze dostępne są jedynie niewielkiej grupie humanistów. W przypadku większości z nich problemem może być sprawne wykorzystywanie programów będących w stanie analizować duże zbiory danych, a także stosowanie programów do automatycznej analizy tekstów, obrazów, filmów, muzyki. Bowiern świadomość ich istnienia i umiejętność wykorzystywania jest niewielka, przynajmniej poza wąskim środowiskiem cyfrowych humanistów. Nieumiejętne posługiwanie się takimi narzędziami i zasobami, a także braki w opanowaniu odpowiednich metod badawczych, mogą przyczynić się do powstawania wiedzy niewiele mającej wspólnego z nauką. Bariern w wykorzystaniu cyfrowych zasobów przez innych badaczy niż tych zaliczających się do cyfrowych humanistów, stanowią nie tylko niskie kompetencje informatyczne, lecz także nienadążanie rozwoju metodologii nauk humanistycznych za rozwojem cyfrowych narzędzi i zasobów.

Od połowy XX w. zauważyć można przemiany w medialnych i naukowych komunikatach związanych z problematyką nauką. Z liniowym tekstem zaczęły konkurować prezentacje infografik, zdjęć i filmów. Tendencje te nasiliły się wraz z rozwojem mediów cyfrowych, w tym Internetu, a w szczególności wraz z powstaniem cyfrowej humanistyki, która podjęła się badania cyfrowych wytworów człowieka. Pojawiła się nawet swoista moda na wizualizację wiedzy abstrakcyjnej, jaką w dużej mierze wytwarza większość dyscyplin humanistycznych – coraz więcej treści naukowych przybiera formę infografik, filmów i animacji. Charakterystyczne dla cyfrowych humanistów wykorzystywanie programów



komputerowych do automatycznych analiz skutkuje pojawianiem się graficznych form prezentacji wyników tychże analiz. Twierdzi się, że skala produkcji danych i informacji zaczyna wymuszać na wszystkich, także humanistach, wizualizację zarówno na etapie analizy materiału badawczego, jaki i prezentowania efektów tychże analiz. Bowiern tradycyjna narracja liniowa i przekaz słowem pisanym nie są w stanie ogarnąć i opisać zjawisk kryjących się w wielkich zbiorach danych i różnorodnych bazach informacji (Radomski, 2016). Badacze problematyki wizualizacji przekonują, że ta forma przekazu dzięki syntezie, integracji i koncentracji informacji służy skondensowanemu przekazowi wiedzy, zwielokrotniając jej potencjał interpretacyjny w porównaniu z przekazem tekstowym. Zapewniają, że u podstaw tworzenia efektywnych wizualizacji stoją wyniki badań naukowych nad percepcją wzrokową i psychologią percepcji. Twierdzą, że wizualizacje bardziej efektywnie niż tekst wykorzystują nasz mózg, czynią złożone zbiory danych łatwiej zrozumiałymi, pozwalają na ich efektywną interpretację. Dzięki prezentowaniu informacji w atrakcyjnej i innowacyjnej formie przyciągają uwagę, podtrzymują zaangażowanie odbiorcy, pozwalają mu zauważyć coś, czego inaczej nie dostrzegłby i ułatwiają zapamiętywanie informacji (Carswell & Wickens, 1987; Chen, 2006; Tufte, 2001; Ware, 2004). Biorąc pod uwagę rosnącą popularność graficznych form przekazu, a także fakt, że w społeczeństwie informacyjnym i gospodarce opartej na wiedzy znaczna część najnowszej „produkcji naukowej” powinna trafiać nie tylko do innych badaczy, lecz także do przedstawicieli wielu spośród pozostałych grup zawodowych, oczekiwać można, iż coraz większa grupa naukowców będzie decydowała się na prezentowanie wyników badań w formach graficznych.

Moda na posiadanie kont w serwisach społecznościowych i na prowadzenie blogów zapewnia rozwój nieformalnych dyskusji naukowych w Internecie. Powstały nawet specjalistyczne, naukowe serwisy społecznościowe (np. Index Copernicus Scientist Panel, MethodSpace, ResearcherID, ResearchGate). Ułatwiają one zdobywanie informacji o konferencjach, grantach, stażach, nowościach publikacyjnych, także nawiązywanie kontaktów. Jednakże wykorzystanie tych możliwości zależy wyłącznie od postawy samych badaczy. Niestety, jak przekonuje analiza aktywności polskich humanistów (Iwańska-Cieślak, 2016 a; 2016 b; Osiński, 2010), ich obecność jest tam stosunkowo niska. Biorąc pod uwagę przydatność wymienionych powyżej zasobów i narzędzi oraz względną łatwość ich wykorzystania, można podejrzewać, że wśród dużej grupy polskich humanistów istnieją silne bariery mentalne i kompetencyjne uniemożliwiające uczestniczenie w internetowej komunikacji naukowej. Jednakże należy podkreślić, iż możliwości współpracy przez Internet posiadają ogromną zaletę – przełamują bariery czasowo-przestrzenne. W związku z tym pojawiają się inicjatywy polegające na współtworzeniu projektów naukowych online (np. praca zbiorowa *Writing History in the Digital Age*, która powstała w nietypowy sposób, jako efekt oddolnej współpracy grupy autorów prowadzonej w Internecie: formę cyfrową miała zarówno współpraca – kontakty w Internecie – jak też jej efekt, czyli sama książka). Nieobecność badacza w świecie cyfrowym staje się takim samym czynnikiem ograniczającym jak nieznanomość języka angielskiego czy niski poziom kompetencji informatycznych. Internet stał się kluczowym obszarem pośredniczenia w komunikacji naukowej, więc nieobecni sami wykluczają się z tej komunikacji.



## 8. Zakończenie

Na zakończenie należy podkreślić, że stosunkowo szybki i wszechstronny rozwój cyfrowej infrastruktury informacyjnej humanistyki stwarza duże możliwości w zakresie badania tempa, kierunków, prawidłowości i skutków tego procesu. Autorzy zaproponowali informatologiczną metodykę takich badań, na którą składa się model tej infrastruktury uwzględniający realia humanistyki, wskazujący na to, co należy badać. Następnie zaproponowali dwie grupy parametrów pozwalających w miarę precyzyjnie badać poszczególne składniki infrastruktury. Wskazali też na potencjalne błędy, które mogą zostać popełnione w trakcie takich badań. A także wykazali, że cyfrowa infrastruktura informacyjna może w określony sposób kształtować przyszłe praktyki badawcze i publikacyjne humanistów.

## Bibliografia

- Batorowska, H. (2009). *Kultura informacyjna w perspektywie zmian w edukacji*. Warszawa: Wydaw. SBP.
- Bliźniuk, G., Szafranski, B. (2006). *Interoperacyjność i bezpieczeństwo systemów informatycznych administracji publicznej*. Katowice: Wydaw. PTI.
- Bowker, G.C., Baker, K., Millerand, F., Ribes D., (2010). Toward Information Infrastructure Studies: Ways of Knowing in a Networked Environment. In: J. Hunsinger, L. Klastrup & M. Allen, (eds.). *International Handbook of Internet Research* (97–117). Dordrecht: Springer Science+Business Media.
- Carswell, M., Wickens, C. D. (1978). Information Integration and the Object Display: An Interaction of Task Demands and Display Superiority. *Ergonomics*, 30(3), 511–527.
- Chen, Ch. (2006). *Information Visualization: Beyond the Horizon*. London: Springer-Verlag.
- Ciborra, C.U., Hanseth O. (1998). From Tools to Gestell. Agendas for Managing Information Infrastructures. *Information Technology & People. Special Issue on Heidegger and Information Technology*, 11(4), 305–327.
- Goliński, M. (1997). *Poziom rozwoju infrastruktury informacyjnej społeczeństwa. Próba pomiaru*. Warszawa: Akademicka Oficyna Wydawnicza PLJ.
- Górny, M., Głowacka, E., Kisilowska, M., Osiński, Z. (2017). *Mechanisms of the Formation and Evolution of Personal Information Spaces in the Humanities*. Poznań: Wydaw. Rys.
- Hanseth, O. (2002). *From Systems and Tools to Networks and Infrastructures – From Design to Cultivation. Towards a Theory of ICT Solutions and Its Design Methodology Implications*, [online] [15.06.2017], [http://heim.ifi.uio.no/~oleha/Publications/ib\\_ISR\\_3rd\\_resubm2.html](http://heim.ifi.uio.no/~oleha/Publications/ib_ISR_3rd_resubm2.html)
- Iwańska-Cieślak, B. (2016 a). Informacja o nowych publikacjach polskich bibliologów i informatologów w przestrzeni sieciowej (część I). *Toruńskie Studia Bibliologiczne*, 1(16), 211–238.
- Iwańska-Cieślak, B. (2016 b). Informacja o nowych publikacjach polskich bibliologów i informatologów w przestrzeni sieciowej (część II). *Toruńskie Studia Bibliologiczne*, 2(17), 179–200.
- Kisilowska, M. (2011). Przestrzeń informacyjna jako termin informatologiczny. *Zagadnienia Informatyki Naukowej*, 2, 35–52.
- Lau, J. (2011). *Kompetencje informacyjne w procesie uczenia się przez całe życie*. Wtyczne [online]. Stowarzyszenie Bibliotekarzy Polskich, [21.12.2017], <https://www.ifla.org/files/assets/information-literacy/publications/ifla-guidelines-pl.pdf>
- Lubacz, J. (2002). Development of Information Infrastructure. In: W. Cellary (ed.). *Human. Development Report. Poland and the Global Information Society*. Warsaw: United Nations Development Program.
- Oleński, J. (2006). *Infrastruktura informacyjna państwa w globalnej gospodarce*. Warszawa: Wydaw. UW.

- Osiński, Z. (2010). Nauka 2.0 w środowisku historyków najnowszych dziejów Polski. *Praktyka i Teoria Informacji Naukowej i Technicznej*, 4, 25–31.
- Osiński, Z. (2014). Europejskie czasopisma historyczne w bazach Scopus i Web of Science w kontekście oceny dorobku naukowego historyków w Polsce. *Zagadnienia Informacji Naukowej*, 52(2), 47–91.
- Osiński, Z. (2015). Open access w środkowoeuropejskiej historiografii – perspektywa informatologiczna. *Zagadnienia Informacji Naukowej*, 53(2), 31–64.
- Pironti, J.P., (2006). *Key Elements of a Threat and Vulnerability Management Program* [online]. ISACA, [21.12.2017], <http://iparchitects.com/wp-content/uploads/2016/07/Key-Elements-of-a-Threat-and-Vulnerability-Management-Program-ISACA-Member-Journal-May-2006.pdf>
- Radomski, A. (2016). Wizualne analizy, wizualne narracje. W: R. Bomba, A. Radomski, E. Solska (red.). *Humanistyka cyfrowa. Badanie tekstów, obrazów i dźwięku* (147–158). Lublin: Wydaw. E-Naukowiec.
- Report (1994). Report on Europe and the Global Information Society: Recommendations of the High-level Group on the Information Society to the Corfu European Council. *Bulletin of the European Union* [online], Suppl. 2, [15.06.2017], <http://aei.pitt.edu/1199/>
- Smith, L. C. (1981). Citation Analysis. *Library Quarterly*, 30(1), 83–106.
- Stellner, F., Vokoun, M. (2014). Internet, Social Sciences and Humanities. *Human Affairs*, 24, 492–510.
- Tufte, E. R. (2001). *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press.
- Ware, C. (2004). *Information Visualization: Perception for Design*. San Francisco, Morgan Kaufman.

---

## Impact of Information Infrastructure Digitization on Research Processes in the Humanities. Selected Aspects of the Phenomenon

### Abstract

**Purpose/Thesis:** Exploration of changes in the digital information infrastructure for the humanities; developing the model of this infrastructure, and describing both visible and expected impact of the model and the infrastructure itself on research and publishing behavior of the scholars; suggesting methods for future studies on digital information infrastructure.

**Approach/Methods:** Collection and analysis of data concerning selected elements of the digital information infrastructure and data on its usage by selected groups of the humanities' scholars.

**Results and conclusions:** The model of the digital information infrastructure for the humanities was developed; the parameters for exploring this infrastructure were defined; the relationship conjoining the infrastructure usage, the level of information literacy and the state of information awareness was proved to exist; the current usage of digital infrastructure by the humanities scholars was described, and the mechanisms of the infrastructure potential impact on research practices of the humanists were discovered.

**Originality/Value:** The original model of digital information infrastructure and parameters for the study of this infrastructure were developed; the potential obstacles for such research and the possible impact of the infrastructure development on research practices were identified.

### Keywords

Digital information infrastructure. Humanities' research. Information awareness. Information literacy.

---

*Dr hab. EWA GŁOWACKA – profesor Uniwersytetu Mikołaja Kopernika w Toruniu i dyrektor Instytutu Informacji Naukowej i Bibliologii UMK. Jej zainteresowania naukowe skupiają się wokół metodologii oceny*

*jakości zasobów i usług biblioteczno-informacyjnych, zarządzania jakością i wiedzą. Wybrane publikacje: E. Głowacka, M. Górny, M. Kisilowska, Z. Osiński (2015). Model indywidualnej przestrzeni informacyjnej w procesie badawczym – wstępne założenia, Zagadnienia Informacji Naukowej, 53(2), 18–30; E. Głowacka (2016). Badania wykorzystania e-booków w bibliotekach: obszary, metody, problemy. W: M. Górala, A. Wandel (red.), Metody i narzędzia badań piśmiennictwa cyfrowego i jego użytkowników (151–161). Wrocław: Wydaw. UWr; E. Głowacka (2016). Propozycja modelu badań funkcjonalności serwisów www bibliotek cyfrowych opartego na zasadach architektury informacji. Biblioteka Nostra, 2(44), 23–32; E. Głowacka (2016). Polskie repozytoria instytucjonalne jako miejsce dla otwartych zasobów naukowych i edukacyjnych. Zagadnienia Informacji Naukowej 55(1), 44–54.*

*Kontakt z autorką:*

*egt@umk.pl*

*Instytut Informacji Naukowej i Bibliologii*

*Uniwersytet Mikołaja Kopernika w Toruniu*

*ul. Władysława Bojarskiego 1*

*87-100 Toruń*

*Dr hab. MIROSŁAW GÓRNY – profesor UAM, kierownik Zakładu Systemów Informacyjnych w Instytucie Językoznawstwa UAM. Wybrane publikacje: M. Górny (2009). From the Old-Fashioned Library to the Public Library: Changes in the Cultural Functions of Polish Academic Libraries. In: W. Graves III et al., Advances in Library Administration and Organization (109–122). Emerald. Bingley 2009, vol. 27; M. Górny, P. Wierzchoń (2010). Polish Digital Libraries as a Philologists' Tools. Poznań; M. Górny, J. Mazurek (2012). Key Users of Polish Digital Libraries. The Electronic Library, 30(4), 543–556; M. Górny, J. Catlow, J. Mazurek (2015). Evaluating Polish Digital Libraries from the Perspective of Non-Academic Users. The Electronic Library, 33(4), 714–729.*

*Kontakt z autorem:*

*mgorny@amu.edu.pl*

*Instytut Językoznawstwa*

*Uniwersytet im. Adama Mickiewicza w Poznaniu*

*Collegium Novum*

*al. Niepodległości 4*

*61-874 Poznań*

*Dr hab. MAŁGORZATA KISIŁOWSKA – adiunkt na Wydziale Dziennikarstwa, Informacji i Bibliologii UW, zajmuje się zagadnieniami kulturowej funkcji bibliotek i informacji, a także kompetencji informacyjnych. Wybrane publikacje: M. Kisilowska (2016). Kultura informacji. Warszawa: Wydaw. SBP; M. Kisilowska, M. Paul, M. Zajac (2016). Jak czytają Polacy? Warszawa: Centrum Cyfrowe Projekt: Polska; M. Kisilowska, J. Jasiewicz, A. Mierzecka (2016). Relativity is the Key: The Family of Digital Competencies' Catalogues and Their Potential Applications. In: S. Kurbanoglu et al. (eds.) Information Literacy: Key to an Inclusive Society. Communications in Computer and Information Science Series, vol. 676 (294–303). Springer International Publishing AG.*

*Kontakt z autorką:*

*mdkisilo@uw.edu.pl*

*Katedra Badań nad Bibliotekami i Innymi Instytucjami Kultury*

*Wydział Dziennikarstwa, Informacji i Bibliologii*

*Uniwersytet Warszawski*

*ul. Nowy Świat 69*

*00-927 Warszawa*

*Dr hab. ZBIGNIEW OSIŃSKI – profesor w Instytucie Informacji Naukowej i Bibliotekoznawstwa UMCS w Lublinie; specjalności: informatologia, historia najnowsza Polski, metodyka i jakość kształcenia. Najnowsze publikacje: Z. Osiński (2016). Zasoby internetu jako podręcznik akademicki – studium przypadku przedmiotu organizacja i zarządzanie informacją. e-Mentor, 3, 35–45; Z. Osiński (2016). Dyskusje o polskiej*

Akademii w odbiorze sceptycznego, cyfrowego humanisty i dydaktyka. *Kronos. Metafizyka. Kultura. Religia*, 3, 54–70; M. Górny, E. Głowacka, Z. Osiński, M. Kisilowska (2016). *The Importance of the Study of Individual Information Space for Library Science*. In: *Proceedings of the 2016 International Conference on Library and Information Science (LIS 2016) Kyoto, Japan 12–14.07.2016*, (86–103). *International Business Academics Consortium, Taipei*; Z. Osiński (2016). *Historia Polski w czasopiśmie indeksowanych w bazie Scopus*. W: E. Głowacka, M. Jarocki, N. Pamuła-Cieślak (red.), *Współczesne oblicza komunikacji i informacji (191–212)*. *Przestrzeń informacyjna nauki*. Toruń: Wydaw. UMK.

Kontakt z autorem:

[zbigniew.osinski@gmail.com](mailto:zbigniew.osinski@gmail.com)

*Instytut Informacji Naukowej i Bibliotekoznawstwa*

*Uniwersytet Marii Curie-Skłodowskiej w Lublinie*

*pl. Marii Curie-Skłodowskiej 4*

*20-031 Lublin*

# Postawy i zachowania młodych naukowców – doniesienie z pierwszego etapu badań międzynarodowych, ze szczególnym uwzględnieniem polskich respondentów

Marzena Świgoń

*Uniwersytet Warmińsko-Mazurski w Olsztynie*

David Nicolas

*CIBER Research Ltd.*

---

## Abstrakt

**Cel/Teza:** W artykule omówiono wybrane wyniki międzynarodowego projektu badawczego „Harbingers”, których celem jest prognoza zmian w komunikacji naukowej na podstawie analizy postaw i zachowań współczesnych młodych pracowników naukowych. Uwagę skupiono w szczególności na wynikach uzyskanych w badaniach przeprowadzonych na grupie polskich respondentów, porównując je z wynikami uzyskanymi dla całej międzynarodowej grupy badawczej.

**Koncepcja/Metody badań:** Projekt „Harbingers” obejmuje badania prowadzone w siedmiu krajach w latach 2016–2018 metodą pogłębionych wywiadów indywidualnych z wykorzystaniem znormalizowanego kwestionariusza. Dodatkowo na zgromadzonych danych szczegółowych prowadzone są analizy statystyczne. W pierwszym etapie badań udział wzięło 116 młodych naukowców, w tym 10 z Uniwersytetu Warmińsko-Mazurskiego w Olsztynie.

**Wyniki i wnioski:** Większość respondentów z badanych krajów zwracała uwagę na przeciążenie obowiązkami zawodowymi, wysokie wymagania oraz niepewne warunki pracy. Widoczne było ich skupienie na finalnej części pracy naukowej, a mianowicie publikowaniu jak największej liczby artykułów w najbardziej prestiżowych czasopismach. Nie stwierdzono znaczących różnic w zachowaniach i postawach związanych z komunikacją naukową uczestniczących w badaniach polskich młodych naukowców, w stosunku do zachowań i postaw młodych badaczy z innych krajów.

**Oryginalność/Wartość poznawcza:** Prowadzone badania są pierwszymi na świecie długookresowymi studiami związanymi ze zmianami w systemie komunikacji naukowej wśród młodych pracowników nauki.

## Słowa kluczowe

Badania międzynarodowe. Komunikacja naukowa. Młodzi naukowcy.

*Otrzymany: 20 maja 2017. Zrecenzowany: 30 czerwca 2017. Poprawiony: 9 października 2017.  
Zaakceptowany: 28 grudnia 2017*

---

## 1. Wprowadzenie

Młodzi pracownicy nauki są przedmiotem szczególnego zainteresowania w kontekście badań nad przyszłością komunikacji naukowej. Doktoranci i doktorzy przed 35. rokiem życia (ang. *early career researchers*) są bowiem nie tylko nową, ale też najliczniejszą grupą badaczy na świecie. Komunikacja naukowa zmienia się dynamicznie, głównie za sprawą

rozwoju nowych technologii, m.in. mediów społecznościowych, społeczności wirtualnych i ruchu otwartej nauki. Interesujące i ważne wydają się obserwacje postaw i zachowań przedstawicieli tej generacji naukowców także z uwagi na fakt, że pracują oni w prekariackim środowisku, w warunkach niestalego zatrudnienia, czasowych kontraktów i niepewnej przyszłości (np. po skończeniu studiów doktoranckich).

Niniejszy artykuł dotyczy badań nad komunikacją naukową wśród młodych naukowców, prowadzonych w ramach międzynarodowego projektu *Early Career Researchers – the harbingers of change?* Celem tego projektu jest prognoza zmian, jakich można spodziewać się w komunikacji naukowej, na przykład w odniesieniu do Otwartej Nauki czy Nauki 2.0, która to idea jest wynikiem wpływu rozwoju technologii informacyjnych na globalny system komunikacji naukowej i edukacji (Nicholas et al., 2015b).

Na potrzeby projektu przyjęto umowną definicję młodych naukowców (wspomniani na wstępie doktoranci i doktorzy przed 35. rokiem życia), którą sformułowano na podstawie analizy innych określeń *early-career researcher* i jego rozumienia w światowej literaturze przedmiotu. Istnieją bowiem różne kryteria wyodrębniania „młodych pracowników nauki”, należą do nich m.in. czas pracy naukowej (pierwsze pięć lat pracy naukowej, do 10 lat po doktoracie, zob. Bazeley, 2003; Bridle et al., 2013) lub wiek (doktor do 35 lat, zob. Müller, 2014). W przyjętej definicji połączono te dwa najczęściej występujące kryteria.

W literaturze przedmiotu dotyczącej komunikacji naukowej wśród młodych naukowców podkreśla się, że są to bardzo kreatywni i pełni energii badacze, mogący odegrać kluczową rolę w ekonomii wiedzy na całym świecie (Friesenhahn & Beaudry, 2014), którzy, niestety, pracują w szczególnie trudnych, prekariackich, niesprzyjających warunkach (Corkery et al., 2013; Graham et al., 2014; Müller, 2014a, 2014b; Fransman, 2014; James et al., 2009; Waaijer, 2016). Droga do niezależności naukowej, bardzo różnie rozumiana w różnych krajach (może nią być habilitacja, jak w Polsce albo stały etat na uczelni, tzw. *tenure*, jak w Stanach Zjednoczonych), jest dzisiaj dla młodych naukowców na całym świecie bardzo niepewna. Zatrudniani są na umowy czasowe, co stało się swoistą cechą współczesnego systemu edukacji wyższej w wielu krajach (Teichler & Cummings, 2015). Niewielkie wynagrodzenia sprawiają, że zmuszeni są do podejmowania dodatkowego zatrudnienia, są przeciążeni i przepracowani (Bennion & Locke, 2010; Brechelmacher et al., 2015).

Dotychczasowe analizy postaw i zachowań młodych badaczy w komunikacji naukowej dowodzą, że stanowią oni raczej konserwatywną grupę. Jest to ściśle związane ze stawianymi im wymaganiami, a szerzej z całym formalnym systemem oceny, promocji i awansów w środowisku naukowym. Przede wszystkim odczuwają znacznie większą presję publikowania w prestiżowych czasopiśmie niż czuli ją starsi naukowcy w przeszłości (a także dzisiaj), co sprawia, że żyją pod presją czasu i w stanie permanentnego niepokoju (Müller, 2014b; Friesenhahn & Beaudry, 2014). W licznych badaniach dowiedziono, że współcześni młodzi naukowcy są niezmiernie skoncentrowani na liczbie publikowanych artykułów w czasopiśmie prestiżowych, w tym posiadających Impact Factor, oraz na liczbie cytowań, jaką osiągają ich teksty, ponieważ od tych czynników zależy zatrudnienie i dalsze etapy kariery (Harley et al., 2010; Housewright et al., 2013; Mulligan and Mabe, 2011; Van Dalen and Henkens, 2012; Waaijer, 2016). Wydaje się więc, że przed osiągnięciem niezależności naukowej raczej nie odważą się na jakieś rewolucyjne idee w zakresie przekształcania form komunikacji naukowej (Fransman, 2014; James et al., 2009; Housewright et al., 2013; Nicholas et al., 2015a; Watkinson et al., 2016).



Zważywszy jednak na możliwości, jakie w komunikacji naukowej wiążą się z rozwojem Nauki 2.0, mediów społecznościowych czy platform naukowych (takich jak Research Gate), oczywiste wydaje się, że zmiany są nieuchronne. Badacze coraz większą wagę przywiązują do widoczności w sieci (Bar-Ilan, 2012) i wykorzystują media społecznościowe na wszystkich etapach cyklu badań naukowych, od identyfikowania potencjalnych współpracowników, przez konsultowanie metod do rozpowszechniania wyników (Nicholas & Rowlands, 2011; Rowlands et al., 2011; Tenopir et al., 2013). Zmiany następują i można zastanawiać się, kiedy zostaną uwzględnione w nowych systemach oceny i promocji naukowców. Pozostaje tylko pytanie, jaki udział we wprowadzaniu zmian systemowych będą miały różne pokolenia badaczy. Wydaje się, że dużo zależy przede wszystkim od naukowców z ugruntowaną pozycją w nauce, którym nie zagraża już utrata przez lata budowanej pozycji (Harley et al., 2010). Jednakże obserwacja zachowań młodych badaczy w nauce może być ważnym przyczynikiem do określenia kierunków przyszłych zmian systemowych w komunikacji naukowej.

Poniżej przedstawiono cele i organizację badań podjętych w ramach wspomnianego projektu „Harbingers”, a także ogólne wyniki i wnioski z pierwszego ich etapu, zakończonego w 2016 r. Badania empiryczne objęły: Stany Zjednoczone, Chiny, Malesję, a w Europie: Wielką Brytanię, Hiszpanię, Francję i Polskę.

Trzeba mocno podkreślić, że w tym artykule skupiono się jedynie na wybranych aspektach i wynikach badań, a pełniejszą analizę całego przedsięwzięcia czytelnik znajdzie w raporcie z pierwszego etapu badań (CIBER Report, 2016) oraz kilku innych publikacjach zespołu, ilustrujących bliżej niektóre kwestie (Nicholas et al., 2017a, 2017b, 2017c; Rodríguez Bravo et al., 2017). Niniejszy artykuł odróżnia od publikacji zespołu w szczególności mocniejsze zaakcentowanie udziału w badaniach polskich respondentów.

## 2. Cele, organizacja i metodyka badań

Nadrzędnym celem długookresowych badań w ramach projektu „Harbingers” jest prognoza zmian w przyszłej komunikacji naukowej, uwypuklona w tytule międzynarodowego przedsięwzięcia. Zaplanowano trzy roczne fazy projektu (w 2016, 2017 i 2018 roku), po zakończeniu których będzie można stwierdzić, czy i jakie zmiany zaobserwowano oraz czy i jakich zmian można jeszcze oczekiwać. Warto podkreślić, że nikt wcześniej nie prowadził w tym zakresie długookresowych, międzynarodowych badań wśród młodych naukowców, co potwierdził przegląd literatury przedmiotu dokonany w 2015 r. na potrzeby projektu „Harbingers”. Literaturę poddaną analizie w tym przeglądzie wybiórczo przywołano we wprowadzeniu do tego artykułu, natomiast pełny stan badań w interesującym nas zakresie został przedstawiony we wspomnianym raporcie z pierwszego etapu badań (CIBER Report, 2016). W przeglądzie najnowszej literatury również nie znaleziono analogicznych studiów.

Pytania badawcze związane są z wybranymi kwestiami z zakresu komunikacji naukowej, jak m.in.: wykorzystanie mediów społecznościowych i społeczności wirtualnych przez młodych naukowców, zaangażowanie w otwartą naukę, strategia publikowania, recenzowanie i ocenianie, dzielenie się wiedzą i współpraca, budowanie reputacji, a także rola wydawców i bibliotek w komunikacji naukowej (szczegółowe pytania zawarto w raporcie i publikacjach zespołu). To właśnie m.in. w tym obszarach ściśle związanych z Nauką 2.0,

jak stwierdzono na podstawie światowej literatury przedmiotu, można spodziewać się pewnych zmian.

W badaniach wykorzystano metody jakościowe i ilościowe. Metody jakościowe obejmują wywiady indywidualne, pogłębione, z wykorzystaniem znormalizowanego kwestionariusza obejmującego łącznie 12 tematycznych grup pytań-dyspozycji. Obok zagadnień wyżej wymienionych zbierane są też informacje o dorobku naukowym i przebiegu kariery respondentów. Oprócz rozmów bezpośrednich wywiady prowadzone są także za pomocą Skype'a oraz telefonicznie. Grupą badaczy w projekcie „Harbingers” pochodzących z sześciu krajów kieruje David Nicholas (dyrektor niezależnego ośrodka badawczego CIBER Research), a w jej skład wchodzi: Anthony Watkinson i Eti Herman (także pracownicy CIBER Research, Wielka Brytania), Cherifa Boukacem-Zeghmouri (Université Claude Bernard Lyon 1, Francja), Blanca Rodríguez Bravo (Universidad de León, Hiszpania), Jie Xu (Wuhan University, Chiny), Abrizah Abdullah (University of Malaya, Malezja) i Marzena Świgoń (Uniwersytet Warmińsko-Mazurski w Olsztynie, Polska). Wymienieni badacze przeprowadzają wywiady z respondentami. W zależności od kraju wywiady są prowadzone w języku angielskim lub językach narodowych, a następnie tłumaczone na język angielski. Na potrzeby badań przygotowany został specjalny rodzaj kodowania otrzymanych odpowiedzi, czyli zestaw wskazówek, pomocniczych pytań i kwestii, na które prowadzący wywiady mają zwrócić uwagę oraz umowne symbole cyfrowe ułatwiające klasyfikowanie odpowiedzi respondentów. Trzeba powtórzyć, że opis tego narzędzia, jak i całej metodologii badań, jest dostępny w sieci w przywoływanym wcześniej raporcie.

W badaniach ilościowych wykorzystano wyżej wspomniany system kodowania (CIBER Report, 2016, 67). Analizy statystyczne polegały na obliczaniu odsetków odpowiedzi danego rodzaju, co z kolei umożliwia do pewnego stopnia, zważywszy na różnice kulturowe i w polityce naukowej różnych krajów, porównania w zakresie płci respondentów, statusu, rodzaju nauk czy pomiędzy krajami, choć z dużymi ograniczeniami. Dlatego należy podkreślić, że analizy ilościowe pełnią tylko uzupełniającą rolę, drugorzędną w porównaniu z analizą interpretatywną, która wynika z dominującego jakościowego charakteru podjętych badań i ograniczonej próby badawczej.

W pierwszej fazie wywiadów udział wzięło łącznie 116 młodych naukowców z siedmiu krajów. Dokładny rozkład respondentów przedstawiono w tabeli 1.

Tab. 1. Udział respondentów z siedmiu krajów w badanej zbiorowości młodych naukowców (źródło: <http://ciber-research.eu/harbingers.html>)

| Kraj              | Liczba respondentów | Odsetek w próbie badawczej |
|-------------------|---------------------|----------------------------|
| Stany Zjednoczone | 28                  | 24.1                       |
| Wielka Brytania   | 21                  | 18.1                       |
| Hiszpania         | 18                  | 15.5                       |
| Francja           | 14                  | 12.1                       |
| Chiny             | 13                  | 11.2                       |
| Malezja           | 12                  | 10.3                       |
| Polska            | 10                  | 8.6                        |
| Razem             | 116                 | 100.0                      |

Wszystkich młodych pracowników nauki łączyło mniejsze lub większe doświadczenie w pracy badawczej i projektowej (chodziło o udział w projektach badawczych w przeszłości i/lub w trakcie badań). Większość respondentów (76%) miała doktoraty, zdecydowanie mniejszą część całej próby stanowili doktoranci (24%), ale wśród respondentów z Polski proporcje te rozłożyły się równomiernie (po 50%). Dziesięciu polskich respondentów związanych było (w okresie prowadzenia wywiadów) z Uniwersytetem Warmińsko-Mazurskim w Olsztynie. W całej badanej zbiorowości większość młodych badaczy (urodzonych w latach 1982–1994) przekroczyła 30. rok życia (69%), co dotyczyło też respondentów z Polski (60%). Mężczyźni przeważali zarówno w całej próbie (58%), jak i w grupie polskich respondentów (60%).

Uczestnicy wywiadów reprezentowali różne dyscypliny naukowe, jednak przeważnie nauki ścisłe i biologiczne (75.8%), rzadziej nauki społeczne (24.2%); dokładny rozkład przedstawiono w tabeli 2. Polscy respondenci związani byli w zdecydowanej większości (80%) z naukami biologicznymi i rolniczymi oraz inżynieryjnymi i technicznymi, tylko dwoje reprezentowało nauki społeczne.

Tab. 2. Udział respondentów z różnych dyscyplin w badanej zbiorowości młodych naukowców (źródło: <http://ciber-research.eu/harbingers.html>)

| Rodzaj nauk                     | Liczba respondentów | Odsetek w próbie badawczej |
|---------------------------------|---------------------|----------------------------|
| Nauki biologiczne, rolnicze     | 25                  | 21.6                       |
| Nauki medyczne i o zdrowiu      | 15                  | 12.9                       |
| Nauki inżynieryjne i techniczne | 12                  | 10.3                       |
| Chemia                          | 10                  | 8.6                        |
| Informatyka                     | 10                  | 8.6                        |
| Psychologia                     | 6                   | 5.2                        |
| Inne nauki społeczne            | 22                  | 19.0                       |
| Inne nauki ścisłe               | 8                   | 6.9                        |
| Razem                           | 116                 | 100.0                      |

W kolejnej części artykułu scharakteryzowano niektóre obserwacje poczynione w pierwszej fazie badań, czyli podczas wywiadów przeprowadzonych w 2016 r. Pełniejszy opis uzyskanych rezultatów znajduje się w przywołanych we wprowadzeniu raporcie i publikacjach zespołu (prace opublikowane zawarto w bibliografii, inne pozostają w druku).

Dokonując wyboru wyników badań prowadzonych w ramach międzynarodowego projektu „Harbingers”, autorzy artykułu mieli na uwadze nie tylko troskę o niepowtarzanie omawiania rezultatów opublikowanych w języku angielskim, ale przede wszystkim skupienie uwagi na wynikach uzyskanych w badaniach prowadzonych na polskich respondentach oraz porównaniu ich z wynikami uzyskanymi dla całej, międzynarodowej próby. Celem artykułu jest też rozpropagowanie wyników rozpoczętych badań wśród polskich czytelników (temu służyła też prezentacja omawianego tu tematu na konferencji „Nauka o informacji (informacja naukowa) w okresie zmian”, zorganizowanej w Warszawie przez Katedrę Informatologii Wydziału Dziennikarstwa, Informacji i Bibliologii Uniwersytetu

Warszawskiego w maju 2017 r.)<sup>1</sup>. Tak więc poniższa część niniejszego artykułu poświęcona wynikom zespołowych badań, jest służącym tym celom wyborem dokonanych obserwacji i wyciągniętych wniosków.

### 3. Wybrane rezultaty i wnioski z pierwszego etapu badań z akcentem na polskich respondentów

Zdecydowana większość ze 116 respondentów, którzy wzięli udział w wywiadach w pierwszej fazie badań, będąc na czasowym kontrakcie bądź studiach doktoranckich, bardzo chciałaby znaleźć stałe zatrudnienie w instytucji naukowo-badawczej (ang. *tenure*). Dotyczyło to także polskich respondentów, z których tylko jedna osoba (10%) wiązała przyszłość z przemysłem, a nie z uczelnią. Większość (ogółem oraz w polskiej grupie) podkreślała swoje duże zainteresowanie pracą naukową i podjętą tematyką badawczą, wskazując je jako główną motywację zaangażowania w działalność naukową (w porównaniu do innego rodzaju motywacji, np. finansowej).

Z uzyskanych odpowiedzi w siedmiu krajach wynikało jasno, że najważniejszą częścią pracy naukowej jest dla młodych badaczy publikowanie wyników w **czasopismach posiadających wskaźnik wpływu Impact Factor (IF)** lub innych uznanych w danym środowisku lub dyscyplinie za prestiżowe. Jak wspomniano we wprowadzeniu, współcześni prekariusze nauki są mocno skoncentrowani na budowaniu swojej reputacji w oparciu właśnie o publikacje w renomowanych czasopismach, wierząc przy tym, że to one są gwarantem rozwoju ich przyszłej kariery. Z uzyskanych odpowiedzi w międzynarodowej grupie wynika, że niekiedy nie chodzi tylko o zasadę „publikuj albo gin” (ang. „*publish or perish*”) ani nawet o publikowanie w czasopismach z IF. Celem młodych badaczy jest publikowanie w czasopismach z jak najwyższym IF (sic!). Taka postawa była widoczna częściowo także w polskiej grupie, a ściślej wśród osób z doktoratami z nauk ścisłych i przyrodniczych. Dla przykładu, jedna z polskich respondentek wspomniała, że nie jest zadowolona z artykułów za dwadzieścia kilka punktów, pragnie raczej tych za czterdzieści.

Tak więc w kontekście poszukiwania oznak zmian w komunikacji naukowej i stwierdzając wyraźne nastawienie młodych badaczy na publikowanie w czasopismach z różnego rodzaju rankingów, uznać można całą międzynarodową zbiorowość respondentów raczej za naśladowców czy konserwatystów niż innowatorów. W odniesieniu do polskiego środowiska badaczy, także mocno skoncentrowanych na publikowaniu, należy dopowiedzieć, że trend publikowania na ministerialnej liście A, czyli w czasopismach rejestrowanych w bazie Web of Science, nie jest jeszcze wiodący, a przynajmniej nie dominuje w naukach humanistyczno-społecznych. Ponadto sama wysokość współczynnika IF była dużo bardziej ważna dla respondentów z Wielkiej Brytanii, Stanów Zjednoczonych, Francji i Hiszpanii niż tych z pozostałych krajów, czyli, oprócz Polski, także w Malezji i Chinach. Oczywiście, z jednej strony pewne znaczenie ma tutaj fakt istnienia w niektórych krajach preferowanych przez rząd list czasopism, ale z drugiej strony, takie listy tworzone są m.in. w oparciu o Journal Citations Reports, czyli bazę danych o znaczeniu globalnym.

<sup>1</sup> Sprawozdanie z IV Międzynarodowej Konferencji „Nauka o informacji (informacja naukowa) w okresie zmian: Innowacyjne usługi informacyjne” opublikowane zostało w ZIN 2017, 55 (1), 181–185 [red.].

Respondenci z projektu „Harbingers” okazali się płodnymi autorami, ponieważ każdy średnio opublikował do czasu wywiadu około 10 artykułów w czasopiśmie oraz sześciu tekstów w tomach pokonferencyjnych, a łącznie było to ponad 1200 tekstów w grupie 116 młodych badaczy. Zróżnicowanie w każdej grupie narodowej było dość duże, a w polskiej grupie zdarzyły się osoby z kilkudziesięcioma publikacjami, jak i bez żadnego opublikowanego tekstu (ta ostatnia sytuacja dotyczyła najmłodszych doktorantów, którzy zgłosili swoje teksty do druku).

W całej badanej populacji międzynarodowej nie było widać symptomów zmiany w odniesieniu do wiodącej od dawna pozycji artykułu w renomowanym czasopiśmie, jako podstawowym kanale formalnej komunikacji naukowej. Do specyfiki wyników uzyskanych w wywiadach prowadzonych z polskimi respondentami dodać jednak należy, że część ich dorobku stanowiły artykuły w krajowych czasopiśmie uwzględnionych w ministerialnej liście B, w języku polskim, więc właściwie nieobecnych w międzynarodowej komunikacji naukowej. Wagę podobnych list czasopism uznawanych (przez ministerstwo lub uczelnię) widać było w odpowiedziach nie tylko polskich respondentów, ale także chińskich i malezyjskich. Trzeba jednak dodać, że we wszystkich nieanglojęzycznych krajach uwzględnionych w badaniach publikowanie w języku angielskim jest normą, jeśli nie priorytetem.

Wypatrując zmian w przyszłej komunikacji naukowej wydawać by się mogło, że pozycja **czasopism w otwartym dostępie** (ang. *Open Access journals*) powinna być szczególnie wysoka wśród nowej generacji badaczy. Niestety, uzyskane wyniki – tak ogółem, jak i w polskiej grupie – w ogóle tego nie potwierdziły. Generalnie najważniejszym kryterium doboru czasopisma był dla respondentów nie otwarty dostęp, a indeksowanie w bazach: Web of Science, Scopus lub obecność na krajowych listach czasopism. W świetle wypowiedzi młodych naukowców o publikowaniu w otwartym dostępie można powiedzieć, że ich postawy wobec tej drogi komunikacji naukowej są bardzo pozytywne. Jednak zachowania młodych badaczy w tym zakresie nie są już innowacyjne, tzn. w wyborze czasopisma do publikacji swoich prac respondenci nie kierowali się zasadą otwartego dostępu. Wynika to z pewnością ze specyfiki systemów formalnej komunikacji naukowej oraz obowiązujących kryteriów oceniania badań naukowych i zasad ich finansowania. Z wyjątkiem Wielkiej Brytanii i Stanów Zjednoczonych, gdzie złota droga otwartego dostępu wspierana jest przez rządy państw i fundatorów badań, w pozostałych krajach publikowanie w otwartym dostępie – przynajmniej w świetle uzyskanych odpowiedzi – nie należy do priorytetów ani nie jest wyznacznikiem strategii publikacyjnej młodych naukowców. Zresztą także dla respondentów z Wielkiej Brytanii i Stanów Zjednoczonych priorytetem jest raczej, jak wcześniej powiedziano, obecność czasopisma w Journal Citation Reports i wysokość wskaźnika IF.

Publikowanie dużej liczby artykułów indeksowanych w WoS jest dużo łatwiejsze, jeśli pracuje się w **grupach projektowych**, często wirtualnych, w które zaangażowana była część badanych w różnych krajach, w tym w Polsce. Ponadto bycie członkiem takiej grupy, szczególnie na poziomie międzynarodowym, wpływało znacząco zdaniem respondentów na budowanie prestiżu w nauce, a w niektórych krajach było ściśle związane z ofertami zatrudnienia. Na przykład młodzi naukowcy z Francji twierdzili zgodnie, że nie tylko publikacje, ale także szeroka międzynarodowa współpraca wpływa pozytywnie na ich karierę zawodową. Co ciekawe, wśród polskich respondentów, zarówno w naukach przyrodniczych, jak i społecznych, niebędących członkami takich międzynarodowych zespołów, wyraźnie widać było ową potrzebę znalezienia zagranicznych partnerów do badań.

Zdecydowana większość badanych przyznała też, że dla budowania prestiżu i pozycji w świecie nauki ważne jest, oprócz publikowania artykułów w renomowanych czasopiśmie, także **dzielenie się wiedzą** w sieci, w mediach społecznościowych, na platformach społeczności wirtualnych, np. Research Gate. Owo dzielenie się wiedzą było jednak dosyć wąsko pojmowane, w dużej mierze ograniczało się bowiem do aktywności w sieci „po publikacji”.

Do **mediów społecznościowych** najczęściej wykorzystywanych przez badanych 116 respondentów należały: Research Gate (RG) – cieszące się największą popularnością we wszystkich siedmiu krajach, LinkedIn i Twitter – wykorzystywane w porównaniu do RG rzadziej i nie we wszystkich krajach. Wykorzystywanie *social media* polegało na ogół na udostępnianiu własnych tekstów lub informowaniu o ich publikacji. Przed ukazaniem się (bądź przyjęciem do druku) oficjalnej publikacji (w czasopiśmie) raczej nie dzielono się – w szerokim zakresie – nową wiedzą i informacjami. Innymi słowy, media społecznościowe były przez respondentów wykorzystywane do dzielenia się nowymi pomysłami czy wynikami trwających badań, ale tylko w wąskich grupach projektowych czy z najbliższymi współpracownikami.

Natomiast dosyć powszechnie wykorzystywano ogromny potencjał mediów społecznościowych związany z rozpowszechnianiem opublikowanych (w czasopiśmie i książkach) tekstów, a także z szukaniem i pozyskiwaniem informacji (informacji o publikacjach i pdf-ów). W polskiej grupie w pierwszym roku prowadzenia badań siedmiu respondentów (70%) było aktywnymi użytkownikami Research Gate, tzn. posiadało konta ze zdjęciem oraz udostępniało swoje teksty, jak też korzystało z tekstów tam udostępnianych. Dla porównania tylko trzech (30%) używało serwisu Academia.edu. Podobną dużą przewagę korzystania z Research Gate w porównaniu do Academia.edu zaobserwowano też wśród respondentów z pozostałych sześciu krajów.

Z innych mediów społecznościowych w komunikacji naukowej pewne znaczenie ma, jak wspomniano, Twitter oraz LinkedIn, głównie w Stanach Zjednoczonych, Wielkiej Brytanii oraz Hiszpanii, mniej w Chinach i Malezji. Co charakterystyczne, Twitter oraz LinkedIn nie był używany przez respondentów z Polski i Francji; za to tylko w tych krajach wspomniano o YouTube. O Facebooku w kontekście pozyskiwania informacji wspomniano jedynie w Stanach Zjednoczonych i w Hiszpanii, a o Wikipedii tylko w Malezji.

W pozyskiwaniu informacji naukowej prym wiodą, w świetle udzielonych odpowiedzi, Google, Google Scholar, Web of Science oraz Scopus (w naukach medycznych PubMed), jak też Research Gate. Wspomnieć trzeba o różnicach politycznych widocznych w odpowiedziach, np. chińscy respondenci twierdzili, że preferują krajowe źródła i komunikatory, co można traktować jako pewną postawę nacechowaną politycznie.

**Repozytoria** nie cieszyły się zbyt dużym zainteresowaniem w badanej grupie młodych naukowców, choć wydawać mogłoby się, że będą chcieli je wykorzystywać jako dobrą okazję do zwiększenia swojej widoczności w sieci. Abstrahując od różnic narodowych i w obrębie dyscyplin, można powiedzieć, że młodzi badacze umieszczają w nich swoje teksty, jeśli jest to obligatoryjne. Niektórzy twierdzili, że sami nie przeglądają ich zawartości, inni przyznawali się nawet, że nie orientują się, czy ich instytucja takowe prowadzi. Polscy respondenci nie byli tu wyjątkiem. Wprawdzie UWM nie posiada instytucjonalnego repozytorium, ale respondenci nie wskazywali też na korzystanie z innych repozytoriów. Ogółem respondenci z siedmiu krajów podkreślali, że najważniejsze dla nich jest archiwizowanie swoich



tekstów na **Research Gate**, doceniali też tę platformę jako miejsce poszukiwania pełnych tekstów z interesującego ich zakresu.

Uczestniczący w badaniach młodzi naukowcy generalnie okazali się nie być entuzjastami idei **otwartej nauki**. Niektórzy nawet nie byli zbyt dobrze zorientowanych co dokładnie wchodzi w zakres tego pojęcia. Jak już wspomniano, młodzi naukowcy nie byli zaangażowani w publikowanie w czasopismach w otwartym dostępie (ang. *open access*), w żadnym razie nie czynili z tego strategii publikacyjnej. Ponadto raczej nie byli zaangażowani w udostępnianie danych z własnych badań (ang. *open data*), ani w prowadzenie blogów.

**Wskaźniki altmetryczne** także wydają się jeszcze nie mieć dla przedstawicieli badanego pokolenia zbyt dużego znaczenia. Większość podkreślała, że głównym ograniczeniem tego rodzaju danych jest fakt, że nie są brane pod uwagę w ich uczelniach w procesie formalnej oceny ich pracy. Jednakże niektórzy przyznali, że niekiedy zwracają uwagę na liczbę ściągnięć czy udostępnień. Stwierdzili też, że wskaźniki te mają pewien potencjał w określaniu oddziaływania czy **wpływu** (ang. *impact*) danej publikacji lub autora. Jednak większość utożsamiała taki wpływ raczej z tradycyjnymi cytowaniami wynikającymi z publikowania tekstów w wysoko punktowanych czasopismach. Szczególnie miejsce zajmują tu badacze z Wielkiej Brytanii, którzy zobligowani swoim systemem Research Excellence Framework, muszą nie tylko publikować w renomowanych czasopismach, ale także walczyć o szerokie oddziaływanie w społeczeństwie, do czego wykorzystują przeróżne środki, w tym m.in. media społecznościowe.

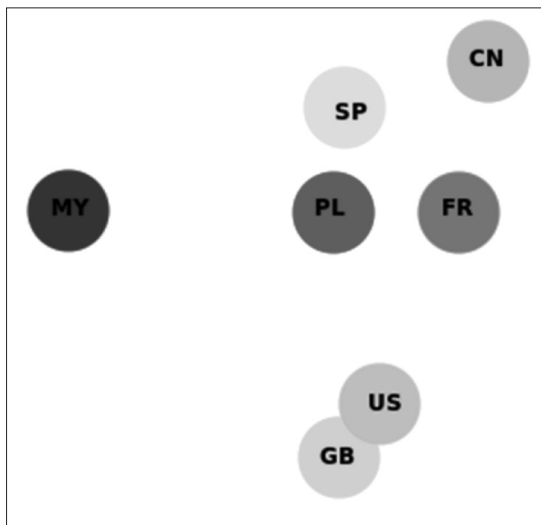
Pewnym zaskoczeniem w uzyskanych rezultatach było stosunkowo niewielkie korzystanie przez respondentów ze **smartfonów** w dzieleniu się informacją naukową. Zważywszy na popularność tego urządzenia w pokoleniu młodych osób, spodziewano się szerszego ich wykorzystywania do celów naukowych. Tymczasem okazało się, że nawet respondenci podłączeni do sieci przez całą dobę wykorzystują smartfony raczej do celów prywatnych, nie czytają pełnych tekstów naukowych, rzadko przeglądają nawet abstrakty. Najbardziej wykorzystywali je Chińczycy, a najmniej Polacy, tylko trzech polskich respondentów okazało się użytkownikami smartfonów do celów naukowych.

Innym ważnym aspektem badań była próba określenia przyszłej **roli bibliotek** naukowych w komunikacji naukowej i w pozyskiwaniu informacji naukowej. Trzeba przyznać, że respondenci, choć korzystali z platform bibliotecznych, nie przywiązywali zbyt dużej wagi do samego istnienia bibliotek, szczególnie w obliczu systematycznego korzystania z Google Scholar i Research Gate. Pewnym wyjątkiem okazali się tu polscy respondenci, którzy w przeważającej większości (osiem osób) przewidywali, że biblioteki będą nadal pełnić ważną rolę i byli świadomi, że dostęp do wielu baz, w tym Web of Science, mają poprzez stronę biblioteki uczelnianej. Ogółem jednak, dla międzynarodowej społeczności biblioteki stają się coraz bardziej niewidoczne.

Odrębną kwestią była przyszła rola **platform wydawców**. Chociaż młodzi pracownicy nauki bazują na czasopismach, rzadko wiedzą, kto jest wydawcą preferowanych przez nich tytułów i jeszcze rzadziej odwiedzają ich strony w celu znalezienia potrzebnej im informacji naukowej (regularnie korzystają z przewodników dla autorów). Nazwy wydawców pojawiły się w tym kontekście w wypowiedziach jedynie 13 ze 116 respondentów.

Projekt „Harbingers” oparty został na badaniach o charakterze jakościowym, jednak zebrany materiał po skategoryzowaniu odpowiedzi umożliwił przeprowadzenie również pewnych analiz ilościowych z wykorzystaniem metod statystycznych.

W badaniach testowano 26 hipotez (ich wykaz znajduje się w CIBER Report (2016, 62–66), które sformułowano na podstawie przeglądu literatury i konsultacji w środowisku badaczy komunikacji naukowej. Otrzymane wyniki ilustrowały m.in. podobieństwa i różnice pomiędzy odpowiedziami respondentów z różnych krajów (Rys. 1).



Rys. 1. Podobieństwa i różnice pomiędzy odpowiedziami respondentów z różnych krajów (źródło: <http://ciber-research.eu/harbingers.html>)

Analizując schemat przedstawiony na rysunku 1, nie sposób nie zwrócić uwagi na centralną pozycję symbolu PL. Świadczy to przede wszystkim o największej liczbie wspólnych cech w postawach i zachowaniach młodych polskich badaczy i badaczy z pozostałych sześciu krajów, w tym szczególnie z Hiszpanii. Wzajemna bliskość Stanów Zjednoczonych i Wielkiej Brytanii wydaje się oczywista, podobnie jak ich odległość od Chin czy Malezji oraz pomiędzy tymi krajami. Warto jeszcze raz podkreślić, że różnice narodowościowe są w całej metodologii badań drugorzędne, zbyt małe były bowiem liczebności grup narodowych, aby wyciągać uogólnione wnioski.

Niemniej warto dodać, że ogółem 20% z testowanych hipotez zostało potwierdzonych i 20% odrzuconych. Tylko dwie hipotezy zostały potwierdzone we wszystkich siedmiu krajach:

- środowisko, w którym pracują młodzi naukowcy, jest prekariackie;
- młodzi pracownicy nauki preferują system anonimowych koleżeńskich recenzji (ang. *double blind peer review*) jako najbardziej sprawiedliwych.

Kolejne trzy hipotezy zostały potwierdzone lub częściowo potwierdzone w siedmiu krajach:

- dopóki akademicy będą w tradycyjny sposób zatrudniani, oceniani i finansowani na podstawie głównie listy publikacji i wskaźnika cytowań, nie będzie możliwa zmiana zachowań związanych z komunikacją naukową;
- młodzi pracownicy nauki podporządkowują się zastanym zasadom w komunikacji naukowej;

- młodzi pracownicy nauki, którzy są zainteresowani zdobywaniem kolejnych szczebli kariery naukowej, czują się „niewolnikami” metrycznego (punktowego) systemu oceny ich pracy, ściśle związanego z publikowaniem w renomowanych czasopismach.
- Dwie hipotezy zostały odrzucone we wszystkich uwzględnionych w badaniach krajach:
- młodzi pracownicy nauki publikują w czasopismach w otwartym dostępie, ponieważ łatwiej jest się do nich dostać;
  - młodzi pracownicy nauki są zainteresowani bardziej mediami społecznościowymi i wskaźnikami altmetrycznymi niż tradycyjnymi cytowaniami, na które trzeba bardzo długo czekać.

Wiele było też widocznych różnic w zachowaniach respondentów pochodzących z poszczególnych krajach, o niektórych wspomniano już w tym artykule. Jednakże przytoczone powyżej potwierdzone i odrzucone hipotezy pokazują, jak wiele przede wszystkim łączy młodych reprezentantów różnych nauk z różnych państw. Wszędzie borykają się z podobnymi problemami funkcjonowania w prekariackim i konkurencyjnym środowisku. Dostosowują się do zastanego systemu, ponieważ najbardziej zależy im na utrzymaniu się na obranej ścieżce kariery, która zgodna jest z ich zainteresowaniami i ambicjami. Nie planują rewolucji, ponieważ nie czują, że mogą coś zmienić. Raczej koncentrują się na spełnianiu stawianych im wymagań, chociaż podkreślają, że są one zbyt wygórowane.

Kończąc prezentację wybranych rezultatów projektu „Harbingers”, trzeba wspomnieć o ograniczeniach rozpoczętego cyklu tych międzynarodowych badań. Jednym z nich jest zasadniczo niewielka grupa respondentów, która nie pozwala na uogólnianie wyników na całe pokolenie młodych badaczy, nawet gdyby poszerzyć lub zawęzić umownie przyjętą definicję (doktoranci i doktorzy do 35 lat). Reprezentacje poszczególnych krajów w próbie badawczej również są niewystarczające do wyciągnięcia wniosków o charakterze cech narodowych ani o narodowych systemach komunikacji naukowej. W każdym kraju w obrębie różnych uczelni i dyscyplin mogą występować znaczące różnice, które nie są w podjętych badaniach brane pod uwagę. Wydaje się jednak, że pionierski zamysł zespołu „Harbingers”, polegający na długookresowych (a rozważane jest przedłużenie czasu badań z trzech lat do pięciu) obserwacjach tych samych młodych naukowców, śledzenia ich losów w świecie nauki i sprawdzania opinii o kluczowych dla komunikacji naukowej kwestiach, może zaowocować ciekawymi wynikami i ważnymi wnioskami (choćaby w obliczu aktualnie dyskutowanej reformy polskiej nauki).

#### 4. Zakończenie

Z przeprowadzonych międzynarodowych badań, a ściślej pierwszej z trzech zaplanowanych ich części (CIBER Report, 2016; Nicholas et al., 2017abc; Rodríguez Bravo et al., 2017 i inne publikacje zespołu w druku), których wycinkową i fragmentaryczną charakterystykę zaprezentowano w tym artykule, wynika, że choć młodzi pracownicy nauki widzą możliwości zmiany systemu komunikacji naukowej, nie korzystają z nich w takim stopniu jak można by się spodziewać.

Możliwości te są związane, jak wspomniano we wprowadzeniu, przede wszystkim z rozwojem nowoczesnych technologii, mediów społecznościowych i wskaźników altmetrycznych, z ideą otwartego dostępu do informacji naukowej i otwartej nauki. Niestety, nie idą

one w parze ze zmianami systemu finansowania i oceniania pracy badawczej, a właściwie systemów w różnych krajach, nie tylko tych uwzględnionych w projekcie „Harbingers”.

W ubieganiu się o granty wciąż ważne są tradycyjne sposoby oceniania potencjału badaczy, czyli publikacje w najbardziej prestiżowych czasopismach oraz tradycyjne cytowania. Presja związana z takimi publikacjami, powiązana z nadmiarem innych obowiązków zawodowych, a także z niepewnością wynikającą z warunków zatrudnienia (prekariusze) na ogół nie sprzyjają zainteresowaniu młodych naukowców mniej formalnymi aktywnościami naukowymi, nie przekładającymi się bezpośrednio na ich ocenianie. Tymczasem prowadzenie blogów naukowych, dyskusji na forach jest niezwykle ważne dla idei otwartej nauki i edukacji, a ponieważ jest bardzo czasochłonne i na ogół nie liczy się w ocenie, niewielu badaczy bierze aktywny w nich udział.

Jednakże były w badanej międzynarodowej grupie osoby, które pragną zmian w komunikacji naukowej i nawet deklarują w tym zakresie swoje zaangażowanie, ale dopiero kiedy zdobędą pozycję lub stanowisko umożliwiające bardziej ryzykowne postawy i zachowania, a co za tym idzie wprowadzanie zmian. W tym kontekście polscy respondenci mieli świadomość ogromnej międzynarodowej konkurencji i konieczności pewnych reform w polskim systemie szkolnictwa wyższego (zwracali uwagę m.in. na długą drogę do samodzielności naukowej). Niektórzy przyznali się do marzeń o zatrudnieniu w zachodnich uczelniach, które postrzegali jako bardziej sprzyjające pracy naukowej (wystarczy doktorat) i byli świadomi, że teraz muszą przede wszystkim budować swoją reputację w oparciu o publikacje indeksowane w Web of Science i Scopus.

Patrząc jednak globalnie, a przynajmniej na kraje uwzględnione w niniejszych badaniach, można na zakończenie tylko powtórzyć, że w każdym z nich panuje duża konkurencja i niepewność wśród młodych badaczy.

Być może wywiady w kolejnych latach przyniosą wyraźniejszy obraz zmian, jakich możemy się spodziewać w przyszłej komunikacji naukowej.

## Źródło finansowania i współpracownicy:

Projekt Harbingers jest finansowany przez Publishing Research Consortium, a realizowany przez CIBER Research Ltd. Do pozostałych współpracowników, obok autorów artykułu, należą: Anthony Watkinson, Cherifa Boukacem-Zeghmouri, Abriyah Abdullah, Blanca Rodríguez Bravo, Jie Xu i Eti Herman.

## Bibliografia

- Bar-Ilan, J., Haustein, S., Peters, I., Priem, J., Shema, H. and Terliesner, J. (2012). *Beyond Citations: Scholars' Visibility on the Social Web* [online]. *arXiv preprint*, [22.12.2017], arXiv:1205.5611.
- Bazeley, P. (2003). Defining Early Career in Research. *Higher Education*, 45(3), 257–79.
- Bennion, A., Locke, W. (2010). The Early Career Paths and Employment Conditions of the Academic Profession in 17 Countries. *European Review*, 1These 8(S1), S7–S33.
- Brechelmacher, A., Park, E., Ates, G., Campbell, D. F. (2015). The Rocky Road to Tenure–Career Paths in Academia. In: T. Fumasoli, G. Goastellec, B.M. Kehm (eds.) *Academic Work and Careers in Europe: Trends, Challenges, Perspectives* (13–40). Springer International Publishing.
- Bridle, H., Vrieling, A., Cardillo, M., Araya, Y., Hinojosa, L. (2013). Preparing for an Interdisciplinary Future: A Perspective from Early-Career Researchers. *Futures*, 53, 22–32.

- CIBER Report (2016). *Early Career Researchers: The Harbingers of Change?* [online]. Final Report CIBER, Year 1, [9.10.2017], [http://ciber-research.eu/download/20161120-ECR\\_Year\\_1\\_final\\_report\\_071116.pdf](http://ciber-research.eu/download/20161120-ECR_Year_1_final_report_071116.pdf)
- Corkery, C., Mitchell, J., Walker, V., Annan, R., Goel, N., Harvey, L., ... Vilches, S. L. (2013). *The 2013 Canadian Postdoc Survey: Painting a Picture of Canadian Postdoctoral Scholars* [online]. CASP-ACSP and Mitacs [26.12.2017], [https://www.mitacs.ca/sites/default/files/caps-mitacs\\_postdoc\\_report-full\\_oct22013-final.pdf](https://www.mitacs.ca/sites/default/files/caps-mitacs_postdoc_report-full_oct22013-final.pdf)
- Fransman, J. (2014). *Becoming Academic in the Digital Age: Negotiations of Identity in the Daily Practices of Early Career Researchers* [online]. Connected Communities and Early Career Researchers Workshop, City University, May 2014, [9.10.2017], [https://www.srhe.ac.uk/downloads/FRANSMAN\\_Final\\_Report.pdf](https://www.srhe.ac.uk/downloads/FRANSMAN_Final_Report.pdf)
- Friesenhahn, I., Beaudry, C. (2014). *The Global State of Young Scientists – Project Report and Recommendations*. Berlin: Akademie Verlag.
- Graham, H., Hill, K., Matthews, P., O'Brien, D., Taylor, M. (2014). Connecting Epistemologies: Methods and Early Career Researchers in the Connected Communities Programme [online]. [9.10.2017], <https://earlycareerresearchers.files.wordpress.com/2014/10/connecting-epistemologies-report.pdf>
- Harley, D., Acord, S.K., Earl-Novell, S., Lawrence, S., King, C.J. (2010). *Assessing the Future Landscape of Scholarly Communication: An Exploration of Faculty Values and Needs in Seven Disciplines* [online]. UC Berkeley: Center for Studies in Higher Education. [9.10.2017], <https://escholarship.org/uc/item/15x7385g>
- Housewright, R., Schonfeld, R. C., Wulfson, K. (2013). *Ithaka S+R US Faculty Survey 2012* [online]. Ithaka S+R [26.12.2017], [http://www.sr.ithaka.org/wp-content/uploads/2015/08/Ithaka\\_SR\\_US\\_Faculty\\_Survey\\_2012\\_FINAL.pdf](http://www.sr.ithaka.org/wp-content/uploads/2015/08/Ithaka_SR_US_Faculty_Survey_2012_FINAL.pdf)
- James, L., Norman, J., De Baets, A. S., Burchell-Hughes, I., Burchmore, H., Philips, A., Sheppard, D., Wilks, L., Wolffe, J. (2009). *The Lives and Technologies of Early Career Researchers* [online]. JISC; CARET, University of Cambridge; The Open University, UK. [9.10.2017], <http://www.webarchive.org.uk/wayback/archive/20140614204612/http://www.jisc.ac.uk/publications/reports/2009/earlycareerresearchersstudy.aspx>
- Müller, R. (2014a). Racing for What? Anticipation and Acceleration in the Work and Career Practices of Academic Life Science Postdocs. *Forum Qualitative Sozialforschung / Forum: Qualitative Social Research* [online], 15(3), [22.12.2017], <http://www.qualitative-research.net/index.php/fqs/article/view/2245>
- Müller, R. (2014b). Postdoctoral Life Scientists and Supervision Work in the Contemporary University: A Case Study of Changes in the Cultural Norms of Science. *Minerva*, 52(3), 329–349.
- Mulligan, A., Mabe, M. (2011). The Effect of the Internet on Researcher Motivations, Behaviour and Attitudes. *Journal of Documentation*, 67(2), 290–311.
- Nicholas, D., Rowlands, I. (2011). Social Media Use in the Research Workflow. *Information Services and Use*, 31(1–2), 61–83.
- Nicholas, D., Boukacem-Zeghmouri Ch., Rodríguez-Bravo, B., Xu, J., Watkinson, A., Abrizah, A., Herman, E., Świgoń, M. (2017a). Where and How Early Career Researchers Find Scholarly Information. *Learned Publishing*, 30(1), 19–29.
- Nicholas, D., Jamali, H. R., Watkinson, A., Herman, E., Tenopir, C., Volentine, R., Allard, S., Levine, K. (2015a). Do Younger Researchers Assess Trustworthiness Differently When Deciding What to Read and Cite and Where to Publish? *International Journal of Knowledge Content Development and Technology*, 5(2), 45–63.
- Nicholas, D., Herman, E., Jamali, H.R. (2015b). Analysis of Emerging Reputation Mechanisms for Scholars [online]. In: R. Vuorikari & Y. Punie (eds.). *Analysis of Emerging Reputation and Funding Mechanisms in the Context of Open Science 2.0* (3–72). Part 1. European Commission, Joint Research Centre, Institute for Prospective Technological Studies. [9.10.2017], <http://publications.jrc.ec.europa.eu/repository/bitstream/JRC94952/jrc94952.pdf>

- Nicholas, D., Rodríguez Bravo, B., Watkinson, A., Boukacem-Zeghmouri, Ch., Herman, E., Xu, J., Abrizah, A., Świgoń, M. (2017c). Early Career Researchers: Their Publishing and Authorship Practices. *Learned Publishing*, 30(3), 205–217.
- Nicholas, D., Watkinson, A., Boukacem-Zeghmouri, Ch., Rodríguez Bravo, B., Xu, J., Abrizah, A., Świgoń, M., Herman, E. (2017b). Early Career Researchers: Scholarly Behaviour and the Prospect of Change. *Learned Publishing*, 30(2), 157–166.
- Rodríguez Bravo, B., Nicholas, D., Herman, E., Boukacem-Zeghmouri, Ch., Watkinson, A., Abrizah, A., Xu, J., Świgoń, M. (2017). Peer Review: the Experience and Views of Early Career Researchers. *Learned Publishing* 30(4), 269277.
- Rowlands, I., Nicholas, D., Russell, B., Canty, N., Watkinson, A. (2011). Social Media Use in the Research Workflow. *Learned Publishing*, 24(3), 183–195.
- Teichler, U., Cummings, W. K. (2015). Forming, Recruiting and Managing the Academic Profession: A Varied Scene. In: Teichler, U., & Cummings, W. K. (eds) *Forming, Recruiting and Managing the Academic Profession* (1–10). Springer International Publishing.
- Tenopir, C., Volentine, R., King, D. (2013). Social Media and Scholarly Reading. *Online Information Review*, 37(2), 193–216.
- Van Dalen, H.P., Henkens, K. (2012). Intended and Unintended Consequences of a Publish-or-Perish Culture: A Worldwide Survey. *Journal of the American Society for Information Science and Technology*, 63(7), 1282–1293.
- Waaiker, C. J., Macaluso, B., Sugimoto, C. R., Larivière, V. (2016). Stability and Longevity in the Publication Careers of US Doctorate Recipients. *PloS one*, 11(4), e0154741.
- Watkinson, A., Nicholas, D., Thornley, C., Herman, E., Jamali, H. R., Volentine, R., ...and Tenopir, C. (2016). Changes in the Digital Scholarly Environment and Issues of Trust: An Exploratory, Qualitative Analysis. *Information Processing & Management*, 52(3), 446–458.
- 

## Attitudes and Behavior of Junior Researchers – Early Findings of an International Study with a Particular Focus on Polish Researchers

### Abstract

**Purpose/Thesis:** The aim of this study is to identify the changes in scholarly communication based on the analysis of attitudes and behaviors of junior researchers.

**Approach/Methods:** The study is being conducted in seven countries over a three-year period from 2016 to 2018 using a deep interview method. During the first year of the study, 116 early-career researchers have participated, including 10 participants from the University of Warmia and Mazury in Olsztyn, Poland.

**Results and conclusions:** The majority of respondents from all countries have drawn attention to the fact that they were overwhelmed with their professional duties and unstable work environment. Their focus on publishing papers in highly-ranked journals was noticeable.

**Originality/Value:** It was first longitudinal and international study on changes in the scholarly communication among junior researchers.

### Keywords

Early career researchers. International study. Scholarly communication.

---



*Dr hab. MARZENA ŚWIGOŃ, prof. UWM specjalizuje się w problematyce zarządzania i dzielenia się informacją i wiedzą. Habilitowała się w zakresie bibliologii i informatologii na Uniwersytecie Wrocławskim w 2013 r. Pracuje w Instytucie Historii i Stosunków Międzynarodowych Uniwersytetu Warmińsko-Mazurskiego w Olsztynie, kieruje Pracownią Badań nad Komunikacją i Zarządzaniem Informacją. Najważniejsze publikacje: M. Świgoń (2013). *Personal Knowledge and Information Management – Conception and Exemplification*. *Journal of Information Science*, 39(6); M. Świgoń (2011). *Library Anxiety Among Polish Students: Development and Validation of the Polish Library Anxiety Scale*. *Library & Information Science Research*, 33(2); M. Świgoń (2011). *Information Limits – Definition, Types and Typologies*. *Aslib Proceedings: New Information Perspectives*, 63(4).*

*Kontakt z autorką:*

*marzena.swigon@uwm.edu.pl*

*Pracownia Badań nad Komunikacją i Zarządzaniem Informacją*

*Instytut Historii i Stosunków Międzynarodowych*

*Wydział Humanistyczny*

*Uniwersytet Warmińsko-Mazurski w Olsztynie*

*ul. Kurta Obiżca 1*

*10–725 Olsztyn*

*DAVID NICHOLAS (PhD. DSc. MPhil.) jest kierownikiem niezależnego centrum badawczego CIBER Research Ltd. i profesorem na University of Tennessee. Autor kilkuset publikacji z obszaru nauki o informacji, komunikacji naukowej i zachowań informacyjnych. Do najnowszych należą m.in.: D. Nicholas, E. Herman, D. Clark (2016). *Scholarly Reputation Building — How Does ResearchGate Fare?* *International Journal of Knowledge Content Development & Technology*,6(2); D. Nicholas, D. Clark, E. Herman (2016). *ResearchGate: Reputation Uncovered*. *Learned Publishing*, 29(3); D. Nicholas, H. R. Jamali, A. Watkinson, E. Herman, C. Tenopir, R. Volentine, S. Allard, K. Levine (2015). *Do Younger Researchers Assess Trustworthiness Differently When Deciding What to Read and Cite and Where to Publish?* *International Journal of Knowledge Content Development & Technology*,5(2).*

*Kontakt z autorem:*

*CIBER Research Ltd.*

*1 Westwood Farmhouse*

*Greenham*

*Newbury*

*RG14 7RU*

---

# Co mogą zaoferować cyfrowej humanistyce biblioteki i ośrodki informacji?

Anna Kamińska

*Katedra Informatologii, Wydział Dziennikarstwa, Informacji i Bibliologii,  
Uniwersytet Warszawski*

---

## Abstrakt

**Cel/Teza:** Celem artykułu jest zaprezentowanie wyników badań dotyczących usług oferowanych przez biblioteki i ośrodki informacji dla cyfrowej humanistyki.

**Koncepcja/Metody badań:** Termin „usługa” został zdefiniowany jako pomoc oferowana cyfrowym humanistom przy realizacji ich badań oraz projektów. Usługi zidentyfikowano na podstawie wyselekcjonowanych abstraktów referatów wygłoszonych podczas międzynarodowej konferencji „Digital Humanities” w 2016 r. w Krakowie.

**Wyniki i wnioski:** Wśród zidentyfikowanych możliwych do zaproponowania lub już proponowanych usług znalazły się usługi z zakresu udostępniania cyfrowych dokumentów oraz metadanych dokumentów, tworzenia archiwów i repozytoriów cyfrowych oraz innych systemów informacyjnych, tworzenia ontologii i taksonomii oraz pomoc przy prowadzeniu kursów cyfrowej humanistyki dla studentów i naukowców. Zwrócono uwagę na problem interoperacyjności cyfrowych narzędzi, takich jak biblioteki cyfrowe czy bibliograficzne bazy danych, oferowanych przez pracowników informacji z aplikacjami spoza środowiska bibliotekarskiego, długoterminowej archiwizacji zasobów cyfrowych oraz na konieczność posiadania przez pracowników informacji wiedzy dziedzinowej z obszaru nauk humanistycznych.

**Oryginalność/Wartość poznawcza:** Identyfikacja i opis usług dla humanistyki cyfrowej przyniesie specjalistom informacji cenną wiedzę na temat nowych perspektyw dla ich zadań zawodowych oraz możliwości rozwoju ich miejsc pracy – bibliotek i ośrodków informacji. Uzyskane wyniki mogą być szczególnie ważne właśnie dla bibliotek, które w czasach zmieniającej się roli biblioteki poszukują nowych zadań i obszarów działania. Warto też zauważyć, że w Polsce humanistyka cyfrowa rozwija się dosyć intensywnie od kilku lat, co może stanowić dobry moment na włączenie się bibliotek i ośrodków informacji w rozwój tego nowego nurtu w naukach humanistycznych.

## Słowa kluczowe

Bibliotekarze. Biblioteki. Cyfrowa humanistyka. Ośrodki informacji. Pracownicy informacji. Usługi. Usługi informacyjne.

*Otrzymany: 3 listopada 2017. Zrecenzowany: 29 listopada 2017. Zaakceptowany: 14 grudnia 2017.*

---

## 1. Wprowadzenie

Cyfrowa humanistyka już od kilkunastu lat budzi zainteresowanie badaczy reprezentujących różne dyscypliny nauk humanistycznych, ale nie doczekała się jeszcze jednej powszechnie przyjętej definicji. Można jednak wyróżnić trzy główne jej cechy, często powtarzane w różnych opisach i proponowanych definicjach:

- (1) Badanie przez przedstawicieli nauk humanistycznych rzeczy czy zjawisk zaistniałych i funkcjonujących w środowisku cyfrowym.

- (2) Wykorzystywanie narzędzi cyfrowych do prowadzenia badań w naukach humanistycznych, używanie nowych technologii do udostępniania w środowisku cyfrowym dzieł kultury analogowej (Maryl, 2017).

Zarówno w Polsce, jak i na świecie, coraz liczniej powstają jednostki realizujące projekty humanistyki cyfrowej, wspierające prace cyfrowych humanistów i zapewniające naukowcom dostęp do potrzebnej infrastruktury badawczej. Wydaje się, że również biblioteki i ośrodki informacji mogą włączyć się w tworzenie takich jednostek oraz zaoferować cyfrowym humanistom usługi przydatne w realizacji ich projektów badawczych.

Celem niniejszego artykułu jest pokazanie, jakie usługi mogą zaoferować biblioteki oraz ośrodki informacji cyfrowym humanistom w ich pracy badawczej. Na tematykę usług postanowiono spojrzeć z szerszej perspektywy, opierając się na definicji tego terminu zawartej w Słowniku Języka Polskiego PWN online (<https://sjp.pwn.pl/>). Słownik ten określa usługę jako *pomoc okazaną komuś*. W przypadku pracowników bibliotek i ośrodków informacji, pomoc okazana naukowcom przy realizacji projektów cyfrowej humanistyki, może dotyczyć dzielenia się specjalistyczną wiedzą i umiejętnościami związanymi z pracą z informacją, dostarczeniem materiałów do badań, narzędzi i technologii tworzących e-infrastrukturę (Starczewski, 2015). Wśród badanych usług znalazły się oczywiście usługi informacyjne, które za Stanisławą Kurek-Kokocińską zostały tutaj zdefiniowane jako

specyficzna dziedzina działalności, której przedmiotem (towarem) jest informacja: wyselekcjonowana, zebrana, opracowana i przedstawiona w określony sposób, dostarczona do użytkownika (Kurek-Kokocińska, 2016, 165).

Usługi informacyjne można podzielić na trzy główne kategorie: związane z poszukiwaniem dokumentów i informacji, udostępnianiem dokumentów i informacji oraz szkoleniem użytkowników w zakresie pracy z informacją. Wydaje się oczywiste, że cyfrowi humaniści korzystają z usług informacyjnych. Cyfryzacja warsztatu pracy humanisty nie spowodowała, że przestał on być naukowcem, który poszukuje i korzysta z dokumentów oraz informacji do swoich badań, a także potrzebuje kompetencji informacyjnych do sprawnego poruszania się w środowisku informacji naukowej. Przedstawiona obserwacja pozwoliła ukierunkować analizę oferowanych usług informacyjnych pod kątem ich dostosowania do specyficznych potrzeb naukowych cyfrowych humanistów.

Identyfikacja i opis usług dla humanistyki cyfrowej oferowanych przez pracowników bibliotek i ośrodków informacji są ważne z perspektywy wykonywanych zadań zawodowych oraz możliwości rozwoju bibliotek i ośrodków informacji. Uzyskane wyniki przeprowadzonej analizy mogą być szczególnie istotne właśnie dla bibliotek, które w czasach zmieniającej się roli biblioteki poszukują nowych zadań i obszarów działania. Warto też zauważyć, że w Polsce humanistyka cyfrowa rozwija się intensywnie dopiero od kilku lat, a więc może to być dobry moment na włączenie się bibliotek i ośrodków informacji w rozwój tego nowego nurtu w naukach humanistycznych.

## 2. Opis koncepcji i metody przeprowadzonego badania

Materiał badawczy, który posłużył do identyfikacji usług oferowanych przez bibliotekarzy i specjalistów informacji dla cyfrowej humanistyki, stanowiły abstrakty referatów wygłoszonych podczas corocznej międzynarodowej konferencji naukowej „Digital Humanities”, która

odbyła się latem 2016 r. w Krakowie (Eder & Rybicki, 2016). Konferencja organizowana jest przez Alliance of Digital Humanities Organizations (ADHO) i należy do najważniejszych naukowych wydarzeń dla przedstawicieli cyfrowej humanistyki. Zaprezentowane w referatach wyniki badań oraz opisy projektów i inicjatyw nakreślają pewien obraz współczesnej cyfrowej humanistyki, obszarów jej najnowszych zainteresowań oraz kierunków rozwoju.

Księga abstraktów konferencji „Digital Humanities 2016”<sup>1</sup> zawiera 23 abstrakty sesji tematycznych składających się z kilku referatów (ang. *panels*), 119 abstraktów wystąpień w formie długich referatów (ang. *long papers*), 135 abstraktów wystąpień w formie krótkich referatów (ang. *short papers*) i 131 abstraktów prezentowanych w formie plakatów (ang. *posters*). Ostatecznie do badania wybrano abstrakty wystąpień w formie długich referatów napisanych w języku angielskim w liczbie 110<sup>2</sup>. Prezentacje w formie długich referatów służyły do przedstawienia znaczących, ukończonych i niepublikowanych jeszcze badań, opisu nowych, znaczących metod czy narzędzi badawczych lub zasobów cyfrowych, jak również do zaprezentowania omówień problemów teoretycznych czy hipotez badawczych. Ich abstrakty odznaczały się znaczną obszernością (mogły zawierać do 1500 słów) oraz dużym stopniem szczegółowości, co pozwoliło uznać tę część abstraktów za wartościowy materiał badawczy. Abstrakty krótkich referatów z założenia opisywać miały projekty narzędzi i oprogramowania będące we wczesnych fazach rozwoju oraz miały prezentować wnioski z nieukończonych jeszcze badań i eksperymentów. Z tego względu ich analiza nie przyniosłaby satysfakcjonujących wyników. Abstrakty sesji tematycznych i posterów nie miały poziomu szczegółowości wystarczającego do przeprowadzenia analizy, przez co nie zostały włączone do badań (DH 2016, 2016).

Na decyzję o ograniczeniu liczby analizowanych abstraktów wpłynął również fakt, iż cały zgromadzony materiał nie stanowił próby reprezentatywnej dla wszystkich projektów humanistyki cyfrowej prowadzonych w 2016 r., co wykluczało podejmowanie próby generalizacji uzyskanych wyników. Analiza wyselekcjonowanych abstraktów pozwoliła natomiast na zidentyfikowanie specyficznych usług dla humanistyki cyfrowej oferowanych przez biblioteki i ośrodki informacji, a także na uchwycenie pewnych interesujących tendencji.

Zgromadzonym abstraktom przyjrano się z perspektywy pól badawczych bibliologii i informatologii. Określono powiązania i relacje pomiędzy obszarami zainteresowań cyfrowej humanistyki zaprezentowanych w referatach a obszarami zainteresowań bibliologii i informatologii, następnie sprawdzono, które z odkrytych powiązań mogą mieć charakter świadczonych usług. Współczesny obraz pól badawczych bibliologii i informatologii stworzono na podstawie kilku uzupełniających się źródeł. Najważniejszym z nich było trzecie, ostatnie wydanie *Encyclopedia of Library and Information Sciences* (Bates & Maack, eds., 2010), która jest najpełniejszym i nadal najbardziej aktualnym źródłem wiedzy o tych dwóch dyscyplinach. Pomocnym uzupełnieniem okazały się również publikacje *Bibliotekarstwo* pod red. Anny Tokarskiej (Tokarska, 2013) oraz *Nauka o informacji* pod red. Wiesława Babika (Babik, 2016), które można uznać za aktualne polskie podręczniki wykorzystywane w kształceniu bibliotekarzy oraz specjalistów informacji.

<sup>1</sup> Księga abstraktów z konferencji „Digital Humanities 2016” jest dostępna na stronie internetowej konferencji pod adresem <http://dh2016.adho.org/abstracts/>. Sprawozdanie z konferencji ukazało się w ZIN 2016, 54(1), 163–166.

<sup>2</sup> Pozostałe dziewięć abstraktów przygotowano w innym języku niż angielski.

Analizując zebrane abstrakty, zwrócono również uwagę na takie kwestie, jak dyscyplina naukowa, z której wywodzi się temat referatu oraz na cel wystąpienia. Te dodatkowe informacje pomogły w identyfikacji usług, a także umożliwiły dokonanie głębszej analizy badanego zagadnienia.

### 3. Wyniki przeprowadzonego badania

#### 3.1. *Charakterystyka zebranego materiału*

Analizując zebrany materiał pod kątem podejmowanych tematów zauważono, że największa część referatów była poświęcona zagadnieniom lingwistycznym (39 referatów) oraz interdyscyplinarnym – wystąpienia przyporządkowane do tej kategorii najczęściej stanowiły opis programu komputerowego czy aplikacji przeznaczonej dla humanistów uprawiających różne dyscypliny naukowe lub dotyczyły problematyki z pogranicza co najmniej trzech dziedzin nauki (18 referatów). Nieco mniejsze grupy tworzyły referaty podejmujące tematy historyczne (14 referatów) oraz metanaukową refleksję nad cyfrową humanistyką – jej istotą, instytucjami czy studiami prowadzonymi w jej zakresie (10 referatów). Prelegenci prezentowali prace również z zakresu muzykologii i bibliologii (po osiem referatów), dydaktyki i informatologii (po siedem referatów), socjologii (cztery referaty), etnologii i historii sztuki (po trzy referaty). Nieliczne wystąpienia dotyczyły zagadnień medioznawczych i o sztuce (po dwa referaty) oraz architektonicznych, geograficznych, kulturoznawczych i archeologicznych (po jednym referacie). Warto wspomnieć, że wśród tematów bibliologicznych i informatologicznych znalazły się takie jak: edycje cyfrowych publikacji – opracowań naukowych i tekstów źródłowych, czytelnictwo, produkcja wydawnicza, rekonstrukcja zniszczonych książek, kompetencje cyfrowe, komunikacja naukowa, zachowania informacyjne, a także organizacja wiedzy.

Za cel referatu prelegenci najczęściej obierali rozwiązanie konkretnego problemu badawczego o charakterze teoretycznym lub praktycznym (49 referatów) oraz przedstawienie konkretnej metody, rozwiązania technologicznego, programu czy narzędzia wspomagającego pracę cyfrowych humanistów (47 referatów). Prezentowane aplikacje i rozwiązania były przeznaczone zarówno dla wąskiej grupy specjalistów (np. muzykologom czy językoznawcom), jak i dla ogółu badaczy zajmujących się humanistyką cyfrową. Część referatów stanowiła opis konkretnej bazy, repozytorium czy archiwum cyfrowego gromadzącego informacje lub zdigitalizowane szeroko rozumiane dziedzictwo kulturowe (12 referatów). Kilka wystąpień miało na celu przedstawienie zajęć lub kursów prowadzonych dla studentów z zakresu cyfrowej humanistyki (pięć referatów). Jeden referat stanowił opis instytucji wspomagającej rozwój cyfrowej humanistyki w konkretnym kraju.

#### 3.2. *Usługi adresowane do cyfrowej humanistyki przez biblioteki i ośrodki informacji – analiza materiału badawczego*

Wśród wspólnych obszarów zainteresowań cyfrowej humanistyki oraz bibliologii i informatologii, w których bibliotekarze i specjaliści informacji mogą wspomóc cyfrowych humanistów, znalazły się:

- (1) biblioteki i archiwa cyfrowe, bazy pełnotekstowe, digitalizacja;
- (2) systemy informacyjne;
- (3) badania użytkowników informacji;
- (4) architektura informacji;
- (5) organizacja informacji i wiedzy;
- (6) edukacja informacyjna i medialna.

W dalszej części artykułu wskazano, jakie konkretnie usługi mogą zaoferować lub już oferują pracownicy informacji w wyżej wymienionych obszarach zainteresowań.

Materiał badawczy w naukach humanistycznych stanowią przede wszystkim różnego rodzaju dokumenty takie jak książki, czasopisma, druki urzędowe, zdjęcia, filmy, obrazy czy nagrania audio. Poddanie takich dokumentów analizie za pomocą specjalistycznych programów komputerowych wymaga, aby miały one postać cyfrową – pierwotną (tzw. dokumenty *born digital*) lub wtórną (poprzez digitalizację). W udostępnianiu zdigitalizowanych dokumentów ważną rolę odgrywają biblioteki i archiwa dzięki digitalizacji posiadanych zbiorów oraz tworzeniu bibliotek i archiwów cyfrowych. W analizowanych abstraktach referatów konferencyjnych znalazły się 74 abstrakty związane z tematem wykorzystania cyfrowych dokumentów i cyfrowych kolekcji. Choć w przypadku części z nich jednoznaczna identyfikacja pierwotnej formy dokumentów (analogowej lub cyfrowej) okazała się niemożliwa, dopuszczalne jest jednak założenie, że powstały one w wyniku digitalizacji<sup>3</sup>. Cyfrowe dokumenty stanowiły materiał badawczy, z myślą o ich analizie tworzono programy i aplikacje, stanowiły uzupełnienie elektronicznych publikacji za pośrednictwem narzędzi Linked Data (Bizer et al., 2011), a także wzbogacały multimedialne publikacje. Temat digitalizacji i tworzenia kolekcji cyfrowych znalazł się również w programie jednego z kursów akademickich dla studentów, przedstawionym w referacie *Read, Play, Build: Teaching Sherlock Holmes through Digital Humanities* autorstwa J. E. Swafford.

Ciekawie prezentuje się kwestia podejścia autorów referatów do korzystania z cyfrowych dokumentów i kolekcji cyfrowych. W 22 abstraktach referatów zaznaczono, że do przeprowadzenia badań czy testowania specjalistycznych narzędzi komputerowych używano cyfrowych dokumentów, ale w tekście abstraktów referatów nie podano w wyraźny sposób informacji o cyfrowej postaci tych dokumentów oraz nie wskazano drogi ich pozyskania. Przedstawione zjawisko występowało szczególnie często w abstraktach referatów poświęconych badaniom z zakresu lingwistyki komputerowej (np. w referatach *Go Set A Watchman while we Kill the Mockingbird in Cold Blood, with Cats and Other People* M. Edera i J. Rybickiego czy *The Dialogic Turn and the Performance of Gender: the English Canon 1782–2011* G. Muzny, M. Algee-Hewitt i D. Jurafsky). W 14 abstraktach wspomniano o digitalizacji dokumentów na potrzeby badań czy projektu. Naukowcy korzystali także z istniejących już kolekcji cyfrowych, wśród których znalazły się British Library's Nineteenth Century Newspapers Collection, Project Gutenberg (którego wykorzystanie wskazano w trzech abstraktach), Europeana, Digital Public Library of America czy The European Library<sup>4</sup>.

<sup>3</sup> Spośród wszystkich badanych abstraktów osiem dotyczyło wykorzystania (jako materiał badawczy) dokumentów cyfrowych, które jednoznacznie uznano za dokumenty typu *born digital*. Wśród nich znalazły się tweety, czy też rekordy z internetowych katalogów bibliotecznych.

<sup>4</sup> Choć Europeana, Digital Public Library of America czy The European Library *de facto* są agregatorami danych pochodzących z różnych bibliotek i archiwów cyfrowych, to postanowiono wymienić je w tym miejscu z powodu ich powszechnej identyfikacji jako kolekcje cyfrowe.



Udostępnianie dokumentów stanowi jedną z podstawowych usług informacyjnych oferowanych przez biblioteki. Z punktu widzenia potrzeb cyfrowej humanistyki należy zwrócić uwagę na trzy kwestie związane z realizacją tej usługi. Pierwsza z nich dotyczy skali, na jaką prowadzone są działania digitalizacyjne. Jak wspomniano, tylko dokumenty w cyfrowej postaci mogą zostać poddane analizie komputerowej. Oprócz tego analiza dużej liczby dokumentów, możliwa dzięki programom komputerowym, pozwala na prowadzenie badań na niespotykaną do tej pory skalę (tzw. zjawisko *big data*, zob. Idzik, 2013). Masowa digitalizacja zasobów bibliotek czy archiwów napotyka jednak bariery finansowe oraz prawne związane z przestrzeganiem autorskich praw majątkowych<sup>5</sup>. Wraz z digitalizacją dziedzictwa kulturowego biblioteki i archiwum powinny zatroszczyć się o ich długoterminową archiwizację. Ważne jest nie tylko zadbanie o przetrwanie zasobów cyfrowych dla kolejnych pokoleń badaczy, ale również umożliwienie ich odczytu w ciągle zmieniającym się środowisku cyfrowym (Gmiterek, 2013).

Kolejna kwestia dotyczy problemu przygotowania cyfrowych dokumentów do analizy komputerowej. Sama digitalizacja to za mało – bez właściwego przygotowania skanów dokumentów do automatycznego przetwarzania zawartych w nich informacji, w tym także automatycznego wnioskowania, sama digitalizacja dla cyfrowych humanistów traci na znaczeniu. Normą stosowaną w bibliotekach i archiwach cyfrowych powinno stać się stosowanie techniki OCR (Optical Character Recognition) do rozpoznawania tekstu z plików graficznych (Bień, 2011)<sup>6</sup>. Kolejny krok powinien polegać na kodowaniu zdigitalizowanych materiałów, np. w standardzie TEI (Text Encoding Initiative) (Vanhouthe, 2004). Przy digitalizacji dokumentów nie można również zapomnieć o zapewnieniu wysokiej jakości wykonanych skanów, szczególnie w przypadku dokumentów graficznych. Warto też dodać, że wydawcy komercyjnych i niekomercyjnych baz pełnotekstowych, w szczególności wydawcy tekstów literackich w postaci ebooków, powinni zwrócić uwagę na opisane powyżej kwestie.

Kolejna płaszczyzna, na której zarysowuje się możliwość współpracy pomiędzy cyfrowymi humanistami a pracownikami informacji, dotyczy systemów informacyjnych. Cyfrowi humaniści są twórcami różnego rodzaju systemów informacyjnych. Jak wcześniej zauważono, 12 abstraktów referatów miało na celu prezentację konkretnej stworzonej w ramach projektu bazy, którą wraz z narzędziami jej przeszukiwania, komunikacji z użytkownikiem i prezentacji wyników wyszukiwania można uznać za system informacyjny. Wśród przedstawionych systemów informacyjnych znalazły się: cyfrowy korpus starożytnych inskrypcji skalnych z terenu dzisiejszej Sycylii, baza z historycznymi informacjami ekonomicznymi z XIX w., baza z materiałami prezentującymi historię Florydy, baza zbierająca informacje o projektach, narzędziach i danych tworzonych w ramach cyfrowej humanistyki, repozytorium tekstów oraz danych z zakresu lingwistyki, cyfrowy korpus chronologiczny polskich tekstów prasowych, biblioteka cyfrowa Early English Books Online, repozytorium dokumentów dotyczących Afryki, cyfrowe archiwum kolekcji muzyki ludowej z Północnej Karoliny, archiwum z dokumentami (nagraniami, filmami, interaktywnymi mapami) na temat

<sup>5</sup> Ciekawym przykładem pokonania bariery prawnej przy digitalizacji zbiorów bibliotecznych jest projekt ACADEMICA realizowany przez polską Bibliotekę Narodową (Academica, b.d.).

<sup>6</sup> Technika OCR jest już stosowana w bibliotekach i archiwach cyfrowych, jednak w niezadowalającym stopniu.

niemieckich obozów pracy i obozów zagłady z lat 1939–1945, czy baza z informacjami na temat gramatyki języków niderlandzkiego i fryzyjskiego. W badanym zbiorze abstraktów znalazły się również cztery abstrakty opisujące portale agregujące dane z innych serwisów czy stron internetowych, które także są pewnego typu systemami informacyjnymi, np. agregator informacji o wydaniach (papierowych i cyfrowych) korespondencji z różnego typu baz czy agregator informacji o dokumentach muzycznych z kolekcji cyfrowych różnych instytucji. Warto też zauważyć, że w siedmiu abstraktach pojawiła się informacja o stworzeniu baz lub repozytoriów na potrzeby realizacji konkretnych badań (prezentacja tych baz nie była głównym celem referatu).

Analiza zakresu tematycznego oraz rodzaju dokumentów gromadzonych i udostępnianych w podanych systemach informacyjnych pozwala na uchwycenie dwóch ciekawych spostrzeżeń. Bibliotekarze oraz specjaliści informacji tworzą bazy pełnotekstowe lub agregatory informacji odznaczające się interdyscyplinarnym lub też ogólnym charakterem gromadzonych treści, w tym treści wykorzystywanych w badaniach cyfrowej humanistyki. Patrząc z tej perspektywy, również pracowników informacji można byłoby zatem uznać za cyfrowych humanistów. Brak specjalistycznej wiedzy dziedzinowej może jednak stanowić problem w tworzeniu wyspecjalizowanych systemów informacyjnych. Pracownicy informacji posiadają za to inną wiedzę i umiejętności, atrakcyjne dla zespołów konstruujących takie systemy – wiedzę i umiejętności z zakresu budowania architektury informacji systemów informacyjnych, z zakresu tworzenia metadanych dotyczących formalnych i rzeczowych cech dokumentów, które umożliwiają identyfikację dokumentów i przetwarzanie informacji o nich w środowisku cyfrowym, a także wiedzę i umiejętności z zakresu dostosowywania systemów informacyjnych do oczekiwań i umiejętności użytkowników (tzw. podejście *user experience* – zob. Mościchowska & Rogoś-Turek, 2015) na podstawie analizy ich potrzeb oraz zachowań informacyjnych. Dzięki pogłębionej współpracy bibliotekarzy i specjalistów informacji oraz cyfrowych humanistów bazy i serwisy tworzone dla naukowców mogą lepiej zaspokajać oczekiwania cyfrowych humanistów oraz zapewnić im wysoki komfort pracy.

Przedstawiciele nowego nurtu badań humanistycznych, jakim jest cyfrowa humanistyka, są nie tylko twórcami, ale też użytkownikami systemów informacyjnych, z których korzystają w codziennej pracy naukowej. Wśród użytkowanych systemów informacyjnych znajdują się systemy gromadzące informacje o dokumentach, takie jak katalogi biblioteczne, katalogi centralne czy bazy bibliograficzne, kartoteki haseł wzorcowych, a także gromadzące same dokumenty, np. wspomniane już wcześniej biblioteki i archiwa cyfrowe czy bazy pełnotekstowe. W kontekście badania usług oferowanych przez biblioteki i ośrodki informacji warto zwrócić uwagę na dwa aspekty użytkowania tych systemów informacyjnych. Pierwszy z nich dotyczy wykorzystania jako materiał badawczy metadanych dokumentów opisywanych i gromadzonych w katalogach czy bazach. W zbiorze analizowanych abstraktów referatów znalazły się opisy dwóch badań przeprowadzonych na metadanych z baz bibliograficznych oraz baz z kolekcjami cyfrowymi. W pierwszym badaniu, materiał badawczy stanowiły opisy rzeczowe obrazów z baz RKDImages, The Montias Database of 17th Century Dutch Art Inventories, Rijksmuseum (badanie dotyczyło tematyki podejmowanej w malarstwie niderlandzkim i flamandzkim), zaś w drugim, opisy bibliograficzne z Fińskiej Bibliografii Narodowej Fennica (badanie dotyczyło produkcji wydawniczej książki fińskiej w latach 1640–1828).

Drugi aspekt korzystania z katalogów, baz bibliograficznych czy pełnotekstowych jest związany z tworzeniem multimedialnych publikacji oraz agregatorów danych. Dzięki narzędziom Linked Data metadane z katalogów bibliotecznych czy bibliotek cyfrowych wzbogacają multimedialne publikacje i agregatory danych. Wśród analizowanych abstraktów referatów znalazły się trzy podejmujące tę tematykę. Warto przyjrzeć się bliżej dwóm z nich. Pierwszy opisywał platformę integrującą informacje o wydaniach (papierowych i cyfrowych) korespondencji z różnego typu baz (correspSearch). Informacje o nadawcach, odbiorcach oraz adresie nadawcy i odbiorcy pozyskano z kartotek haseł wzorcowych Library of Congress, Deutsche Nationalbibliothek, Bibliothèque nationale de France, japońskiej National Diet Library oraz Virtual International Authority File (VIAF). W drugim abstrakcie przedstawiono platformę integrującą informacje biograficzne o ludziach tworzących dokumenty historyczne lub w nich opisanych (SNAC). Dane zgromadzone w SNAC pochodzą z katalogu centralnego OCLC WorldCat udostępniającego informacje o zbiorach wybranych bibliotek z całego świata, kartotek National Archives and Records Administration (NARA), British Library, Smithsonian Institution Archives, New York State Archives oraz Archives Nationales (France).

Systemy informacyjne tworzone przez bibliotekarzy oraz specjalistów informacji, takie jak katalogi biblioteczne czy bazy bibliograficzne mogą znaleźć szerokie wykorzystanie w projektach i badaniach z obszaru cyfrowej humanistyki, jeśli będą w nich stosowane standardy sieciowe, a także rozwiązania technologiczne umożliwiające automatyczne przetwarzanie danych w nich zawartych przez zewnętrzne aplikacje. Dobrym przykładem takiego systemu jest VIAF, który jest w pełni przystosowany do wykorzystania przez aplikacje sieciowe nie tylko z obszaru działań bibliotekarskich.

Usługi związane z organizacją informacji i wiedzy stanowią kolejne pole działalności pracowników informacji, która może wspierać rozwój cyfrowej humanistyki. Przedstawiciele tego nowego nurtu badań humanistycznych do satysfakcjonującego przeprowadzenia swoich badań korzystają z sieciowych systemów organizacji wiedzy (SSOW), takich jak taksonomie czy ontologie informatyczne. Wśród poddanych analizie abstraktów referatów zidentyfikowano 12 takich, które poruszały zagadnienie ontologii informatycznych (trzy opisywały specjalnie stworzone ontologie, w dziewięciu wspomniano o wykorzystaniu ontologii do przeprowadzenia konkretnych badań), pięć dotyczących tematu taksonomii, zaś w dwóch abstraktach wskazano na użycie do badań tezaursów. Tworzone oraz użytkowane SSOW opisywały niewielki obszar wiedzy, np. metody badawcze cyfrowej humanistyki, terminologię z zakresu lingwistyki, interpretację tematów muzycznych czy relacje pracownik-pracodawca (stosunek pracy).

W tym obszarze wspólnych zainteresowań pracownicy informacji służą pomocą cyfrowym humanistom przede wszystkim teoretyczną wiedzą na temat tworzenia taksonomii czy ontologii (brak specjalistycznej wiedzy dziedzinowej znacznie utrudnia pomoc przy ich przygotowaniu od strony praktycznej), jak również dają gotowe narzędzia w postaci konkretnych SSOW. Dzięki stosowaniu formatów danych zgodnych ze standardami sieciowymi (np. standardu Simple Knowledge Organization System, SKOS) coraz więcej SSOW może być wykorzystanych w cyfrowych projektach naukowych<sup>7</sup> (Sosińska-Kalata & Roszkowski, 2016).

<sup>7</sup> Lista SSOW zapisanych w standardzie SKOS dostępna jest na stronie internetowej World Wide Web Consortium <https://www.w3.org/2001/sw/wiki/SKOS/Datasets> (Sosińska-Kalata & Roszkowski, 2016).

Ostatni rodzaj usług, jakie oferują bibliotekarze i specjaliści informacji ma związek z działalnością dydaktyczną. Wiedza i umiejętności pracowników informacji okazują się cenne przy realizacji cyfrowych badań i projektów z zakresu humanistyki, co potwierdzają analizy opisane we wcześniejszej części niniejszego artykułu. Przykłady włączania treści informatologicznych do programów nauczania cyfrowej humanistyki na poziomie akademickim można znaleźć w abstraktach trzech referatów (na pięć opisujących akademickie kursy cyfrowej humanistyki) ze zgrupowanego materiału badawczego. Znalazły się w nich tematy takie, jak tworzenie archiwów cyfrowych, ontologie i zarządzanie wiedzą czy wizualizacja informacji. W innym abstrakcie, opisującym kurs cyfrowej humanistyki, wspomniano o współpracy z biblioteką akademicką przy prowadzeniu kursu (niestety, nie opisano zadań realizowanych przez bibliotekę w ramach prowadzenia kursu). Bliższe poznanie tematów z obszarów bibliologii i informatologii pojawiających się w programach nauczania cyfrowej humanistyki oraz zakresu współpracy z bibliotekami i ośrodkami informacji przy realizacji tych programów wymaga przeprowadzenia bardziej szczegółowych badań.

#### 4. Wnioski

Analiza abstraktów referatów z konferencji „Digital Humanities 2016” pokazała, że biblioteki oraz ośrodki informacji mogą zaoferować i już oferują cyfrowym humanistom szereg usług pomocnych w prowadzeniu badań i realizacji cyfrowych projektów. Wśród nich znalazły się usługi z zakresu dostarczania materiału badawczego oraz materiału uzupełniającego do publikacji multimedialnych w postaci kolekcji cyfrowych dokumentów lub metadanych dokumentów z katalogów, kartotek haseł wzorcowych czy baz bibliograficznych lub pełnotekstowych. Kolejne usługi dotyczą dostarczania narzędzi badawczych w postaci taksonomii, ontologii i innych sieciowych systemów organizacji wiedzy. Pracownicy informacji dzielą się też wiedzą i doświadczeniem przy projektowaniu, tworzeniu, organizowaniu oraz zarządzaniu cyfrowymi kolekcjami i innymi systemami informacyjnymi, jak również przy tworzeniu taksonomii i ontologii. Proponowane usługi często mają charakter usług outsourcingowych, ale przyjmują też inną formę, np. aktywnego uczestnictwa w realizacji określonych projektów i przedsięwzięć. A zatem trafne jest stwierdzenie, że w zespołach badawczych i projektowych cyfrowych humanistów jest miejsce dla bibliotekarzy i specjalistów informacji.

Satysfakcjonująca realizacja usług dla cyfrowej humanistyki oferowanych przez pracowników informacji napotyka kilka ważnych wyzwań. Pierwsze związane jest z posiadaniem specjalistycznej wiedzy informatycznej. Narzędzia oferowane do tej pory przez biblioteki i ośrodki informacji, takie jak elektroniczne katalogi, bazy bibliograficzne, biblioteki i archiwa cyfrowe, systemy organizacji wiedzy powinny zostać wyposażone w innowacje pozwalające na przetwarzanie ich zawartości przez zewnętrzne aplikacje i programy, a także szerokie zastosowanie w środowisku sieciowym. Wiedza i umiejętności informatyczne potrzebne są również do prowadzenia długoterminowej archiwizacji zasobów cyfrowych. Warto zauważyć, że wykorzystanie rozwiązań informatycznych w bibliotekarstwie wkracza na wyższy poziom zaawansowania. Od pracowników informacji trudno wymagać wysoko wyspecjalizowanej wiedzy informatycznej, ich kompetencje w tym zakresie powinny jednak być na tyle zaawansowane, aby mogli wspólnie z informatykami tworzyć innowacyjne narzędzia dla cyfrowej humanistyki.

Nie tylko wiedza z dziedziny informatyki jest potrzebna do rozwijania użytecznej współpracy specjalistów informacji z cyfrowymi humanistami. Również specjalistyczna wiedza z obszaru nauk humanistycznych może okazać się konieczna przy realizacji konkretnych zadań dla cyfrowej humanistyki, np. przy tworzeniu ontologii czy taksonomii. Na tym polu wykazać się mogą bibliotekarze dziedzinowi, którzy taką wiedzę właśnie posiadają. Współpraca z cyfrowymi humanistami mogłaby dać tej grupie bibliotekarzy nowe wyzwania w pracy zawodowej oraz nowe możliwości rozwoju.

Warto zastanowić się także, na ile cyfrowi humaniści są świadomi usług, które mają im do zaoferowania biblioteki oraz ośrodki informacji. Ważne jest, aby bibliotekarze i pracownicy informacji wychodzili naprzeciw potrzebom przedstawicieli tego nowego nurtu badań humanistycznych. Zaangażowanie w prace jednostek tworzących i wspierających rozwój humanistyki cyfrowej – uczelnianych laboratoriów cyfrowej humanistyki czy międzynarodowych stowarzyszeń oraz organizacji (np. ADHO czy DARIAH<sup>8</sup>) może przynieść wiele korzyści dla środowiska pracowników informacji.

Podsumowując można stwierdzić, że możliwości współpracy pomiędzy cyfrowymi humanistami a bibliotekarzami i specjalistami informacji rysują się wyraźnie. Zasadne wydaje się jednak postawienie pytania o realny kształt tej współpracy, co stanowi nowy, warty zgłębienia problem badawczy.

## Bibliografia

- Academica (b.d.). O projekcie [online]. *Academica wypożyczalnia* [1.09.2017], [https://academica.edu.pl/about/new\\_quality](https://academica.edu.pl/about/new_quality)
- Babik, W., red. (2016). *Nauka o informacji*. Warszawa: Wydaw. SBP.
- Bates, M. J., & Maack, M. N., eds. (2010). *Encyclopedia of Library and Information Sciences*. T. 1–7 (3rd Edition.). Boca Raton: CRC Press Taylor & Francis Group.
- Bień, J. S. (2011). *Skanowane teksty jako korpusy* [online]. Biblioteka Cyfrowa KLF UW [1.09.2017], <http://bc.klf.uw.edu.pl/192/>
- Bizer, C., Heath, T., Berners-Lee, T. (2011). Linked Data – The Story So Far. In: A. Sheth (ed.), *Semantic Services, Interoperability and Web Applications: Emerging Concepts* (205–228). Hershey: Information Science Reference.
- Dariah-eu (n.d.). *Dariah-eu* [online] [1.09.2017], <http://www.dariah.eu/>
- DH 2016 (2016). Zaproszenie do udziału w konferencji [online]. *Digital Humanities 2016* [7.08.2017], <http://dh2016.adho.org/cfp-polish/>
- Eder, M., Rybicki, J., eds. (2016). *Digital Humanities 2016 – Conference Abstracts*. Kraków: European Association for Digital Humanities et al.
- Gmiterek, G. (2013). Długoterminowa archiwizacja zasobów cyfrowych. *Res Historica*, 35, 213–235.
- Idzik, P. (2013). Analiza Big Data. Badania niereaktywne w erze Internetu 2.0. W: A. Radomski, R. Bomba (red.), *Zwrot cyfrowy w humanistyce* (153–167). Lublin: E-naukowiec.
- Kurek-Kokocińska, S. (2016). Usługi informacyjne. W: W. Babik (red.), *Nauka o informacji* (157–188). Warszawa: Wydaw. SBP.
- Maryl, M. (2017). Kim są polscy humaniści cyfrowi? *Teksty Drugie*, 1, 286–300.
- Mościchowska, I., Rogoś-Turek, B. (2015). *Badania jako podstawa projektowania user experience*. Warszawa: Wydaw. Naukowe PWN.
- Sosińska-Kalata, B., Roszkowski, M. (2016). Organizacja informacji i wiedzy. W: W. Babik (red.), *Nauka o informacji* (305–357). Warszawa: Wydaw. SBP.

<sup>8</sup> Digital Research Infrastructure for the Arts and Humanities (*Daria-eu*, n.d.).



- Starzewski, M. (2015). e-Infrastruktura: preludium cyfrowej humanistyki. *Czas Kultury*, 185(2), 42–47.
- Tokarska, A., red. (2013). *Bibliotekarstwo*. Warszawa: Wydaw. SBP.
- Vanhoutte, E. (2004). An Introduction to the TEI and the TEI Consortium. *Literary and Linguistic Computing*, 19(1), 9–16.
- 

## What Kind of Services Could Libraries and Information Centers Offer to Digital Humanities?

### Abstract

**Purpose/Thesis:** The purpose of this paper is to present services libraries and information centers could offer to digital humanities in the field of research and projects.

**Approach/Methods:** The abstracts selected from the book of abstracts of the International Conference “Digital Humanities 2016” were used as a test material to identify the services for digital humanities offered by libraries and information centers.

**Results and conclusions:** The analysis of the test material showed that libraries and information centers were capable of providing digital humanities with access to multiple digital documents from digital libraries and archives as well as multiple metadata of printed and digital documents. They could frequently help digital humanists with projects such as the construction of digital libraries, archives and repositories, the design of ontologies and taxonomies and the delivery of courses on digital humanities. The interoperability of the tools offered by libraries and information centers (for instance, digital libraries or bibliographic databases) with the non-library tools, the long-term protection of digital resources and lack of knowledge in the field of humanities are recognized as new problems faced by the librarians and information professionals.

**Originality/Value:** The results of the research showed some new tasks and challenges faced by the librarians and information professionals in their work as well as new directions of development for libraries and information centers. Digital humanities has been thriving in Poland recently and this may be a suitable moment for Polish libraries and information centers to become a promising partner of Polish digital humanists.

### Keywords

Digital humanities. Information centers. Information professionals. Information services. Librarians. Libraries.

---

ANNA KAMIŃSKA jest doktorantką w Katedrze Informatologii na Wydziale Dziennikarstwa, Informacji i Bibliologii Uniwersytetu Warszawskiego. Rozprawę doktorską pisze na temat cyfrowej humanistyki – jej koncepcji, istoty, kierunków rozwoju oraz powiązań z bibliologią i informatologią. Specjalności naukowe: cyfrowa humanistyka, komunikacja naukowa oraz zachowania informacyjne.

Kontakt z autorką:

[a.kaminska8@uw.edu.pl](mailto:a.kaminska8@uw.edu.pl)

Katedra Informatologii

Wydział Dziennikarstwa, Informacji i Bibliologii

Uniwersytet Warszawski

Nowy Świat 69

00–927 Warszawa



## Wśród zagranicznych książek

Jacek Wojciechowski

Przeglądając sygnały publikacji na temat informacji naukowej i praktyki informacyjnej, odnoszę wrażenie postępującej monotonii przedmiotowej i poniekąd autoteliczne skupienie się na wewnętrznych sprawach tej dyscypliny – a wyraźnie mniej na tym, co w rzeczywistości ewidentnie zasługiwałoby na nowe refleksje. Myślę, że są takie obszary i spróbuję tu je częściowo odnotować.

Coś musi zresztą być na rzeczy, skoro pojawiają się głosy o schyłku tej dyscypliny, a nawet wręcz o zapaści<sup>1</sup> – jakkolwiek być może nieco przesadzone. Nie mam dostatecznych kompetencji, żeby sformułować wiążącą opinię, jednak niepokój jest.

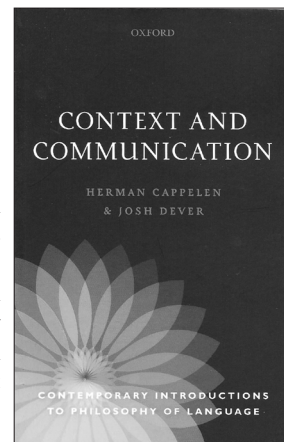
Ze względu na taką niepewność spróbowałem tym razem odnieść się do szerszego kontekstu teorii oraz praktyki informacji naukowej, licząc na przełamanie ograniczonego rejestru tematów. Być może udało mi się to – przynajmniej połowicznie. Ale swoją drogą, podaż opracowań nadmiernie mozaikowa nie jest.

### Informacja, konteksty, znaczenia [\*\*\*\*]

Herman Cappelen, Josh Dever (2016). *Context and Communication*. Oxford: Oxford University Press, pp.199. ISBN 978-0-19-873306-5

Pośród grzechów głównych nauki o informacji – a trochę ich jest – umieszczam niedostatek **rzeczywistych** rozpoznań procesów **odbioru** przekazywanych powiadomień. Jeżeli w ogóle pojawia się dyskurs o jakiegokolwiek recepcji, to przeważnie w prostych kategoriach partycypacji albo absencji w komunikacji informacyjnej, z ewentualnym przejściem kategoryzacyjnych objaśnień socjologicznych. Towarzyszy zaś temu pozaświadome założenie, że odbiór informacji (tymczasem w istocie: sygnałów – a w niepoprawnym utożsamieniu: treści) jest prostą odwrotnością procesu nadania. Nic bardziej mylnego! A to dlatego, że perspektywa odbiorcy bywa w takim ujęciu ignorowana. Nawet znakomite opracowanie Nicolasa Tomaiuolo<sup>2</sup>, chociaż skupia się na użytkownikach informacji, to jednak w rzeczywistości opisuje ich jako nadawców, albo re-nadawców informacji, a nie jako jej **spożywców**.

Z mizernego rozpoznania biorą się zasadnicze nieporozumienia. Ignorujące ewidentny fakt, że każdy odbiór wszelkich komunikatów polega na indywidualnej **interpretacji** treści. I trzeba rozemnać jej prawidłowości, jednak w roztrząsaniach informatologicznych na ogół niczego takiego nie ma. Stąd bierze się paninformacjonizm, uznający każdą transmisję komunikatu za informacyjną, jak też



<sup>1</sup> D. Bowden, L. Robinson (2012). *Introduction to Information Science*. London: Facet Publishing, 335; M. Gorman (2015). *Our Enduring Values Revisited*. Chicago: ALA Editions, 86.

<sup>2</sup> N. G. Tomaiuolo (2012). *U-content. The Information Professionals' Guide to User-Generated Content*. Medford: Information Today Inc.

niedopuszczalne utożsamianie z informacją literatury i sztuki oraz, co najgorsze: brak praktycznego pożytku z bardzo wielu rozważań. To wymaga zmiany.

Poczynając od uświadomienia, że komunikacja jednoznaczna po obu stronach, czyli nadania oraz odbioru, jest możliwa wyłącznie w trybie cyfrowym i tylko pomiędzy maszynami. Jednak na początku oraz na końcu każdego procesu komunikowania znajdują się **ludzie**. Różni – przyjmując, że nie zwracają się sami do siebie. Otóż zawsze dokonują przetworzeń, nazywanych interpretacjami i (również zawsze) robią to po swojemu. Jeśli więc nie rozpozna się tego, to o procesach obiegu informacji będzie wiadomo niewiele, mało lub tyle, co nic.

Oczywiście, z zanagera wyskakuje pytanie: jakim tedy cudem porozumiewamy się jednak i nawet tworzymy wspólne zasoby treści, zarówno informacyjnych, jak też nieinformacyjnych? Oraz w jaki sposób dociekamy obiektywnej **prawdy**, bez której nie byłoby informacji, nauki, ani nawet wiedzy, pojmowanej intersubiektywnie? To powinien być ważny przedmiot dociekań nauki o informacji.

Od dawna zajmują się nim różne inne dyscypliny i subdyscypliny. Filozofia języka, socjolingwistyka i psycholingwistyka, neuronauka, semiozofia oraz socjologia języka. Ale z tego wcale nie wyłania się jednorodny obraz zjawisk. Przeciwnie: powstała mozaika różnych podejść oraz opinii, często sprzecznych wzajemnie, a przeważnie kontrowersyjnych. W dodatku skupiona głównie na procesach komunikowania, a nie odbioru.

Dlatego sygnalizuję tu świeżą publikację na ten temat, która w znacznym stopniu rekapitułuje dotychczasowe opinie, formułowane przez filozofów języka werbalnego. Do ewentualnego rozciągnięcia na inne formy komunikowania – aczkolwiek z ich rozróżnieniem, zwłaszcza pisemnej od digitalnej – mimo całego rejestru wątpliwości, ale jednak częściowo produktywnych. Które warto poddać refleksji, także dla pełniejszego zrozumienia procesów informacyjnych.

Otóż dwaj profesorowie filozofii – **Herman Cappelen** z uniwersytetów w Oslo oraz w szkockim St. Andrews i **Josh Dever** w Austin (USA) – **odświeżyli** wzbogacająco rozważania na temat konstytuowania się znaczeń w komunikacji (werbalnej, ale nie tylko). Dokładniej: mowa o jednoznaczności wieloznaczności, z odniesieniem zatem również do **prawdy**, w zależności od różnych **kontekstów**. Wszystko zaś dotyczy także informacji, która musi być prawdziwa, no i ważne jest, jak kształtują się jej znaczenia. Klarownych wyjaśnień w tej książce nie ma, ale jest zachęta, albo nawet prowokacja, do głębszych przemyśleń.

Rzecz w tym, że znaczenie każdego sygnału znaczącego, komunikatu, wypowiedzi nie jest sztywne, zawsze identyczne, bo może zmieniać się stosownie do kontekstu. Jakiego oraz czyjego? Nadawcy, odbiorcy, otoczenia i okoliczności transmisji. Dlatego kontekstualność nazywa się nieraz okazjonalnością, ponieważ jej wpływ na tworzenie znaczeń bywa rozmaity. Kontekstem może być rzeczywistość, do której odnosi się komunikat, lecz także jej wyobrażenie, ale również ta, która towarzyszy nadaniu bądź odbiorowi. Jednak kontekstami mogą być same składniki wypowiedzi (słowa w różnych zdaniach mają odmienne znaczenia) albo obszary świadomości uczestników procesu komunikowania. Zresztą kontekstem bywa nawet cel sformułowania wypowiedzi albo – inny – jej przyjęcia.

W rezultacie sygnały, słowa, wypowiedzi oraz (właśnie) informacje mogą mieć bardzo szeroką gamę znaczeń. Nie do rozpoznania w rejestrze pełnym, kompletnym. Komunikaty, formułowane z ich pomocą w jednym kontekście, są odbierane w kontekstach innych, też wzajemnie tak dalece różnorodnych, że nie do zidentyfikowania w całości. No więc te okoliczności relatywizują konkretyzację sensów, zatem i to, co uważa się za **prawdę**. A relatywizowana prawda: czy rzeczywiście jest prawdą prawdziwą?

W tych okolicznościach pojawia się kategoria **prawdopodobieństwa**, ale to już jednak nie to samo, co prawdziwość, bowiem podobieństwo to nie identyczność, ani tożsamość. A jednak prawda przecież istnieje, zaś w komunikowaniu zachodzi również częściowa przynajmniej dosłowność, czyli jednoznaczność: identyczność desygnacji po stronie nadania oraz odbioru. Tyle, że nie konstytuuje się automatycznie.

W dyskusjach, które zresztą nie prowadzą do wiążących rozstrzygnięć, pojawiają się wskazania na komunikację cyfrową, rzekomo jednoznaczną, bo dwa to dwa, a nie siedem albo jeden. Ale to

jest rozumowanie nieuprawnione. W czystej komunikacji digitalnej występują jedynie znaki, sygnały, kody, pozbawione konotacji. Denotacji dokonują **zewnątrzni** (poza urządzeniami) uczestnicy procesów komunikowania: ludzie – decydując o desygnacjach oraz o znaczeniach. I (wracając do punktu wyjścia) trzeba przypomnieć, że każdy robi to po swojemu. Dlatego sam opis i model procesu komunikacji lub transmisji informacji objaśnia niewiele albo jeszcze mniej.

Jednak z różnych wypowiedzi nauwców konstatuje się wskazanie na konieczne minimum interkontekstowej stabilności, więc elementarnej **stałości** znaczeń. Jest taki podstawowy obszar semantyczny co najmniej niektórych sygnałów, znaków, słów i wypowiedzi, przynajmniej częściowo interkontekstualny: oznaczający dla wielu (dla wszystkich?) to samo. Słowo **JA** może oznaczać konkretnie każdego nadawcę indywidualnie, ogólnie natomiast zawsze wskazuje na autora komunikatu.

Przydałoby się ustalenie, czy w komunikacji informacyjnej również istnieje obszar znaczeń niezmiennych, interkontekstowych, dla wszystkich ten sam. Na wyczcucie: musi, bo to wszak jest podstawa porozumienia. Ale jak to dookreślić?

Muszę dopowiedzieć, że chociaż rozważania obu autorów odnoszą się nie tylko do komunikacji ściśle informacyjnej, umownie biorąc faktograficznej – jest bowiem w tej książce mowa też o wyobraźni, o emocjach, oraz o innych kojarzonych procesach mentalnych – to jednak całkowicie pomijają komunikację artystyczną oraz literacką. Ta zaś już z natury opiera się na wielodesygnacyjności, nastawiona multikontekstowo oraz na znaczenia pośrednie, metaforyczne i symboliczne.

Oczywiście, jedna książka nie może zajmować się wszystkim, zwłaszcza że z góry jest nakierowana na komunikację werbalną. A jednak przydałoby się chociaż sygnał, wskazujący na istnienie (wszak w trybie oralnym również) komunikacji nieinformacyjnej, co uproszczony paninformacjonizm najwyraźniej ignoruje. Bo to jest zupełnie inna formuła komunikowania: obu nie wolno wzajemnie mylić.

Desygnaty odpowiadające znakom, wypowiedziom, komunikatom, treściom przekazów, mogą być mniej lub bardziej liczne; stąd bierze się wieloznaczność. Bywają konkretne, referencyjne odniesienia do realnej rzeczywistości i wtedy zdarza się, że desygnat jest tylko jeden. Ale są też abstrakcyjne i atrybutywne, wskazujące na cechy **klasy** obiektów: wówczas jest ich więcej, nieraz bardzo dużo i trzeba wybierać. Pomocne w wyborze mogą być nadawcze wskazania, lub uzgadniające omówienia, czasem zaś (lecz tylko w bezpośredniej przestrzeni wspólnej) także to, co znajduje się w otoczeniu. Ale tak w większości procesów informacyjnych nie jest. Nawet w wymianie opinii specjalistów z tego samego zakresu.

Wobec tego **kierunkowe** nastawienia interpretacyjne biorą się z kontekstów, tych jednak jest mnóstwo. Bywają obiektywne bądź subiektywne, nadawcze albo odbiorcze, pragmatyczne lub semantyczne – w końcu kontekstem jest nawet samo mówienie, pisanie, albo klikanie – a egzystują przecież wszystkie razem, obok siebie. Nikt jednak nie potrafi rozpoznać całego kompletu kontekstów, który w dodatku nie jest zamknięty: zawsze mogą pojawić się konteksty nowe, wcześniej nieprzewidywane.

Sztuczne reguły skojarzeniowe, konieczne dla jednoznaczności, mogą ewentualnie odnosić się do jednego konkretnego desygnatu, ale wobec grupy lub klasy desygnowanych obiektów muszą być elastyczne. Dochodzenie zatem do informacyjnej prawdy opiera się na zakładanej wspólnocie kontekstu i zbieżnych odniesień w nadaniu oraz w odbiorze. Lecz jest to możliwe w wymiarze ograniczonym, natomiast w powszechnym: całkiem wyjątkowo.

Pomijając różne inne warianty odbioru, w przyjęciu komunikatów informacyjnych generalnie (choć nie tylko) chodzi o to, żeby dowiedzieć się, co sygnalizuje nadawca oraz konkretny komunikat. Ma też miejsce **weryfikacja** prawdziwości odebranych treści. Ale określenie **przejęcie** oznacza interpretację, czyli przekład tych treści na własne kategorie rozumowania odbiorcy. Jak to wyjaśnić i doprecyzować?

Kluczem jest – jak się wydaje – tryb kontekstualny. Umożliwia mianowicie zestawienie przejmowanych treści z kontekstem rzeczywistości realnej lub prawdopodobnej. W tym drugim przypadku weryfikacyjnym regułem prawdziwości poddawana jest nie tyle sama informacja, co raczej (właśnie: prawdopodobna) intencja nadania. Co oczywiście nie brzmi najlepiej. Można bowiem zapytać, o której

байдз о чыjej рэчывістасці мова, а такжэ прыпамніць, жэ кантэксты аферуюць не адну мажлівасць заставіць, але **сіатку** мажлівасці. Но і на тым polega, ціагале істніеячы, сплот wąплівасці.

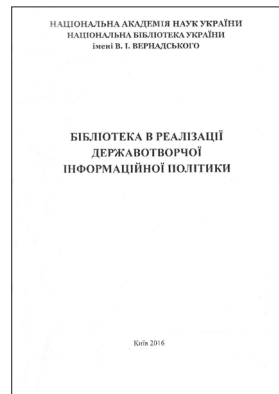
Не без поведу затем в методології познавчэj адрўнна сїе аналітычну канцэпцїе семантычну – доціеканіе ко jest **nadane** (повіедзіане) – од канцэпціj прагматычнэj: усталеніе ко jest **przyjęte** (зрозуміане). І длатего абай аўорцы тей ксіажка далецы сїа од сугестії, жэ всьшткю в тым пакіеціе проблемўв удаło сїе жуз выжаініць.

А дла науки о інфармаціj позостаюць до бліжшэго ропзнанія – коніечніе в прырўнанію до усталеніннх дысцыплін науковух – варункі оптыманей трансмісіj комунікатўв інфармаціjнх **поміежду людзьмі**. Поjмованей в категоріях (і прыз прызмат) једнозначноці ораз правдзівоці. О тым бовіем вядомо становчю за маło.

Мам ачывісїе швіадомасць, жэ рефлексје філозофўв лінгвістыкї не мусяць бїць шэрэj знане спецялістам з zakresу інфармаціj науковей, вїець влїасніе длатего тутай је сыгналізуюць. Односяць сїе бовіем рўвнїець до комунікаціj інфармаціjнэj і сїа дла неj бардо вазане. Длатего ко најмнїеj треба міець је **на оку**. Порзучаюць злўдзенія ко до комунікаціjнх једнозначноці.

### Infocyberwojna [\*\*\*]

T. Dubas, L. Sisina, O. Fedorenko, J. Szłapak, red. (2016). *Biblioteki w realizacji dierżawotworczoj informacijnoj polityki*. Kijiw: Nacjonalna Akademia Nauk Ukraini, Nacjonalna Biblioteka Ukraini im. W. I. Wernadского, ss. 324. ISBN 978-966-02-7909-4



Autoteliczne często nastawienie nauki o informacji, mianowicie głównie na siebie oraz na własne problemy wewnętrzne, w niektórych krajach zostało brutalnie przerwane, a przynajmniej naruszone: z częściowym zamustrowaniem do zupełnie innych zadań. Niekoniecznie akceptowalnych, bowiem paramilitarnych, a nawet w pełni bojowych – od początkowo propagandowych, następnie hakerskich, potem sabotażowych i w końcu ewidentnie agresywnych. Świat zaczął sobie zdawać z tego sprawę z opóźnieniem, niechętnie i bardzo mgliście; dopiero wybory prezydenckie w Stanach Zjednoczonych ujawniły stopień zagrożenia oraz bezradność. Nikt nie był (i nie jest?) przygotowany na coś takiego, na ogół zabrakło więc pomysłów, jak się bronić.

Oto więc niepostrzeżenie w obiegu informacji eksplodowała praktyka dezinformacji oraz napastliwej manipulacji, do której trzeba się odnieść i przygotować co najmniej defensywnie. Lecz wygląda na to, że sama bierna obrona już nie wystarczy.

Sprawy zaszły bowiem znacznie dalej, osiągnając stan informacyjnej **cyberwojny**, którą na równi z artyleryjskim i raketowym ostrzałem, zastosowali agresorzy na Ukrainie. No i okazało się, że ten nie tak mały kraj – brutalnie dopowiem, że jak każdy inny – nie znalazł sposobów zaradczych. Nigdzie bowiem jeszcze nie istnieje obrona antydezinformacyjna. Trzeba ją dopiero wymyślić, stworzyć i praktycznie wdrożyć.

O czymś takim zaczęto myśleć w Kijowie, ale dopiero po czasie, kiedy już utrata Krymu, Doniecka oraz Ługańska stała się faktem – tak jak i ogólny zamęt w głowach. Na początek pojawiły się pomysły nerwowe i proste, jak koncepcja ocenzurowania Internetu i całej komunikacji publicznej. A dopiero teraz generują się przemyślenia głębsze, przy udziale także znawców komunikacji informacyjnej, zmierzające do wypracowania **strategii** obronnej – czego świadectwem jest sygnalizowana tu książka.

Ważna, ponieważ samo zjawisko na pewno nie jest incydentalne i nigdzie nikt nie jest wolny od zagrożeń, nawet jeżeli tak myśli. Natomiast ta ważność nie konkretyzuje się wystarczająco klarownie. I to nie tylko dlatego, że książka jest w języku ukraińskim.

Przed wszystkim bowiem (piszę to z przykrością) jest katastrofalna, a nawet koszmarna, redakcyjnie oraz edytorsko. Wydawane tam książki, odnoszące się do bibliotekarstwa oraz informacji naukowej, już z wyglądu przypominają produkty chałupnicze, mimo że inne wydaje się na Ukrainie jak wszędzie. Jednak tym razem zgroza jest znaczniejsza. Oto bowiem autorów (bez imion) wyliczono tylko... na odwrocie strony tytułowej, a ponieważ nawet podrozdziały bywają wieloautorskie, absolutnie nie wiadomo, co kto napisał. Z powodu zaś autorskiej mnogości te podrozdziałowe niby-całości bywają treściowo niespójne. Trudno zgadnąć, co tam robiła czwórka redaktorów, wymieniona zresztą (dla odmiany?) na ostatniej stronie książki. To jest jakiś edytorski bzik.

Układ całości niewiele ma wspólnego z logiką: to chaos, pełen nawrotów i dziwnych **odskoków** treściowych. Największe jednak zdumienie budzi fakt, że w jednej książce są... dwie książki. Ta druga (70 stron), lekceważąco określona w spisie treści jako  **dodatek**, ma jednak wewnątrz tytuł, autorów oraz zatytułowane segmenty. W obu częściach jest sporo przypisów, jednak zachodnich mniej niż kot napłakał i przeważnie stare. Ponieważ zaś tylko lokalnie, trudno byłoby odnieść się produktywnie do dramatycznej rzeczywistości – jakieś zagraniczne nawiązania są, ale głównie... ze słyszenia.

Jednak najgorszy jest typowo postsowiecki bełkot stylistyczny. W humanistyce sowieckiej najważniejsze było nazwanie czegoś możliwie zawile i niby **uczenie**, a potem operowanie tą nazwą jak cepem. W ten sposób klamka stawała się **kinetycznym urządzeniem drzwiowym** i taki frazes powtarzał się w pełnym brzmieniu co kilka zdań. Otóż tutaj jest podobnie, tyle że po ukraińsku. Dlatego czytanie tej książki jest katogą i boję się, że mało kto zechce przez to przejść.

No i to jest fatalne. Bowiem problematyka tam sygnalizowana ma charakter fundamentalny i to w skali już właściwie globalnej, zaś dramatyczne doświadczenia ukraińskie to **dorobek** absolutnie unikalny. Miałby przejść niezauważony? Zwłaszcza, że w niektórych wypowiedziach nie brakuje przemyśleń oraz sugestii ważnych. Tyle, że trzeba przedzierać się do nich przez zwały sloganów, pustosłowia i edytorskiej oraz pisarskiej nieporadności.

No więc na Ukrainie uświadomiono sobie, że równoległe do działań militarnych, Rosjanie prowadzą wojnę informacyjną i są w niej jeszcze skuteczniejsi niż na polu walki. To wprawdzie nie jedyne imperium, które poczyni sobie w ten sposób, lecz inni nie mogą się równać ani skalą przedsięwzięć, ani profesjonalizmem, ani efektywnością. Dodatkowy dramat polegał zaś na tym, że w pierwszym okresie agresji nikt nie miał pojęcia, co robić. O żadnym rewanzu nie mogło być mowy, ponieważ już od 2000 r. Rosja ma doktrynę narodowego bezpieczeństwa informacyjnego<sup>3</sup> i na ewentualne reakcje zwrotne jest odpowiednio przygotowana. W grę zatem wchodziła i wchodzi wyłącznie obrona, którą jednak dopiero od niedawna zaczyna się obmyślać i przygotowywać.

Do świadomości – przynajmniej znawców i ewentualnie polityków – dotarło, że trzeba zadbać o narodową **tożsamość**. Chronić ją, odbudować albo wręcz zbudować na nowo, przeciwdziałając **dopełnianiu** w sabotażowych intencjach oraz innym formom destrukcji. Jednak co do sposobu praktycznego postępowania nie ma ani jednolitej koncepcji strategicznej, ani zgodnych pomysłów dyrektywnych. Zwłaszcza takich, które dawałyby szansę na sukces.

Wymyślono tyle, że władze muszą zorganizować system narodowej informacji w skali kraju i nawiązać w tym trybie dialog ze społeczeństwem – jednak bez funkcji nadzorca, albo cenzora, bo przed tym przestrzega się na całym świecie<sup>4</sup>. Jednak w konkretnej sytuacji wygląda to mało realnie. Już za późno na działanie w rękawiczkach: jest wojna. I brakuje narzędzi walki.

<sup>3</sup> W. Klujew (2007). *Uprawlenczeskaja ekonomika rossijskoj kultury. Tematiczeskij sbornik izbrannych rabot*. Moskwa: Izdatielstwo FAIR, 225.

<sup>4</sup> IFLA (2016). *IFLA Trend Report 2016 Update* [online]. IFLA, [23.12.2017], <https://trends.ifla.org/files/trends/assets/trend-report-2016-update.pdf>; UNESCO (2015). *Keystones to Foster Inclusive Knowledge Societies. Access to Information and Knowledge, Freedom of Expression, Privacy and Ethics on a Global Internet. Final Study* [online]. UNESCO, [23.12.2017], <http://unesdoc.unesco.org/images/0023/002325/232563E.pdf>



Nawiasem mówiąc, w całej tej wieloautorskiej książce, ani jedna osoba nie odniosła się do języków. W kraju, gdzie znaczny odsetek mieszkańców (niekoniecznie antagonistów władzy) mówi po ukraińsku słabo lub wcale? Do zgody – przynajmniej w tej książce – doszło natomiast w takim zakresie, że cały ten narodowy system informacji trzeba budować w oparciu o biblioteki. A na ile to realne – o tym nieco dalej.

W założeniu, bardzo ogólnym, sugeruje się stworzenie drożnego systemu informacji elektronicznej w wymiarze ogólnokrajowym, o charakterze państwowo-prywatnym (ale nikt nie wyjaśnia, jak to pożenić), który zapewniłby szeroki i sprawny obieg informacji. A także: suwerenność tej informacji oraz odporność na **infekcje** i na wrogie manipulacje. Jednak to tylko katalog koniecznych cech.

Hasłowość tego wszystkiego jest ewidentna, praktycyzm mglisty, a co najgorsze: w świadomości pomysłodawców istnieje informacja **wyłącznie** internetowa. To koncepcja całkowicie błędna i niezyciowa. Bezpieczeństwo można bowiem informacji zapewnić – zresztą jedynie częściowe – tylko przez **wielosemiotyczne** (równoległe digitalne oraz piśmiennicze) jej rozpowszechnianie. Przy jednym systemie, akurat elektronicznym, nie ma na to szans.

Na razie powołano Ministerstwo Polityki Informacyjnej Ukrainy, które ma dążyć do **konsolidacji informacji narodowej** oraz do **patriotycznego wychowania młodzieży** (jedno i drugie nie brzmi najlepiej), a także dbać o bezpieczeństwo informacyjne i przeciwdziałać wrogim ingerencjom w tym zakresie. Niestety, są to nie tylko drętwe hasła, bez wskazań praktycznych, ale na dodatek sygnalizują biurokratyzację i uwłaszczenie informacji publicznej na szczeblu rządowym. A jak zmieni się władza – bo tak przecież bywa – to wszystko trzeba będzie przerabiać na nowo? To nie jest dobra perspektywa.

Poza tym, nie bez racji, postuluje się przedsięwzięcia informacyjne wobec zagranicy. W tym celu ma być powołany specjalny Instytut im. Tarasa Szewczenki, a na razie pochwały zbierają telewizyjne programy na użytek zewnętrzny, w rodzaju „Ukraine Today” i „Ukraine Tomorrow”. Owszem, tylko czy ktoś w Europie o nich słyszał? Obawiam się, że efektywność jest nikła. Żeby zostać spostrzeżonym, trzeba **wepchnąć się** w serwisy już istniejące i popularne.

Tymczasem – co przyznają niektórzy autorzy – świat nie jest zorientowany w ukraińskich nieszczęściach. Według Google, tereny opanowane przez separatystów są **sporne**, ale w wersji rosyjskojęzycznej już rosyjskie, a FIFA (piłkarskie mistrzostwa świata w 2018 r. zlokalizowano w Rosji) od dawna uważa Krym za część Rosji. Na razie więc dyplomacja Ukrainy zabiega, żeby w zagranicznych mediach miasta utracone określano jednak jako ukraińskie. Ale tak nie zawsze jest.

Trudności informacyjne wystąpiły nawet w obiegu wewnętrznym – choćby terminologiczne. Na ogół unika się tam terminu wojna, stosując eufemizm **wojskowy konflikt**, a wzmianki o separatystach sygnuje się określeniem **zagraniczni najemnicy**. Zaś regiony odseparowane od Ukrainy nie mają wspólnej nazwy: rosyjski termin **Noworosja** jest oczywiście zakazany.

Zarazem jednak – żeby nie urażać mieszkających tam ludzi, którzy przeważnie w niczym nie zawinili – próbuje się eliminować rozmaite nazwy, a właściwie przezwiska, uznane za szczególnie obelżliwe, jak **Ługanda** albo **Ługambia**. Przy okazji: to jest praktyczne potwierdzenie informatologicznej opinii, że nazwy oraz terminy nie są tylko znakami, lecz w systemach organizacji wiedzy mogą występować również jako kategorie poznawcze<sup>5</sup>. Szkoda, że potwierdza to akurat taki kontekst.

Jak sugerują autorzy, w wymuszonym przez okoliczności poszukiwaniu sposobów ochrony narodowej informacji, początkowo nie doceniano potencjalnych możliwości bibliotek. Dopiero z czasem uświadomiono sobie, że biblioteczna sieć jest gęsta i biblioteki ze swoją darmową ofertą mogą być efektywnymi, a zarazem chronionymi, transmitterami informacji. Jednak potrzebują do tego znacznych środków dodatkowych i sprawnych partnerów.

<sup>5</sup> J. Woźniak-Kasparek (2015). *Terminologia jako mapa świata*. W: G. Gmiterek, M. Ochmański, M. Roszkowski (red.), *LaTeI. Z badań nad wykorzystaniem technologii informacyjnych w bibliologii i informatologii* (177–182). Warszawa: Wydaw. SBP.



Są bowiem w tej książce również doniesienia z życia, na tyle drastyczne, że kruszą nadmierne iluzje. Oto bowiem 5% tamtejszych bibliotek publicznych na wsi obywa się bez elektryczności, wiele nie ma ogrzewania (choć podany odsetek 60% {!!!} nie wydaje mi się prawdziwy), a prawie wszystkie wymagają pilnego remontu. No więc nie wystarczy, że sieć jest liczna; liczniejsza zresztą niż u nas. Jest jeszcze pytanie o jej kondycję. Opracowano program przedsięwzięć, na razie bardzo ogólny, który jednak też nie pozostawia złudzeń: trzeba zabrać się za rewitalizację całej bibliotecznej bazy. W obecnych okolicznościach to naprawdę brzmi dramatycznie.

Poza tym (powtórzę, bo to ważne) na Ukrainie – jak wszędzie – pomysł na publiczny obieg informacji powiązany prawie wyłącznie z siecią digitalną. To ogromny błąd. Który popełniają wszyscy i gdziekolwiek. Ale obecnie na Ukrainie nie jest tak, jak gdziekolwiek. Odwrotnie – gdziekolwiek może być (oby nie) tak jak na Ukrainie. Taki system bowiem da się zdemolować w ciągu minut, a najwyżej godzin. Z drugiej strony zaś: o ile w tamtejszych miastach korzysta z Internetu 75% mieszkańców, to na wsi zaledwie 24%. Więc nawet z tego powodu, przez samą elektroniczną sieć informacyjną nie zwojuje się zbyt wiele.

Jest wprawdzie w tej książce również tekst o prasie drukowanej jako narzędziu informacji publicznej, lecz jakoś dziwnie nieśmiały. Otóż na Ukrainie jest 309 gazet ogólnokrajowych oraz 1538 regionalnych i lokalnych. Mimo że niektóre przedają bardzo cienko, w sumie zatem to może być po ważny transmiter informacji do ewentualnego wykorzystania. W dodatku niektóre tytuły ukazują się w językach innych niż ukraiński i to tym bardziej powinno być brane pod uwagę. Także w trybie bibliotecznej prenumeraty, która na razie jest mizerna.

Szczególnie ważna wydaje się tamtejsza prasa rosyjskojęzyczna. Spora część ludności włada bowiem głównie tym właśnie językiem i jest jednostronnie indoktrynowana, bez skutecznej przeciwwagi. To niemały odsetek mieszkańców Ukrainy, siłą rzeczy optujących po stronie separatystów. Potrzeba więc specjalnej informacyjnej strategii. Niestandardowej.

Na razie bogatym rejestrem przedsięwzięć proinformacyjnych, w zakładanej skali ogólnokrajowej, może wykazać się kijowska Biblioteka Narodowa. Tworzy mianowicie liczne i rozmaite bazy danych, dostępne w sieci, w tym zwłaszcza **Ukrainiana**, oraz organizuje dostęp do materiałów światowych. Wszystko ma jednak charakter wyłącznie digitalny, toteż odporności na zakłócenia i sabotaż nie ma żadnej, a i praktyczna dostępność nie jest nieograniczona. W imieniu tej biblioteki pojawiła się deklaracja udziału w wychowaniu narodowym, ale bez konkretów.

Brakuje ich również wśród sygnałów, że w podobne zamierzenia zaangażowała się Fundacja Prezydentów Ukrainy oraz Narodowa Biblioteka Prawnicza. Generalnie: widać stan mobilizacji oraz mnogość intencji, natomiast relacje z rozwiązań praktycznych są skromne.

Jest też doniesienie, że regionalne ośrodki (biblioteki?) oferują specjalne materiały przez sieć osobom, które musiały wyemigrować z Krymu, Doniecka oraz Ługańska. Niby to słuszne, ale... wszyscy ci ludzie przenieśli się z laptopami? I wiedzą, że jest dla nich specjalny serwis? Rozwiązania sieciowe są bardzo wygodne, ale głównie dla nadawców; najwyższy już czas, żeby (wszędzie) zacząć myśleć o odbiorcach.

Wszystko to, niestety, wskazuje, że formalizacja i deklaratywność mają wpływ na przedsięwzięcia oraz na pomysły. Trochę w stylu niby postsowieckim, lecz nie ma pewności, czy gdziekolwiek indziej, w podobnych okolicznościach, byłoby inaczej.

Poza tym wygląda na to, że informacja polityczna rozwija się tam własnym trybem. Niedawno istniały na Ukrainie 262 partie polityczne i każda miała w sieci swój własny materiał. Do lansowania również przez biblioteki, czy jedynie do zasynalizowania? Trudno to rozstrzygnąć arbitralnie.

Wnioski z tej książki, może niezbyt liczne, natomiast dramatycznie ważne, wskazują na konieczność stworzenia na Ukrainie – a w rzeczywistości wszędzie – systemu narodowej informacji publicznej, z zasadniczym oparciem o biblioteki, które jednak trzeba (tam) pierwiej reanimować. I chociaż nikt tego tak nie wyraził, ten program trzeba wdrożyć nieomal od zera, generując również informacyjne programy regionalne, także dla obszarów aktualnie odseparowanych. Łatwo zauważyć, że w obecnych okolicznościach, szanse realizacji są **niebywale** utrudnione. Ale nie próbować po prostu nie można.

Mój poważny niepokój dotyczy jednak **elektronicznej** wyłączności referowanych projektów. Czy nikt nie widzi, że to jest informacyjne oraz polityczne **samobójstwo**? W takim trybie w jakiegokolwiek powodzenie trudno uwierzyć.

Odrębny i dramatyczny kłopot, tyle że już o charakterze uniwersalnym, polega na przypisaniu zadań koordynacyjnych centralnym władzom państwowym. W obliczu cyberwojny to niby naturalne, że tworzy się jedno wspólne dowództwo naczelne. Ale taki stan rzeczy nawet tam nie jest wieczny, a gdzie indziej na ogół go nie ma. Tymczasem polityczna centralizacja systemu informacyjnego, jeśli raz wdrożona, może okazać się trwała. A to przecież nic innego, jak knebel oraz cenzura, już ze swojej natury antydemokratyczne. W istocie niszczą w zarodku nawet samą ideę Internetu. Także **okrzyczana** nowa idea, określana jako **Internet 2**, wcale nie jest od podobnych wątpliwości wolna. Nie twierdzą, że wiem, jak te sprzeczności pogodzić, ale problem jest zasadniczy i z całą pewnością nie lokalny.

Tak ułożyły się losy, że akurat Ukraina okazała się pierwszą poważną ofiarą zmasowanej cyberwojny, która – oby nie! – może jednak zaskoczyć wszystkich i wszędzie. Byłoby więc katastrofą, gdyby dramatyczna **lekcja** ukraińska przeszła przez świat niezauważona.

To, co tam widać wymaga starannego przemyślenia i wypracowania pomysłów zaradczych. Nie tylko dla samej Ukrainy, którą trzeba wesprzeć, ale w równym stopniu na użytek własny. Bo na tym co jest, nie skończy się na pewno. A w nieszczęściu i w zaskoczeniu radzą sobie tylko przygotowani.

Otóż trudno wyobrazić sobie produktywnie przygotowanie w **tym zakresie**, bez aktywnego i mądrego współdziałania informatologów oraz informatyków – obok innych znawców. Tak w skali globalnej, jak też w wielu krajach: indywidualnie.

## Information literacy [\*]

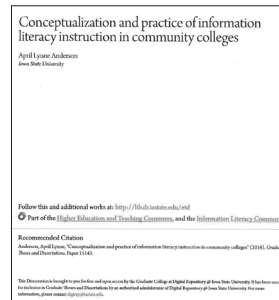
A. L. Anderson (2016). *Conceptualization and Practice of Information Literacy Instruction in Community Colleges* [online]. Ames: Iowa State University. Graduate These and Dissertations 15143, [23.12.2017], <http://lib.dr.iastate.edu/etd/15143>

Szukając informatologicznych tekstów niestandardowych do omówienia, odstających od sztampy i przez to ewentualnie ciekawszych, natrafiłem na repozytoryjną publikację rozprawy doktorskiej **April Anderson** z Uniwersytetu Stanu Iowa w Ames. I... opadła mi szczęka. Ale po kolei.

Uniwersytet w Ames (29 tys. studentów) należy wprawdzie w Stanach Zjednoczonych do grupy uczelni trzeciej kategorii, ale na światowych listach rankingowych zajmuje 421. lokatę, więc wyższą niż polskie najlepsze uczelnie, toteż porównania nie są bezzasadne. Tyle, że główną **wizytówką** tego uniwersytetu jest (pochodzący z Izraela) prof. Dan Shechtman, laureat Nagrody Nobla z chemii w 2011 r., natomiast u nas naukowego noblisty nie ma już od dawna. Ale przywołana tu rozprawa doktorska – w Polsce wszak równoważna z habilitacyjną – częściowo powinna rozładować kompleksy.

Mianowicie nie mam wątpliwości, że u nas nie ostałaby się nawet jako licencjacka, mimo że w obu dyscyplinach INB przyświeca nam wysoka tolerancja i spolegliwość. Otóż na 119 raptem stronach, z wielostronicowym opisem **narzędzi** badawczych oraz literatury przedmiotu (rejestr katastrofalny: nie ma ani jednej publikacji poważnej), doktorantka wielokrotnie powtarza jedno i to samo, zresztą nic odkrywczego oraz wplata sentymtalne reminiscencje z własnego dzieciństwa, jak też z rodzinnego życia. Czy osoby odpowiedzialne za promocję były przytomne i czy ktoś to w ogóle przeczytał? To jest absolutna parodia nauki.

Na swój sposób pouczająca i dlatego warta uwagi. Mamy bowiem powody przypuszczać, że naszej nauce, w tym właśnie dyscyplinom INB, oraz naszej edukacji, trochę lub sporo brakuje do światowego



poziomu. Jednocześnie funkcjonuje uzasadnione przeświadczenie, że amerykańska nauka dominuje w skali globalnej. Owszem, okazuje się jednak, że tylko w wąskim wydaniu elitarnym. Natomiast przywołany tu przykład zdaje się świadczyć, że przeciętna, standardowa młocka uczelniana, nie wygląda na lepszą niż nasza, a być może bywa gorsza. Sygnalizowana tu rozprawa doktorska świadczy o tym nie tylko sama sobą, ale także zawartością treściową. Bardzo to smutne, przykre i niepokojące.

Sam pomysł na dysertację nie był zapewne bez sensu. To miała być analiza realizacji przysposobienia informacyjnego (ang. *Information Literacy*, IL) na poziomie przeduniwersyteckim. Jednak wykonanie prezentuje się **fatalnie**, po części przez niefortunny wybór miejsca badań, ale też wyraźnie za sprawą autorskiej nieporadności.

April Anderson wybrała do **badania** (to jednak nie są prawdziwe badania) pięć szkół pomaturalnych, nazywanych tam **community college**, które w cyklu dwuletnim przygotowują co słabszych absolwentów szkół średnich do studiów wyższych. Tamtejsi uczniowie to najgorzej wykształcona zbiorowość młodociana: przeważnie ze wsi oraz z małych miast (55% to biali Amerykanie) i nawet po takiej czterosemestralnej preparacji, 40% nie nadaje się na żadne studia.

Przedstawiona przez autorkę panorama umiejętności – chociaż właściwie nieumiejętności – tych uczniów jest przygnębiająca. A trzeba dopowiedzieć, że nauka w tych kolegiach odbywa się nie tylko w trybie stacjonarnym, lecz także w zdalnym oraz zaocznym (kompletnie nieefektywne!), zaś wiek uczących się jest wyraźnie wyższy niż **statystyczny** wiek studentów uniwersyteckich. Tak więc całe to edukacyjne przedsięwzięcie prezentuje się kiepsko, albo bardzo kiepsko. I wprawdzie jakość edukacji szkolnej w ogóle jest w skali globalnej oceniana bardzo krytycznie<sup>6</sup>, jednak aż takiej mizerii nie sygnalizuje nikt. Zwłaszcza w odniesieniu do Stanów Zjednoczonych.

Wybrane do analizy kolegia, są ze sobą administracyjnie powiązane, ale żadnej łączności ani programowej wymiany nie ma. Co więcej: w ramach programu nauczania żadnego przysposobienia informacyjnego też nie ma, bo dyrekcje kolegiów nie były i nie są tym zainteresowane. Wdrażanie zatem umiejętności odbioru informacji jest dobrowolnym zajęciem bibliotek z tych kolegiów – więc tylko dla uczniów, którzy tego chcą, a z wyłączeniem w ogóle osób uczących się zdalnie oraz zaocznie. W istocie tedy autorka podjęła próbę przeanalizowania czegoś, co istnieje tylko w postaci szczątkowej. I taka też jest cała ta analiza.

Punkt wyjścia rozważań stanowi ocena uczniowskich kompetencji informacyjnych, które miały być rozwinięte w stopniu przynajmniej dostatecznym, w trakcie dotychczasowej nauki szkolnej. Dopowiem, że zbieżne założenie towarzyszyło również polskiemu programowi szkolnemu, dopóki istniały ścieżki edukacyjne<sup>7</sup>. Obecnie to zadanie próbują na własną rękę realizować biblioteki szkolne, publiczne oraz pedagogiczne. Czy z takim efektem jak (sygnalizowany) w Stanach Zjednoczonych? Trudno myśleć o tym bez irytacji.

No bo to, co referuje Anderson, to horror. Jej zdaniem uczniowie (absolwenci szkół średnich!) nie umieją wyszukiwać informacji, z komputerów korzystają całkowicie bezmyślnie, a nie radzą sobie z urządzeniami mobilnymi. Portret jest tak dramatyczny, że najchętniej zasygnalizowałbym brak dowodów – żadnej własnej analizy empirycznej autorka nie dokonywała – lecz obawiam się, że nie jest to czysta abstrakcja. W tym zestawieniu zaś, szumne hasła **społeczeństwa informacyjnego** lub **opartego na wiedzy** konkretyzują się absurdalnie.

Smutnemu wizerunkowi uczniowskich niekompetencji informacyjnych, odpowiada w niczym nie lepszy, bo równie mizerny, wygląd tego doktorskiego przewodu. To jest poznawcza kompromitacja, aprobowana jednak w autorytecie uczelni.

Ze względu na brak IL w dydaktycznych programach tych akurat kolegiów (po co więc autorka je wybierała i kto to zatwierdził?), Anderson przeprowadziła pięć wywiadów z bibliotekarkami z tychże pięciu kolegiów, dodając do tego nieklarowną analizę dokumentów oraz notatki z obserwacji

<sup>6</sup> G. Thomas (2013). *Education. A Very Short Introduction*. Oxford: Oxford University Press.

<sup>7</sup> M. Drzewiecki (2005). *Edukacja biblioteczna i informacyjna w polskich szkołach*. Warszawa: Wydaw. SBP.

doraźnych zajęć IL w bibliotekach. Natomiast z samymi uczniami żadnych wywiadów ani rozmów nie było (!).

Oprócz wielokrotnego powtarzania w tekście tych samych, nieskomplikowanych definicji, rozumowanie w przewodzie polegało na też powtarzalnym, bo pięciokrotnym – bez żadnej zbiorczej syntetyzacji – przyrównywaniu usłyszanych opinii oraz dostrzeżonych faktów, do schematu nauczania IL, opracowanego przez American College and Research Libraries (stowarzyszenie sprzężone z ALA). Schemat ów, który zresztą nie zawiera niczego nadzwyczajnego, został potraktowany jak nienaruszalny i superwzorcowy kanon. Dlaczego? Nie da się zgadnąć.

Określa mianowicie, że trzeba studentów (tu: uczniów) nauczyć, jak rozpoznawać własne potrzeby informacyjne oraz jak pozyskiwać potrzebne informacje, następnie zaś jak oceniać je krytycznie i pożytecznie przyswajać oraz jeszcze: pożytkować etycznie i w zgodzie z prawem. Nic tam nie wykracza zatem ponad oczywistości. Jednak autorka uczyniła z tego zgrzebnego spisu wzornik, do którego następnie przyrównała realizację bibliotecznych szkoleń IL przez pięć obserwowanych bibliotek. Z ogólną oceną pozytywną, której uzasadnienie jest żadne. Gorzej: spośród pięciu bibliotek, dwie w ogóle nie odniosły się do praktyki cytowania oraz weryfikowania informacji przyjętych.

W praktyce, w obserwowanych bibliotekach, w ramach IL dokonano demonstracji korzystania z katalogów, wyszukiwarek i baz, przede wszystkim EBSCO. Przeznaczono zaś na to raptem jedną sesję, trwającą 60–70 minut, z wykorzystaniem laptopów lub dużego ekranu centralnego. Bez żadnej oceny efektów. A ponieważ udział w takich zajęciach miał charakter dobrowolny, uczestniczył w nich mało kto.

Cały ten materiał wskazuje pozaintencjonalnie, że w wybranych kolegiach i w ich bibliotekach, przysposobienie IL było w istocie pozorowane. Jednak takiej opinii Anderson nie sformułowała, wdając się w banalne wywody, pełne sloganów i oczywistości. Nie da się zgadnąć, jakim prawem, dlaczego i po co uznano ten wywód za naukową dysertację, a co najgorsze: zaakceptowano jako podstawę doktorskiej promocji.

Nie ma tam żadnej wykreowanej wartości poznawczej. To, co ewentualnie zaprzęta uwagę – i skłania do smutnej refleksji – wynika z doniesień pośrednio i nie ma związku z autorskimi zamierzeniami. No i co szczególnie smutne: to nie jest wydarzenie jednostkowe, wyjątkowe, niepowtarzalne. Dlatego je tu przywołałem.

Trzeba bowiem chronić naukę przed takimi **dokonaniami**, ale musieliby chcieć tego wszyscy zainteresowani. Tymczasem wcale tak nie jest.

*Nadesłano: 26 kwietnia 2017*

## Przegląd polskich nowości wydawniczych

Anna Stanis

*Biblioteka Uniwersytecka w Warszawie*

G. Gmiterek, S. D. Kotuła (2017). *Aplikacje mobilne nie tylko w bibliotece*. Warszawa: Wydaw. SBP, ss. 291. Seria „Nauka, Dydaktyka, Praktyka” 175. ISBN 978-83-64203-84-8

Na świecie wzrasta coraz szybciej popularność technologii mobilnych, również w Polsce coraz więcej osób korzysta z mobilnego dostępu do zasobów i usług sieciowych. W literaturze z dziedziny bibliologiczno-informatologicznej nie było dotąd tak wszechstronnego opracowania poświęconego aplikacjom mobilnym. Publikacja jest pierwszym w Polsce naukowym przewodnikiem po mobilnych darmowych aplikacjach (także *open source'owych*) przeznaczonych do pracy w systemie Android. Skierowana głównie do humanistów, szczególnie bibliotekarzy i użytkowników bibliotek, a także wszystkich innych zainteresowanych tą problematyką.

Na treść książki składają się cztery rozdziały. Pierwszy poświęcono metodologii przeprowadzonych badań, wskazując na dużą niestabilność obszaru badawczego, wybór kryterium popularności poszczególnych aplikacji, sposób analizy aplikacji i strukturę ich opisu.

Rozdział drugi poświęcono szkicowemu zarysowi historii rozwoju komunikacji w ludzkiej cywilizacji zwracając uwagę na mobilny charakter informacji.

Rozdział trzeci zawiera opis wykorzystania urządzeń mobilnych przez czytelników i bibliotekarzy. Przedstawiono przegląd mobilnych usług bibliotecznych w bibliotekach polskich i zagranicznych (mobilne strony internetowe, mobilne aplikacje) oraz wyszukiwanie i gromadzenie danych bibliograficznych, a także dostęp do cyfrowych zasobów bibliotecznych (mobilne OPAC-i, dostęp do cennych zbiorów bibliotek oraz do treści książek i czasopism). Omówiono nowe technologie: rozszerzonej rzeczywistości (ang. *Augmented Reality*) i technologię komunikacji bliskiego pola NFC (ang. *Near Field Communication*). W rozdziale tym znalazły się także teksty o udostępnianiu czytelnikom urządzeń przenośnych oraz o otwartych danych.

Rozdział czwarty zawiera katalog aplikacji mobilnych uporządkowanych w ośmiu kategoriach głównych: Biblioteki, Czytanie, Produktywność, Nauka i edukacja, Pisanie, Multimedia, Media społecznościowe, Komunikacja. W obrębie kategorii zastosowano bardziej szczegółowe podziały (odnoszące się do poszczególnych funkcji aplikacji). Każda z aplikacji połączona jest odsyłaczem z tabelą „Parametry dotyczące popularności analizowanych aplikacji wraz z informacją o lokalizacji ich charakterystyk w książce” umieszczoną na końcu publikacji. Wszystkie wymienione aplikacje dostępne są w sklepie Google Play. Wybrano 269 aplikacji (w tym 55 *open source*), które mogą być przydatne w bibliotece. Pracę uzupełnia bardzo bogata bibliografia przedmiotu (489 poz.) oraz „Spis sieciowych źródeł aplikacji”.

**D. Drzewiecka, B. Drzewiecki, red. (2016).** „Archiwista cyfrowy” i infobroker - kształcenie, sylwetka, wyzwania. Kraków: Uniwersytet Pedagogiczny im. Komisji Edukacji Narodowej. Instytut Historii i Archiwistyki, ss.124. ISBN 978-83-60154-44-1.

Szybko postępująca komputeryzacja prac archiwalnych i bibliotecznych stawia nowe wyzwania przed archiwistami i bibliotekarzami. Przyszłość zdigitalizowanych archiwów i coraz powszechniej wprowadzanych w biblioteki dokumentów elektronicznych związana jest z ukształtowaniem nowego modelu kompetencji zawodowych absolwentów studiów bibliotekoznawczych i kształcenia w zakresie archiwistyki.

W książce przedstawiono 12 tekstów poświęconych archiwom i powstającemu zawodowi „archiwisty cyfrowego”, a także usankcjonowanemu już zawodowi infobrokera.

Pierwszy tekst o charakterze teoretycznym autorstwa Haliny Robótki, pt. *Archiwa cyfrowe i cyfrowi archiwiści w świetle nauki i wybranych dokumentów. Rozważania czysto teoretyczne* zawiera obecny stan terminologii w zakresie pojęć: „archiwum”, „archiwum cyfrowe”, „archiwista cyfrowy”, „dokument elektroniczny”. Tekst drugi, Henryka Niestroja, pt. *Racjonalizacja pracy archiwisty. Próba modelowego przedstawienia archiwisty połowy XXI wieku* zasobu przedstawia model archiwisty i archiwum w kontekście szybkiego postępu technologicznego na przykładzie wymaganych kompetencji pracownika Narodowego Archiwum Państwowego w dobie digitalizacji jego.

Artykuł Katarzyny Peplowskiej pt. *Archiwa cyfrowe – przyszłość czy rzeczywistość? Przegląd wybranych archiwów cyfrowych na świecie* zawiera rozważania na temat archiwów cyfrowych i przegląd wybranych archiwów (archiwa narodowe Wielkiej Brytanii, Szwajcarii i Estonii).

Omówienie możliwości kształcenia archiwistów, tak aby kwalifikacje zdobyte w trakcie studiów odpowiadały rozwijającej się w szybkim tempie technologii informacyjnej przedstawiła Dorota Drzewiecka w artykule zatytułowanym *Możliwości kształcenia w zakresie systemów elektronicznego zarządzania dokumentacją*.

Wanda K. Roman w tekście pt. *Komu i dlaczego potrzebne jest zarządzanie informacją?* Określa, co oznacza zarządzanie informacją i w jaki sposób nią zarządzać.

Kolejny blok czterech tekstów: Wiesława Babika: *Broker informacji – komercyjny bibliotekarz?*, Marty Dzienkiewicz: *Infobroker w małej firmie*, Justyny Trzeciakowskiej: *Infobroker – czy na pewno niezależny?*, Izabeli Gury i Angeliki Filipczyk: *Kompetencje zawodowe archiwisty w profesjonalnym wyszukiwaniu informacji – infobrokeringu* dotyczą problemów brokerów informacji i infobrokeringu.

Ostatnia część książki, trzy teksty: Huberta Chudzio: *Łączenie pokoleń. Archiwum Centrum Dokumentacji Zsytek, Wypędzeń i Przesiedleń Uniwersytetu Pedagogicznego w Krakowie*, Małgorzaty Tatarczuch i Marii Wierzbak: *Kolekcje dokumentów cyfrowych w zbiorach Biblioteki Instytutu Historii i Archiwistyki Uniwersytetu Pedagogicznego im. KEN w Krakowie*, Jarosława Sadowskiego i Mateusza Wyżgi: *Społeczności lokalne wobec wyzwań współczesności. Przykład gminy Michałowice w Małopolsce i archiwum społecznego „Michaś”*, przedstawia przegląd wybranych archiwów i repozytoriów cyfrowych działających na terenie Krakowa i Małopolski.

**B. Taraszkiewicz, red. (2016).** *Ekologia informacji a kultura informacyjna społeczeństwa*. Słupsk: Wydaw. SBP, Biblioteka Uczelniana Akademii Pomorskiej, Pedagogiczna Biblioteka Wojewódzka w Słupsku, ss. 253. Seria „Ekologia informacji”. ISBN 978-83-88783-22-7

W serii „Ekologia informacji” ukazała się już piąta publikacja poświęcona tej tematyce (pierwsza z tej serii ukazała się w 2012 r.). Książka zawiera 19 artykułów poświęconych szeroko rozumianej ekologii informacji w kontekście kultury informacyjnej społeczeństwa. Otwiera ją artykuł będący podsumowaniem pięciu edycji Ogólnopolskiej Konferencji Naukowej z cyklu „Ekologia Informacji”



zawierający analizę referatów wygłoszonych na tych konferencjach (Marcin Karwowski). Kolejny artykuł, autorstwa Wiesława Babika, poświęcony został diecie informacyjnej – zrównoważonemu podejściu do „tsunami informacyjnego” – tak aby nie wywołać w sobie efektu FOMO (ang. *Fear of Missing Out*), czyli lęku przed przegapieniem informacji. W zbiorze artykułów swoje odzwierciedlenie znalazły takie tematy, jak: etyka i kultura informacji (Oleksander Tverdokhlib, Jolanta Laskowska, Dorota Rak), umiejętności informacyjne humanistów (Katarzyna Bartosiak, Kamil Banaszewski), potrzeby, zachowania, kompetencje informacyjne (Justyna Partyka, Aneta Urbańska; Ewa Dobrogowska-Schlebusch; Maria Bosacka, Iwona Osmulka).

Interesujące jest ujęcie zagadnienia ekologii informacji z punktu widzenia botanika opisującego jak wiele nieprawdziwych, a nawet szkodliwych, informacji o ziołach znajdziemy w Internecie (Zbigniew Sobisz). Analogię pomiędzy ekologią organizmów żywych i ich gatunków a ekologią publikacji naukowych i ich gatunków wykorzystał w swoim artykule Marek Nahotko.

Omówiono także nowelizację ustawy o prawie autorskim (Krystyna Michniewicz-Wanik), technologii XXI w. w dziedzinie edukacji (Zbigniew Meger), rozpatrzono rolę biblioteki akademickiej w kreowaniu wizerunku uczelni (Małgorzata Szubrycht). W książce pojawiły się także artykuły na temat stresu informacyjnego (Radosław Molenda), mowy nienawiści na portalach społecznościowych (Jacek Prądziński) oraz poszukiwania informacji na temat ekologii informacji w Internecie (Jolanta Betkowska).

Ostatni z artykułów poświęcono sztuce *book art* (Anna Kubacka, Beata Taraszkiewicz). Różnorodność przedstawionej tematyki podkreśla znaczenie ekologii informacji i jej roli w kształtowaniu kultury informacyjnej we współczesnym społeczeństwie informacyjnym.

**M. Rowińska, I. Wiśniewska, red. (2017).** *Format MARC 21 rekordu bibliograficznego dla książki*. Warszawa: Wydaw. SBP, ss. 265. Seria „Formaty, Kartoteki” 23. ISBN 978-83-64203-88-6

Na jubileusz 15-lecia istnienia NUKAT (Centralnego Katalogu Polskich Bibliotek Naukowych i Akademickich) ukazało się nowe, zmienione, czwarte już wydanie polskiej wersji *Formatu MARC 21 rekordu bibliograficznego dla książki*. W odróżnieniu od poprzedniego wydania nowy format uwzględni aktualne światowe trendy w opracowaniu zbiorów wprowadzając elementy nowego standardu katalogowania RDA (ang. *Resource Description and Access*). Nowy standard katalogowania zastępuje drugie wydanie anglo-amerykańskich zasad katalogowania AACR2, spełnia wymogi katalogowania zasobów cyfrowych oraz tradycyjnych. Jedną z jego zalet jest możliwość prezentacji relacji pomiędzy dziełami i ich twórcami poprzez odpowiednie grupowanie rekordów bibliograficznych. Zainteresowanych standardem RDA jest wiele polskich bibliotek, szczególnie tych, które współkatalogują w NUKAT. Opracowana instrukcja katalogowania może być stosowana w bibliotekach wykorzystujących różne systemy biblioteczne.

Nowy format rozpoczyna zmiany w opisie bibliograficznym, które nastąpią przy realizacji założeń standardu RDA. Część zmian takich, jak zniesienie „zasady trzech” czy wypełnianie pól 3XX została już wprowadzona w opisach bibliograficznych NUKAT.

Zmian w formacie w stosunku do wydania poprzedniego jest bardzo dużo, począwszy od zmian terminologicznych, a na wprowadzeniu nowych pól formatu kończąc.

Zmianą terminologiczną jest zastąpienie terminu „hasło” terminem „punkt dostępu” i analogicznie główny punkt dostępu, punkt dostępu rzeczowego, dodatkowy punkt dostępu. Te terminy już od dłuższego czasu występują w literaturze przedmiotu (w przypadku punktu dostępu rzeczowego również jako punkt dostępu treściowego lub punkt dostępu do przedmiotu). W nowym formacie dodano definicje i wyjaśnienia różnych pojęć lub je doprecyzowano (np. kod przeznaczenia czytelniczego). Treść główną poprzedzono opisem źródeł informacji identyfikujących zasób. Wprowadzono

nowe, w stosunku do poprzedniego wydania, pola formatu: 013 Dane kontrolne patentu, pole 336 Typ zawartości, 337 Typ medium, 338 Typ nośnika (uzupełnione listami kodów rdamedia, rda-content, różnych typów nośników), pole 880 Zawartość pola wyrażona innym systemem pisma. Zmiany nastąpiły również w podpolu \e Określenie rodzaju współpracy (pole może być powtarzalne i występować również w polu 1XX, stosuje się w tym podpolu rozwinięte skróty). Dodano Aneksy: 1) Rekord analityczny 2) Podstawowe zasady dotyczące stosowania skrótów (generalnie rezygnuje się ze stosowania skrótów, podano wykaz skrótów, których użycie dopuszcza się w strefie uwag). Różni się znacznie Aneks zatytułowany Przykłady. W poprzednim wydaniu zawierał przykłady opisów bibliograficznych różnych typów dokumentów, w aktualnym wydaniu przykłady ilustrują zapis wybranych danych w formacie MARC 21 oraz w standardzie RDA. Drobiazgową analizą porównawczą obu wydań nie jest celem tej notatki, ale warto odnotować, że na rynku bibliotekarskim ukazała się istotna pozycja warsztatowa.

Format wydano w wersji skróconej, która już wcześniej zyskała akceptację użytkowników.

**S. Cisek, red. (2016).** *Inspiracje i innowacje: zarządzanie informacją w perspektywie bibliologii i informatologii*. Kraków: Biblioteka Jagiellońska, ss. 270. ISBN 978-83-943816-4-6 [Wersja elektroniczna: [http://ruj.uj.edu.pl/xmlui/bitstream/handle/item/31976/cisek\\_inspiracje\\_i\\_innowacje\\_zarzadzanie\\_informacja.pdf?sequence=1&isAllowed=y](http://ruj.uj.edu.pl/xmlui/bitstream/handle/item/31976/cisek_inspiracje_i_innowacje_zarzadzanie_informacja.pdf?sequence=1&isAllowed=y)]

Książka dedykowana jest pamięci Profesor Wandy Pindlowej (1933–2015), dlatego w pierwszej części znalazła się kompletna bibliografia jej publikacji (191 poz.) autorstwa Haliny Grzywacz i Justyny Buczyńskiej-Łaby. W bibliografii zawarto w porządku chronologicznym artykuły, fragmenty prac zbiorowych, publikacje naukowe, dydaktyczne, metodyczne, recenzje, sprawozdania, tłumaczenia i wywiady profesor Wandy Pindlowej. Pozostałe dwie części książki poświęcono: Zarządzaniu informacją – aspektem teoretycznym, metodologicznym, dydaktycznym i prawnym oraz Zarządzaniu informacją – aplikacjom, zjawiskom, źródłom (cz. III). Tworzy je 15 artykułów poświęconych zagadnieniu zarządzania informacją w bardzo różnorodnych ujęciach.

W cz. II znalazły się artykuły o zarządzaniu informacją z punktu widzenia bibliologii i informatologii (Marzena Świągół), o *big data* jako nowych trendach w zarządzaniu informacją (artykuły Katarzyny Majerskiej i Magdaleny Wójcik), ekologii informacji (Wiesław Babik), a także o modelach zachowań informacyjnych wraz z zaproponowaniem własnego modelu (Marek Nahotko). Problem wykorzystania techniki oprowadzania (ang. *guided tour*) w badaniach indywidualnego zarządzania informacją poruszył Remigiusz Sapa. Monika Krakowska przedstawiła najważniejsze założenia wirtualnej mobilności. Cz. II kończą próby odpowiedzi na pytanie o potrzeby i kompetencje seniorów w zakresie korzystania z nowoczesnych technologii (Dorota Rak, Marcin Petrowicz, Weronika Pokojńska) oraz uwarunkowania prawne zarządzania informacją w obszarze nauki w Polsce (Diana Pietruch-Reizes).

Część III zawiera omówienie tworzenia cyfrowych archiwów źródeł mówionych (Agnieszka Długosz-Pysz), dwa artykuły na temat bibliotek cyfrowych (Małgorzata Janiak; Aneta Januszko-Szakiel, Wojciech Kowalewski, Leszek Szafranski). Małgorzata Jaskowska omówiła polskie systemy informacyjno-wyszukiwawcze w kontekście ich dostępności dla użytkowników zagrożonych wykluczeniem cyfrowym. Maria Kocójowa podjęła problem zarządzania informacją z konferencji (studium przypadku na przykładzie działalności Wandy Pindlowej). Część III zamyka artykuł Anny Sitarskiej poświęcony systemowemu współdziałaniu bibliotek z grupami użytkowników.

Nadesłano: 12 października 2017

## Europejska, Śródziemnomorska i Bliskowschodnia Konferencja Systemów Informacyjnych – EMCIS (Coimbra, Portugalia, 7–8 września 2017 r.)

European, Mediterranean and Middle Eastern Conference on Information Systems (EMCIS) to doroczna konferencja naukowa poświęcona problemom badawczym dotyczącym systemów informacyjnych (SI). Konferencja, zainicjowana w 2004 r., gromadzi teoretyków i praktyków zajmujących się technicznymi, organizacyjnymi, ekonomicznymi i społecznymi aspektami technologii informacyjnych. Tegoroczna edycja konferencji, która odbyła się w dniach 7–8 września, zorganizowana została przez Wydział Nauki i Technologii Uniwersytetu w Coimbrze (Faculdade de Ciências e Tecnologia da Universidade de Coimbra) i zgromadziła prelegentów z 32 krajów.

Wykład inauguracyjny pt. *Organizational identity and self-awareness: creating convergence between enterprise engineering and organizational design* wygłosił prof. José M. Tribolet (Technical University of Lisbon, Portugalia). Właściwa część konferencji odbywała się w trzech równoległych blokach sesji, poświęconych następującym grupom problemów:

- Healthcare Information Systems,
- Management and Organisational Issues in Information Systems,
- Digital Services, Social Media and Digital Collaboration,
- Information Systems Security and Information Privacy Protection,
- Cloud Computing,
- Big Data and Semantic Web,
- e-Government
- IT Governance.

Dodatkowo przeprowadzono: warsztat na temat *Computer Security, Vulnerability Exploits and Ethical Hacking* oraz panel *How to Succeed on your Ph.D.*

Najciekawszą częścią konferencji z informatologicznego punktu widzenia były sesje poświęcone problematyce e-government. W jej trakcie przedstawiono pięć referatów spośród zaplanowanych siedmiu.

Muhammad Akmal Javaid (Quaid-i-Azam University, Islamabad, Pakistan) i Muhammad Irfanullah Arfeen z (United Nations University-Operating Unit on Policy Driven Electronic Governance, Guimarães, Portugalia) w swoim wystąpieniu zatytułowanym *Impact of eGovernment on Citizen Satisfaction: A Case of Federal Government Agencies in Pakistan* przedstawili wyniki badania poziomu satysfakcji użytkowników usług publicznych oferowanych przez instytucje federalne w Pakistanie. Na potrzeby badania przyjęli pięć hipotez badawczych, w ramach których założyli, że pozytywny wpływ na poziom satysfakcji użytkowników mają: 1) architektura stron internetowych, 2) zaufanie do elektronicznych usług publicznych, 3) *e-readiness* (którą rozumieją wężiej niż przyjęto w *Networked Readiness Index*, ponieważ ograniczają się tylko do infrastruktury), 4) jakość elektronicznych usług publicznych oraz 5) ich poziom zabezpieczeń. W badaniu wykorzystano kwestionariusz ankiety elektronicznej oraz przeprowadzono wywiady pogłębione (zarówno wśród użytkowników usług, jak i pracowników instytucji federalnych). W konkluzjach autorzy przedstawili zalecenia dla zwiększenia poziomu satysfakcji użytkowników z korzystania z wybranych usług

W prezentacji *Is e-Government Serving Companies or Vice-Versa?* Anton Manfreda (University of Ljubljana, Słowenia) położył nacisk na fakt, że e-government to nie tylko informatyzacja sektora publicznego, ale również biznesowy model zarządzania procesami. Bazując na studium przypadku

anonimowej firmy zarządzającej odpadami w Europie Środkowej, omówił wybrane koncepcje digitalizacyjne oraz ich wpływ na wzrost efektywności organizacji.

W wystąpieniu pod tytułem *Public Participation and Regulatory Public Policies: An Assessment from the Perspective of Douglasian Cultural Theory*, przygotowanym przez zespół w składzie Gustavo C. Garcia (Pontifical Catholic University of Rio de Janeiro, Brazylia), Flavio Saab, Jonathan S. Pereira i Paulo Bermejo (University of Brasilia, Brazylia), przedstawiono ewaluację publicznych konsultacji w kontekście zaangażowania grup interesariuszy reprezentujących różne punkty widzenia. Prelegenci w przeprowadzonym badaniu wykorzystali socjologiczną *cultural theory* Mary Douglas, która dzieli ludzi na cztery grupy: indywidualistów, hierarchistów, fatalistów i egalitarian. Za pomocą tej teorii przeanalizowali wpływ konsultacji społecznych na tworzenie polityki w zakresie regulacji rynku żywności w Brazylii. Jak wskazują ich badania, proces konsultacji angażuje jedynie niektóre grupy interesariuszy (egalitarian i hierarchistów, w niewielkim stopniu również indywidualistów), a instytucje publiczne powinny kłaść większy nacisk na pluralizm pozyskiwanych opinii. Referat ten otrzymał *Best Paper Award* w trakcie uroczystej kolacji kończącej pierwszy dzień obrad.

Mariusz Luterek z Uniwersytetu Warszawskiego, w referacie *e-Government and e-Inclusion. The Role of the Public Libraries* zaprezentował wyniki badania opinii bibliotekarzy w zakresie możliwości pośredniczenia przez biblioteki publiczne w dostępie do e-government w Polsce. Na potrzeby zaprezentowanego badania przyjęto hipotezę, iż biblioteki publiczne mogą pełnić funkcję pośrednika w dostępie do informacji i usług publicznych dla obywateli. Wyniki badania ankietowego, przedstawione w kontekście systemów pośredniczenia w dostępie do informacji publicznej i usług publicznych online w Wielkiej Brytanii, Francji, Danii i USA, pokazały, że poziom zainteresowania pełnieniem funkcji pośredniczącej w tym zakresie wśród bibliotekarzy w Polsce, niestety, jest niski.

Wreszcie, Pedro de Barros Leal Pinheiro Marino, Grazielle Isabele Cristina Silva Sucupira, Wender Rodrigues de Siqueira oraz Paulo Bermejo (Universidade de Brasília, Brazylia) w referacie *Public Transparency in the Brazilian Context: An Integrative Review* przedstawili wyniki badania bibliometrycznego w obszarze przejrzystości sektora publicznego w Brazylii. Dla zidentyfikowania różnych form transparentności, zakładanych i osiągniętych rezultatów, przeanalizowali literaturę z lat 1990–2017, dostępną w pięciu bazach elektronicznych: Scielo, Spell, Redalyc, Web of Science, oraz Science Direct. Autorom udało się zidentyfikować trzy typy transparentności: administracyjną, budżetową oraz polityczną. Uzyskane wyniki wskazują na zdecydowaną przewagę publikacji o charakterze empirycznym, prezentujących wyniki badań ilościowych.

Z punktu widzenia zainteresowań informatologów na uwagę zasługują również referaty zaprezentowane w ramach sesji *Digital Services, Social Media and Digital Collaboration*, w której przedstawiono sześć referatów. W pierwszym, zatytułowanym *MOOCS Potential for Democratizing Education: An analysis from the perspective of access to technology*, Valéria F Moura, Cesar A Souza, José D Oliveira Neto oraz Adriana B N Viana (University of São Paulo, Brazylia) przeanalizowali potencjał MOOC jako narzędzia demokratyzującego jakość edukacji. W przeprowadzonym badaniu sondażowym zebrali informacje o doświadczeniach 4784 studentów z 27 uczelni.

Następnie Jooyeon Won i DongBack Seo (Chungbuk National University, Cheongju, Republika Korei) w referacie *Relationship between self-disclosure and cyberbullying on SNSs*, wykorzystując socjologiczną teorię wymiany, na przykładzie Korei Południowej omówili zagrożenia wynikającego z nadmiernej otwartości użytkowników w sieciach społecznościowych. Z kolei Petr Štěpánek, Mouzhi Ge i Leonard Walleżký (Masaryk University, Brno, Czechy), w wystąpieniu *IT-enabled Digital Service Design Principles – Lessons Learned from Digital Cities* przeanalizowali rozwiązania zastosowane w Barcelonie w kontekście możliwości ich zastosowania w innych miastach. Warto zwrócić uwagę, że referat ten wywołał gorącą dyskusję dotyczącą różnicy pomiędzy koncepcjami *digital city* oraz *smart city*.

W ostatniej części sesji wygłoszono dwa referaty. Bruno Faria i Rui Dinis Sousa (University of Minho, 4804–533 Guimarães, Portugalia), w wystąpieniu *Enterprise Social Networks Goals and Success*

*Measurement: Evidence from a Multinational Organization* omówili wpływ narzędzia Yummer (pełni ono podobne funkcje do ogólnodostępnego serwisu LinkedIn, ogranicza się jednak do grona pracowników danej organizacji oraz oferuje dodatkowe rozwiązania dla wsparcia pracy grupowej), udostępnionego w 2014 r. przez Microsoft, na działanie organizacji. Natomiast Christer Carlsson, Anna Sell, Camilla Walden, Pirkko Walden (Åbo Akademi University, Turku, Finlandia), Siw Lundqvist oraz Leif Marcusson (Linneus University, Kalmar, Szwecja) zaprezentowali referat pt. *Go Vendla Go! Creating a Digital Coach for the Young Elderly*. Omówili w nim wyniki realizowanego od 2014 r. projektu, którego celem jest aktywizacja osób z grupy 75+ przy wykorzystaniu dedykowanego elektronicznego trenera nazwanego Vendla.

Kolejna edycja konferencji, zaplanowana na 4–5 października 2018 r., odbędzie się w Limassol na Cyprze.

Mariusz Luterek

*Katedra Informatologii*

*Wydział Dziennikarstwa, Informacji i Bibliologii*

*Uniwersytet Warszawski*

*Nadesłano: 28 grudnia 2017*

## Europejska Konferencja Edukacji Informacyjnej (Saint-Malo, Francja, 18–21 września 2017 r.)

W dniach 18–21 września 2016 r. w Saint-Malo odbyła się 5. edycja European Conference on Information Literacy (ECIL), zorganizowana przez Wydział Nauk o Informacji i Komunikacji Uniwersytetu w Zagrzebiu oraz Wydział Zarządzania Informacją Uniwersytetu Hacettepe w Ankarze, we współpracy z Information Literacy Association (InLitAs).

W konferencji uczestniczyło 300 osób z 50 krajów. ECIL to, obok angielskiego LILAC, największe i najważniejsze odbywające się w Europie coroczne spotkanie bibliotekarzy i badaczy zajmujących się kompetencjami informacyjnymi i edukacją informacyjną. Podczas czterodniowych obrad odbyło się pięć warsztatów, wygłoszono 211 referatów (w sesjach równoległych), 12 krótkich prezentacji (tzw. pecha-kucha), zaprezentowano 26 posterów, siedem prezentacji z badań doktoranckich, odbyły się cztery dyskusje panelowe. Wybrane referaty zostaną opublikowane w recenzowanym tomie pokonferencyjnym, nakładem wydawnictwa Springer w ramach serii *Communications in Computer and Information Science* (Kurbanoğlu et al., eds. 2017).

Referaty w sesjach zostały podzielone na 18 bloków tematycznych:

- Information literacy and education,
- Information literacy research,
- Information literacy in school education,
- Information literacy in the workplace,
- Information literacy for different groups,
- Information seeking and information behaviour,
- Information literacy and law,
- Information literacy in higher education,
- Information literacy instruction,
- Research data literacy,
- Information literacy and democracy,
- Science literacy,



- Information literacy and related concepts,
- Information literacy, research and education,
- Health information literacy,
- Information literacy education,
- Information literacy and ethical issues,
- Information literacy in different contexts.

W sprawozdaniu zostały omówione jedynie wybrane referaty. Szczegółowy program konferencji oraz abstrakty dostępne są na stronie ECIL 2017, <http://http://ecil2017.ilconf.org/>.

Organizatorzy zaprosili jako głównych referentów: dr Bonnie Cheuk, Stéphana Goldsteina, dr Andrew Whitwortha oraz Jean-Philippe Accarta.

Dr Bonnie Cheuk z Euroclear wygłosiła referat podczas sesji otwierającej. Mówiła w nim o roli kompetencji informacyjnych w miejscu pracy; o tym jak ważne jest przekonanie kadry zarządzającej do tworzenia środowiska pracy opartego na informacji. Przekonywała, że specjaliści od edukacji informacyjnej potrzebni są w każdej instytucji czy przedsiębiorstwie, aby szkolić pracowników i pogłębiać ich kompetencje informacyjne.

Wystąpienie Stéphana Goldsteina z InformAll również odnosiło się do kompetencji informacyjnych w miejscu pracy. Prelegent nakreślił przyszłość środowiska pracy, które już w tej chwili zmienia się w bardziej elastyczne, pozbawione ścisłej rutyny i określonego fizycznego miejsca, skupione na pracy projektowej. Według referenta wysokie umiejętności i kompetencje informacyjne są tym bardziej potrzebne w takim niestałym środowisku pracy.

Dr Andrew Whitworth z Uniwersytetu w Manchesterze przedstawił pojęcie *różnicy* – ważne ze względu na nasze osądy wiedzy wyprodukowanej w różnych kontekstach czy uwarunkowaniach. Zaprezentował podejście do koncepcji *information literacy* oparte właśnie na osadzeniu w kontekście. Zaznaczył jednak, że w takim podejściu należy za wszelką cenę unikać relatywizmu i dyletantstwa.

Z kolei Jean-Philippe Accart z Biblioteki Szkoły Hotelarskiej w Lozannie zaprezentował zastosowanie szwajcarskich standardów edukacji informacyjnej na przykładzie dwóch różnych uczelni: Uniwersytetu w Genewie oraz Szkoły Hotelarskiej w Lozannie. Pierwsza to typowa akademicka szkoła wyższa, gdzie studenci otrzymują gruntowne wykształcenie i gdzie kładziony jest duży nacisk na kompetencje informacyjne wspierające proces dydaktyczny i niezbędne do samodzielnego zgłębiania wiedzy. Druga – to wyższa szkoła zawodowa, gdzie studenci też otrzymują tytuł licencjata lub magistra, jednak szkoła ta ma przygotowywać ich do konkretnej pracy na stanowiskach kierowniczych w branży turystyczno-hotelarskiej. Do tego też niezbędne są kompetencje informacyjne, jednak w zupełnie innym kontekście: nie środowiska naukowego, tylko środowiska pracy.

Sesje plenarne i równoległe były podporządkowane określonym obszarom badań i praktyki *information literacy*.

Sirje Virkus i Sigrid Mandre z Uniwersytetu w Tallinie omówiły wyniki badań dotyczących sposobów radzenia sobie z przeciążeniem informacją (ang. *information overload*) przez studentów reprezentujących różne dyscypliny nauki. Dowodziły, że rozbudowane kompetencje informacyjne pomagają w opracowaniu skutecznych strategii w radzeniu sobie z tym zjawiskiem.

Danuta Nitecki z Drexel University przedstawiła współczesną bibliotekę akademicką jako miejsce pracy. Tradycyjna przestrzeń biblioteczna coraz częściej traktowana jest jako *learning center*, czyli miejsce pracy osób uczących się. Wykorzystując badania zachowań informacyjnych użytkowników bibliotek oraz *information literacy* rozumianą tutaj jako zestaw praktyk uwarunkowanych społecznie, Nitecki postuluje przededefiniowanie przestrzeni uczenia się i przekształcenie jej do potrzeb użytkowników.

Inne, równie ciekawe, podejście do kompetencji informacyjnych zaprezentowała Peggy Nzomo z Kent State University. Zaproponowała ona koncepcję *Multilingual Information Literacy*. Opierając się na badaniach prowadzonych na uniwersytetach kanadyjskich, pokazała jak niewystarczająca znajomość języka danego kraju ogranicza kompetencje informacyjne zagranicznych studentów przyjeżdżających tam na studia.



O potrzebie stworzenia wielokulturowego i wielojęzycznego kursu MOOC dla edukacji informacyjnej mówili Lyn Robinson i David Bawden z City University of London. W ramach międzynarodowego projektu finansowanego ze środków Erasmus+ powstaje platforma do wymiany dobrych praktyk z zakresu dydaktyki edukacji informacyjnej. Będą tam zarówno materiały dotyczące teorii, modeli czy definicji, jak i konkretne materiały dydaktyczne. Projekt jest w trakcie realizacji.

Olivier Le Deuff z Uniwersytetu Bordeaux Montaigne przedstawił koncepcję *Search Engine Literacy*. Prelegent zaczął od stwierdzenia, że wszyscy wykorzystują wyszukiwarkę Google, ale niewiele wie, na jakiej zasadzie ona działa. Zaproponował opracowanie programu dydaktycznego, który pomoże studentom zrozumieć schemat działania wyszukiwarek internetowych i krytycznie ocenić rezultaty wyszukiwania. Tym bardziej, że Google powstał z inspiracji modelami naukometrycznymi, więc bazuje na tradycji nauki o dokumentacji i informatologii. Le Deuff postuluje, by *search engine literacy* została uznana za jedną z tzw. *transliterations*, ponieważ ściśle łączy się z kompetencjami cyfrowymi, informacyjnymi i komunikacyjnymi.

Nicole Johnston i Jodi Neindorf z australijskiego Edith Cowan University oraz Meggan Houlihan z New York University w Abu Dhabi przedstawiły wyniki badań przeprowadzonych wśród zagranicznych studentów obu uniwersytetów, dotyczących ich oczekiwań względem usług informacyjnych oferowanych przez biblioteki. Studentów zagranicznych na uczelniach jest coraz więcej, w związku z powszechnie panującym trendem na umiędzynarodowianie uniwersytetów, tymczasem badań dotyczących potrzeb informacyjnych tej grupy użytkowników nie przeprowadza się wiele. Autorki zachęciły do skorzystania z ankiet i metodyki przez nie wypracowanych, przeprowadzenie badań w wielu krajach i porównanie wyników. Postulowały, że międzynarodowa współpraca bibliotek akademickich w tym zakresie jest ważna, aby wypracować jak najlepsze standardy wspierania zagranicznych studentów.

Na zagadnienie różnic w dyscyplinach naukowych i ich wpływie na to, jak studenci prowadzą wyszukiwania w źródłach informacji zwrócili uwagę w swoim wystąpieniu Jean-Pierre V. M. Héroubel i Clarence Maybee z Purdue University. Zaproponowali oni, aby szkolenia z zakresu edukacji informacyjnej połączyć z nauczaniem podstaw metodologii danej dyscypliny, tak aby od początku studenci wiedzieli, jak efektywnie wyszukiwać informację w danym kontekście.

Lindsay Roberts z University of Colorado zaprezentowała, oparty na ciekawym podejściu teoretycznym, program szkoleń z zakresu edukacji informacyjnej dla doktorantów. Założyła ona, że skoro studenci byli wystarczająco zmotywowani, by rozpocząć studia doktoranckie, należy tę motywację wykorzystać i podtrzymać, wspierając ich w pracy naukowej. Program szkoleń został oparty na teorii kompetencji informacyjnych w środowisku pracy, poczuciu własnej skuteczności, *metaliteracy*, skuteczności własnych kompetencji informacyjnych oraz teorii wyznaczania celów.

Także Hana Janečková z Politechniki w Brnie podjęła tematykę edukacji informacyjnej dla doktorantów. Jest to grupa użytkowników, która od trzech lat znajduje się w centrum oferty szkoleniowej biblioteki. Kurs zaczyna się testem wejściowym, pozwalającym określić stopień kompetencji informacyjnych doktorantów. W trakcie szkolenia, realizowanego zarówno na platformie Moodle, jak i podczas tradycyjnych zajęć, stosowane są kreatywne techniki nauczania, m.in. wykorzystywane są różne gry, tak aby każdy z uczestników był aktywnie zaangażowany w szkolenie.

Angażująca uczestników jest też oferta dydaktyczna z zakresu edukacji informacyjnej, oparta na teorii pedagogicznej *knowing-in-practice* zreferowana przez Alison Hicks z Uniwersytetu w Borås. Jest to oferowane przez bibliotekę szkolenie z zakresu edukacji informacyjnej, zakończone recenzowaną publikacją w Open Access, tzw. *open class book*, udostępnioną w uczelnianym repozytorium. Zbiorowa publikacja wszystkich studentów z danej grupy jest ustrukturyzowaną syntezą poświęconą wybranym zagadnieniom. Taka forma zaliczenia zajęć angażuje studentów, pozwala im zapoznać się z procesem wydawniczym, ponadto publikacja jest dostępna online i może być przykładem zdobytych umiejętności, który można wpisać np. do CV.

Tematyce *fake news* poświęcone było m.in. wystąpienie Johna N. Gathegi z Uniwersytetu Południowej Florydy. Przypominał on, że nie jest to zjawisko nowe, ale z pewnością znów stało się o nim głośno

w ostatnim czasie – przy okazji Brexitu oraz wyborów prezydenckich w Stanach Zjednoczonych. Gathega przedstawił koncepcję *fake news*, modele badania wpływu fałszywych informacji na opinię publiczną, sposoby edukowania obywateli w rozpoznawaniu rzetelnych informacji. Podkreślił rolę, jaką odgrywają wysokie kompetencje informacyjne w dobie nagromadzenia fałszywych informacji.

Zagadnienia poruszane na konferencjach ECIL odzwierciedlają najnowsze tendencje w badaniach kompetencji informacyjnych oraz edukacji informacyjnej i ich zastosowaniu w praktyce. W tym roku szczególnie dużo mówiono o zarządzaniu danymi naukowymi. Po raz pierwszy w historii konferencji ECIL odbyła się sesja specjalna poświęcona zarządzaniu danymi badawczymi (ang. *research data management*). Wygłoszono 20 referatów podsumowujących badania przeprowadzone w 19 krajach. W tej sesji Zuza Wiorogórska wygłosiła referat dotyczący badań (realizowanych wspólnie z Ewą Rozkosz z Dolnośląskiej Szkoły Wyższej i Jędrzejem Leśniewskim z Politechniki Wrocławskiej) przeprowadzonych wśród społeczności akademickiej Uniwersytetu Warszawskiego i Politechniki Wrocławskiej, które miały pokazać stopień świadomości respondentów, co do właściwego podejścia do danych badawczych: szczegółowego planowania procesu zarządzania nimi przed podjęciem prac badawczych, określenia strategii ich przechowywania i udostępniania. Zdefiniowanie luk w wiedzy w tym zakresie środowiska naukowego pozwoli na przygotowanie odpowiednich wytycznych i oferty szkoleniowej wspierającej naukowców w procesie badawczym.

Ciekawym uzupełnieniem wystąpień zaprezentowanych w sesji poświęconej zarządzaniu danymi badawczymi było wystąpienie Rene Schneidera z Genewskiej Szkoły Administracji Biznesem, który przedstawił program warsztatów z zakresu podstaw zarządzania danymi badawczymi w różnych dyscyplinach, ze szczególnym naciskiem na humanistykę cyfrową. Warsztaty składające się z pięciu modułów przeznaczone są dla osób, które w przyszłości same będą szkoliły innych w tym zakresie (ang. *train-the-trainers*).

Polskę na ECIL 2017 reprezentowały: Monika Krakowska i Paloma Korycińska (UJ) oraz Justyna Jasiewicz, Małgorzata Kisilowska, Anna Mierzecka i Zuza Wiorogórska (UW).

ECIL 2018 odbędzie się w dniach 24–27 września 2018 r. w fińskim Oulu.

Zuza Wiorogórska

*Katedra Informatologii*

*Wydział Dziennikarstwa, Informacji i Bibliologii*

*Uniwersytet Warszawski*

*Nadesłano: 19 grudnia 2017*

## XIV Krajowe Forum Informacji Naukowej i Technicznej (Zakopane, 20–22 września 2017 r.)

W dniach 20–22 września 2017 r. w Zakopanem, w salach konferencyjnych hotelu Starmary odbyło się XIV Krajowe Forum Informacji Naukowej i Technicznej. Głównym organizatorem tej cyklicznej konferencji jest Polskie Towarzystwo Informacji Naukowej (PTIN). Do grona współorganizatorów zaproszono w tym roku Zakład Zarządzania Informacją Instytutu Informacji Naukowej i Bibliotekoznawstwa Uniwersytetu Jagiellońskiego oraz Zakład Zarządzania Informacją Instytutu Bibliotekoznawstwa i Informacji Naukowej Uniwersytetu Śląskiego. Przy organizacji współpracowało także Międzynarodowe Towarzystwo Organizacji Wiedzy (International Society for Knowledge Organization – ISKO).

W 2017 r. przypadła 25. rocznica powstania PTIN, dlatego też podczas Forum zorganizowano specjalną sesję jubileuszową poświęconą działalności Towarzystwa, w której zaprezentowano pięć referatów. Pierwszy z nich, wygłoszony przez prezes Zarządu PTIN dr hab. Dianę Pietruch-Reizes, został poświęcony działalności PTIN w ujęciu historycznym oraz planom na przyszłość. Następne dwa referaty dotyczyły osobistości PTIN. Dr Sabina Cisek (UJ) zaprezentowała sylwetkę prof. Wandy Pindlowej – omówiła jej działalność naukową, dydaktyczną, poglądy naukowe, a także aktywność w społecznym ruchu naukowym. Drugie wystąpienie, przygotowane przez prof. Wiesława Babikę (UJ) było formą osobistego wspomnienia sylwetki prof. Eugeniusza Ścibora. Dr Renata Frączek (UŚ) zaprezentowała rys historyczny wszystkich spotkań, które odbyły się w ramach Krajowego Forum Informacji Naukowej i Technicznej – ich organizację, hasła przewodnie, poruszane, nierzadko ponadczasowe tematy. Na zakończenie tej sesji prof. Wiesław Babik – jako krajowy koordynator ISKO, prezydent ISKO-PL, omówił aktywność Międzynarodowego Towarzystwa Organizacji Wiedzy w Polsce (ISKO PL), jego formy działania oraz współpracę z Polskim Towarzystwem Informacji Naukowej. Na zakończenie tej sesji obecni członkowie PTIN oraz pozostali uczestnicy konferencji wymienili się wrażeniami i wspomnieniami z poprzednich lat.

Popołudniowa sesja plenarna otwierała merytoryczną część konferencji. Poprowadził ją prof. Wiesław Babik. W ramach tej sesji zaprezentowano cztery referaty. W pierwszym z nich prof. Barbara Sosińska-Kalata (UW) omówiła zagadnienia recepcji europejskiej koncepcji nauki o dokumentacji i informacji w Polsce na przykładzie dzieł Paula Otleta i Suzanne Briet. W referacie zaprezentowane zostały najważniejsze tezy koncepcji Otleta i Briet oraz ich powiązania ze współczesną nauką o informacji i usługami informacyjnymi. Autorka przedstawiła też wyniki swoich poszukiwań śladów recepcji tych koncepcji w pracach polskich badaczy bibliologii i informatologii, starając się ustalić jaki wpływ miały one na kształtowanie się tej dyscypliny w Polsce. W kolejnym wystąpieniu, prof. Jadwiga Woźniak-Kasperek (UW) zaprezentowała refleksje na temat bibliografii w kontekście otwartej nauki. Prelegentka dokonała analizy miejsca bibliografii we współczesnym obiegu dokumentów naukowych. Jej zdaniem bibliografia z bytu samoistnego staje się (a dokładniej: może i powinna stać się) częścią osadzoną w większej całości komunikacyjnej, bibliografią zagłębioną i zaangażowaną. Zagadnienia bezpieczeństwa informacji były tematem następnego referatu, wygłoszonego przez prof. W. Babikę. Autor przeanalizował bezpieczeństwo jako przedmiot badań nauki o informacji oraz jako istotny element działalności informacyjnej i funkcjonowania człowieka w cyberprzestrzeni, a także zaznaczył, że potrzeba szerszego zajęcia się problemem bezpieczeństwa w nauce o informacji jest oczywista, tak samo jak bezpieczeństwo informacji w społeczeństwie informacji i wiedzy. Ostatni referat w tej sesji został wygłoszony przez dr hab. D. Pietruch-Reizes (UJ). Autorka poświęciła go Europejskiej przestrzeni badawczej (EPB) w kontekście przepływu i transferu wiedzy naukowej stwierdzając, że stanowi ona najważniejszy instrument realizacji polityki badań i innowacji w Unii Europejskiej. W referacie przedstawiono postępy prac związanych z kształtowaniem EPB, zarysowano aktualną sytuację EPB, omówiono krajowe strategie wdrażania B+I (Biznes i Innowacje), ze szczególną uwagą na priorytet optymalizacji przepływu wiedzy naukowej. Autorka przeanalizowała kluczowe dokumenty Unii Europejskiej, w tym akty prawne, sprawozdania i krajowe plany działań. Przeprowadzona analiza potwierdza, że w ostatnich latach poczyniono znaczne postępy w obszarze EPB, chociaż dają się zauważyć rozbieżności między państwami członkowskimi w zakresie wdrażania poszczególnych priorytetów EPB.

W drugim dniu konferencji zaplanowano dwie sesje tematyczne. Pierwsza z nich, wzorem poprzednich spotkań, dotyczyła dydaktyki w obszarze informacji. Podczas tej sesji zaprezentowano osiem referatów. Przedstawiciele Uniwersytetu Jagiellońskiego (Małgorzata Janiak, Sabina Cisek, Diana Pietruch-Reizes), Uniwersytetu Łódzkiego (Mariola Antczak, Grzegorz Czapnik), Uniwersytetu Pedagogicznego w Krakowie (Hanna Batorowska, Stanisław Skórka), Uniwersytetu Marii Curie-Skłodowskiej w Lublinie (Anita Has-Tokarz, Renata Malesa), a także Uniwersytetu Śląskiego (Karol Makles, Renata Frączek) przedstawili nowe kierunki studiów oferowane w ośrodkach akademickich

kształcących w ramach informacji naukowej i bibliotekoznawstwa oraz współczesne wyzwania zawodowe wobec oczekiwań pracodawców.

W pierwszym wystąpieniu tej sesji M. Janiak omówiła założenia kierunku studiów „Zarządzanie informacją” prowadzonego przez Instytut Informatyki Naukowej i Bibliotekoznawstwa Uniwersytetu Jagiellońskiego od roku akademickiego 2014/2015. Autorka skupiła się na różnorodności zagadnień podejmowanych w toku studiów. W drugiej części wystąpienia autorka zaprezentowała problematykę zarządzania programem studiów w perspektywie czynników wewnętrznych i zewnętrznych. S. Cisek na podstawie analizy i krytyki piśmiennictwa (głównie anglojęzycznego) przedstawiła wybrane aspekty zagadnień kompetencji informacyjnych w miejscu pracy. Problematyka ta należy do obszernej dziedziny *information literacy* i dotyczy zdolności pracowników (na różnych stanowiskach) do efektywnego lokalizowania, oceny, organizacji i wykorzystania informacji. Obejmuje to również jej tworzenie, „opakowania” i prezentowanie właściwym odbiorcom, gdy jest to potrzebne do działań biznesowych, wykonywania zadań albo rozwiązywania problemów zawodowych. W kolejnym referacie M. Antczak oraz G. Czupnik zaprezentowali wyniki analizy oczekiwań pracodawców z obszaru zarządzania informacją oraz zaprezentowali koncepcję nowego kierunku studiów II stopnia, planowanego do wdrożenia na Uniwersytecie Łódzkim w roku akademickim 2018/19. W następnym wystąpieniu, H. Batorowska odniosła się do kluczowych kompetencji, jakie należy kształtować w edukacji dla bezpieczeństwa informacyjnego, wskazując na miejsce jakie zajmuje ono w edukacji dla bezpieczeństwa i w edukacji informacyjnej. Autorka przybliżyła szerokie rozumienie pojęcia bezpieczeństwa informacyjnego i jego znaczenie dla rozwoju kultury bezpieczeństwa, wymagającej wykształcenia w młodym i dorosłym pokoleniu wrażliwości dostrzegania każdego negatywnego zjawiska oraz umiejętności oceny jego wpływu na realizację celów jednostki, organizacji i narodu. D. Pietruch-Reizes w swoim referacie przybliżyła kwestie realizacji zajęć z zakresu ochrony własności intelektualnej oraz zagadnień pokrewnych w procesie kształcenia akademickiego specjalistów informacji. Dokonała przeglądu planów i programów studiów oraz efektów kształcenia na kierunkach związanych z informacją naukową i bibliotekoznawstwem oraz zarządzaniem informacją na uniwersytetach w Polsce. Z kolei A. Has-Tokarz oraz R. Malesa zaprezentowały doświadczenia Instytutu Informatyki Naukowej i Bibliotekoznawstwa UMCS w Lublinie związane z opracowaniem i wdrożeniem programu nowego kierunku studiów. Autorki wskazały także trudności, które wystąpiły na etapie jego projektowania (m.in. nierozumienie idei, terminologii i specyfiki dyscypliny przez pracowników Uniwersytetu) oraz wdrażania (m.in. konieczność zapewnienia odpowiedniej infrastruktury, poszerzenia zakresu prowadzonych badań i doskonalenia kompetencji kadry naukowo-dydaktycznej, poszukiwania nowych form promocji kierunku, wypracowania procedur współpracy z nową kategorią interesariuszy zewnętrznych), jak również wypracowane w tym zakresie rozwiązania praktyczne. S. Skórka omówił zagadnienia związane z kształceniem architektów informacji w Uniwersytecie Pedagogicznym KEN w Krakowie z perspektywy potrzeb pracodawców. Wskazał na praktyczny aspekt realizowanego kierunku, a także udział specjalistów – praktyków w procesie dydaktycznym. W ostatnim wystąpieniu w tej sesji, R. Frączek i K. Makles zaprezentowali nowy kierunek studiów „Informacja w instytucjach e-społeczeństwa” realizowany w Instytucie Bibliotekoznawstwa i Informatyki Naukowej Uniwersytetu Śląskiego. Omówili bloki tematyczne (moduły przedmiotów), profil absolwenta oraz osiągnięte kompetencje.

Hasłem przewodnim kolejnej sesji drugiego dnia konferencji był „dostęp do informacji”. Zaprezentowano siedem referatów, w których omówiono wybrane problemy związane z głównym tematem sesji zarówno z punktu widzenia organizacji informacji, jak i jej udostępniania szerokiemu gronu odbiorców. Barbara Kamińska-Czubała z Uniwersytetu Pedagogicznego w Krakowie omówiła możliwości wykorzystania semantycznej teorii informacji w praktyce. S. Skórka z tej samej uczelni scharakteryzował pojęcie *contentu* obejmującego dokumenty, usługi, schematy, metadane udostępniane na stronach WWW. Problematyka europejskiej polityki dostępu do informacji i archiwów była tematem wystąpienia Justyny Adamus-Kowalskiej (UŚ). Anna Matysek (UŚ) skoncentrowała się na

tematyce otwartych czasopism w naukach humanistycznych. Szczegółową analizą objęła bibliologię i informatologię. Dostęp do czasopism elektronicznych na polskim rynku prasowym rozpatrywała Agnieszka Łakomy (UŚ), natomiast Alicja Paruzel (Politechnika Częstochowska) omówiła dostęp do dysertacji naukowych na przykładzie europejskiego portalu DART-Europe. Jolanta Hys (Biblioteka Narodowa), w ostatnim wystąpieniu tej sesji, przybliżyła aktualnie prowadzone prace nad nową edycją tablic skróconych Uniwersalnej Klasyfikacji Dziesiątej.

W ostatnim dniu konferencji również zaplanowano dwie sesje. Pierwsza dotyczyła innowacji w bibliotekach. Wygłoszono podczas niej trzy referaty. Pierwszy referat dotyczył badań użytkowników w kontekście stosowanych technologii. Beata Chrapczyńska z Uniwersytetu Ekonomicznego we Wrocławiu, analizując wyniki przeprowadzonej ankiety, odpowiedziała na pytanie, czy wprowadzenie nowoczesnych rozwiązań w bibliotece wpływa na ocenę biblioteki i bibliotekarzy na uczelni. Z kolei Anna Dodot (Uniwersytet im. Adama Mickiewicza w Poznaniu) omówiła ewaluację procesu szkoleniowego jako narzędzia podnoszenia efektywności biblioteki akademickiej. Autorka zaprezentowała wyniki badania ankietowego przeprowadzonego podczas dwóch edycji konferencji szkoleniowych „Efektywna Komunikacja Naukowa”, zorganizowanych przez Bibliotekę Uniwersytecką w Poznaniu oraz Wydawnictwo Naukowe UAM. W trzecim wystąpieniu, Jolanta Szulc (UŚ) skoncentrowała się na analizie wybranych przypadków zastosowania modeli inteligencji rozproszonej w projektowaniu i rozwijaniu działalności bibliotek. Przedstawiła definicje i przykłady inteligencji rozproszonej w różnych sferach działalności człowieka. Następnie, dla ilustracji zagadnienia zaprezentowała studium przypadku (ang. *case study*) programu badania środowiska opracowanego i przeprowadzonego w Harvard Business School Knowledge and Library Services (KLS) w Bostonie oraz oprogramowania zastosowanego w bibliotece Fortune Institute of Technology (Kaohsiung City, Tajwan).

Druga sesja trzeciego dnia, a zarazem zamykająca część merytoryczną konferencji – została poświęcona działalności instytutów badawczych. Wygłoszono podczas niej pięć referatów. W pierwszym z nich Agnieszka Młodzka-Stybel z Centralnego Instytutu Ochrony Pracy – Państwowego Instytutu Badawczego (CIOP-PIB) omówiła postawy użytkowników wobec zastosowań technologii informacyjnych. Witold Sygocki, również z CIOP-PIB, przedstawił wyniki badań własnych dotyczących obecności publikacji z zakresu bezpieczeństwa i higieny pracy w Google Scholar oraz innych bazach i platformach komunikacji naukowej (w tym Web of Science Core Collection oraz Scopus). W. Sygocki wygłosił również referat w zastępstwie nieobecnej Magdaleny Zdunek (CIOP-PIB), którego tematem była analiza wyników wyszukiwania na temat wypadków przy pracy w wybranych bazach artykułów z czasopism. Joanna Szewczykiewicz (Instytut Badawczy Leśnictwa) przedstawiła innowacje wprowadzone w jej macierzystej instytucji w zakresie organizacji gromadzenia i zarządzania danymi badawczymi. Referat Sebastiana Kotuły (UMCS w Lublinie) związany był z ergonomią pracy, a w szczególności ze zjawiskiem utraty czasu w następstwie wykonywania zbędnych kliknięć myszką w procesie wyszukiwania informacji.

Każdą sesję kończyły dyskusje, niejednokrotnie kontynuowane w ramach spotkań towarzyskich w przerwach lub po zakończeniu obrad. Przebieg XIV Forum podsumowali prezes PTIN dr hab. Diana Pietruch-Reizes i prezydent ISKO-PL, prof. Wiesław Babik.

Jak zawsze – części merytorycznej towarzyszyły spotkanie kularowe, które w tym roku były uświetnione uroczystą kolacją z okazji jubileuszu Polskiego Towarzystwa Informacji Naukowej.

Renata Frączek

Polskie Towarzystwo Informacji Naukowej  
Instytut Informacji Naukowej i Bibliotekoznawstwa  
Uniwersytet Śląski, Katowice



## Wskazówki dla autorów

Redakcja *Zagadnień Informatyki Naukowej – Studiów Informatycznych* przyjmuje wyłącznie teksty wcześniej nieopublikowane i niezłożone do druku w innych czasopismach lub pracach zbiorowych. Przyjmowane są: oryginalne rozprawy i prace badawcze, recenzje oraz sprawozdania z konferencji i innych wydarzeń naukowych.

Teksty artykułów są recenzowane zgodnie z zasadami *double-blind peer review*. Zapewnienie anonimowości tekstów przekazywanych do recenzji wymaga, aby w tekście artykułu w żadnym miejscu nie była umieszczona informacja umożliwiająca identyfikację autora.

Każdy artykuł recenzowany jest na podstawie jednolitego formularza przez dwóch recenzentów doborzonych spośród specjalistów problematyki w nim poruszanej. Każda recenzja zawiera jednoznaczne wskazanie czy tekst rekomendowany jest do publikacji w *Zagadnieniach Informatyki Naukowej*. Podstawowymi kryteriami oceny artykułu są: zgodność tematu z profilem czasopisma, wartość merytoryczna, organizacja logiczna i forma językowa tekstu.

O przyjęciu tekstu do publikacji autorzy informowani są w ciągu 10 tygodni od otrzymania go przez Redakcję. Redakcja przyjmuje wyłącznie teksty przygotowane zgodnie z zasadami przedstawionymi poniżej. Teksty należy nadsyłać na adres e-mail: [zin@uw.edu.pl](mailto:zin@uw.edu.pl)

### 1. Zasady ogólne

#### 1.1. Format

Wszystkie pliki (tekst artykułu, materiały ilustracyjne) należy przesyłać jako dokumenty edytora MS WORD w formacie RTF. Zaleca się stosować w tekście czcionkę Times New Roman 12 pkt, interlinię 1.5. Tytuł artykułu należy wyróżnić czcionką Times New Roman 16 pkt. **Nie należy używać automatycznych stylów.**

Materiały ilustracyjne, wstawione w treść artykułu, dodatkowo należy przesyłać również w formacie JPG. Załączniki powinny być ponumerowane według kolejności występowania w tekście oraz zawierać nazwę, np.: *1. Tab. 1. Poziomy metadanych* albo *3. Rys. 1. Mapa myśli*.

#### 1.2. Długość tekstu

Artykuł nie powinien przekraczać 40 000, a recenzja lub sprawozdanie 14 000 znaków (ze spacjami).

#### 1.3. Strona tytułowa

Autorzy artykułów proszeni są o przygotowanie odrębnej strony tytułowej, zawierającej:

- tytuł artykułu (w językach polskim i angielskim)
- dane autora (imię i nazwisko, afiliacja – w językach polskim i angielskim)
- adres e-mail
- adres do korespondencji
- notę biograficzną autora (patrz niżej)
- abstrakt ustrukturyzowany (patrz niżej)
- słowa kluczowe (patrz niżej)
- oświadczenie o oryginalności tekstu (patrz niżej).

Zgodnie z zasadami przeciwdziałania zjawiskom *ghostwritingu* i *guest authorship* Redakcja prosi również, aby na tej stronie ujawnione zostały nazwiska i afiliacje wszystkich osób, które przyczyniły się do powstania artykułu, ich rola i udział w przygotowaniu publikacji (kto jest autorem koncepcji, założeń, metod itp. wykorzystywanych w pracy zgłoszonej do druku; procentowy udział w przeprowadzonych badaniach i opracowaniu artykułu). Redakcja prosi także o podanie informacji o źródłach finansowania publikacji, wkładzie instytucji naukowo-badawczych, stowarzyszeń i innych podmiotów (*financial disclosure*).

#### 1.4. Nota biograficzna autora / autorów

Na stronie tytułowej należy umieścić zwięzłą notę biograficzną (ok. 70 słów) każdego autora artykułu. Nota powinna zawierać następujące informacje: tytuł / stopień naukowy lub zawodowy autora, aktualne miejsce pracy i zajmowane stanowisko; specjalności naukowe lub zawodowe, najważniejsze publikacje (max. 3). Opisy publikacji powinny być sporządzone zgodnie z zasadami APA Style 6th.



### 1.5. *Abstrakt ustrukturyzowany*

Na stronie tytułowej należy umieścić abstrakt w języku polskim o objętości ok. 100 słów (ok. 1 tys. znaków) oraz jego przekład na język angielski. W abstrakcie należy wyróżnić co najmniej cztery spośród następujących kategorii informacji:

- Cel/Teza | Purpose/Thesis (*obowiązkowo*)
- Koncepcja/Metody badań | Approach/Methods (*obowiązkowo*)
- Wyniki i wnioski | Results and conclusions (*obowiązkowo*)
- Ograniczenia badań | Research limitations (*opcjonalnie*)
- Zastosowanie praktyczne | Practical implications (*opcjonalnie*)
- Oryginalność/Wartość poznawcza | Originality/Value (*obowiązkowo*)

### 1.6. *Słowa kluczowe*

Na stronie tytułowej artykułu należy umieścić od 4 do 10 słów kluczowych, w formie fraz nominalnych w mianowniku liczby pojedynczej, których pierwszy wyraz zapisany jest wielką literą, uporządkowanych alfabetycznie, rozdzielonych kropkami. Słowa kluczowe należy podać w językach polskim i angielskim.

### 1.7. *Oświadczenie o oryginalności tekstu*

Na stronie tytułowej artykułu należy umieścić oświadczenia autora /autorów, że tekst przedstawiany Redakcji *Zagadnień Informatyki Naukowej – Studiów Informatycznych* nie był dotychczas opublikowany ani zgłoszony do publikacji w żadnym innym czasopiśmie lub pracy zbiorowej. Jeśli tekst był prezentowany na konferencji, należy podać jej szczegółowe dane wraz z ewentualnymi informacjami o publikacji materiałów konferencyjnych. Jeśli artykuł jest częścią przygotowywanej do druku książki, należy podać jej dane oraz planowany termin publikacji.

## 2. *Zasady opracowania artykułu*

### 2.1. *Organizacja i podział tekstu*

Tekst artykułu powinien być podzielony na podrozdziały zaopatrzone w tytuły. W pierwszej części pod nagłówkiem **Wprowadzenie** zaleca się umieścić informacje wprowadzające w problematykę prezentowaną w artykule. W części ostatniej – pod nagłówkiem **Wnioski** lub **Zakończenie** – wnioski końcowe i podsumowanie przedstawionych rozważań.

Dopuszcza się stosowanie do trzech poziomów podziału tekstu, każdy wyodrębniony własnym śródtytułem i opatrzonego oznaczeniem numerycznym zgodnie z następującymi regułami:

#### 1. **Pierwszy poziom podziału**

##### 1.1. **Drugi poziom podziału**

###### 1.1.1 **Trzeci poziom podziału**

### 2.2. *Przypisy*

**Nie stosuje się przypisów bibliograficznych.** Odesłania do wykorzystanej literatury należy przygotować zgodnie z edytorskimi standardami tekstu naukowego APA 6<sup>th</sup> (patrz niżej).

Przypisy zawierające komentarze, dygresje, objaśnienia i inne dodatkowe informacje należy umieszczać na dole strony i numerować liczbami arabskimi; zaleca się ograniczenie liczby przypisów do niezbędnego minimum.

### 2.3. *Pisownia tytułów w tekście artykułu*

Tytuły wystaw, konferencji, programów itp. powinny być zapisane w cudzysłowie. Tytuły publikacji (książek, czasopism, artykułów itp.) należy wyróżnić kursywą.

### 2.4. *Wyróżnienia w tekście*

W tekście można stosować wyróżnienia za pomocą czcionki półgrubej (bold).

## 2.5. Materiały ilustracyjne i ich oznaczanie w tekście

Materiały ilustracyjne (tabele, wykresy itp.) powinny być przygotowane w odcieniach szarości lub kolorystyce czarno-białej. Wszystkie tego typu materiały należy oznaczyć wskazaniem rodzaju materiału (np. Tabela, Rysunek, Fotografia, Wykres), jego numeru w tekście oraz jego tytułu (np. Tab. 1. Poziomy metadanych). W odpowiednich miejscach tekstu artykułu należy umieścić odesłania do informacji prezentowanych w formie ilustracji, używając w tym celu skrótu określenia rodzaju ilustracji oraz jej numeru (np. zob. Tab. 1, zob. Wykr. 5).

## 2.6. Cytowanie wykorzystanej literatury w tekście i bibliografia załącznikowa

Cytowania w tekście i bibliografię załącznikową należy przygotować zgodnie ze standardami edytorskim publikacji naukowych APA 6<sup>th</sup>. W bibliografii załącznikowej mogą być umieszczone wyłącznie opisy publikacji cytowanych w tekście artykułu.

Publikacje należy cytować w tekście używając odsyłaczy w formie: (nazwisko, rok wydania), np. (Dembowska, 1991); gdy publikacja ma dwóch autorów należy podać obydwa nazwiska połączone znakiem ampersand (nazwisko1 & nazwisko2, rok), np. (Cisek & Sapa, 2007); gdy publikacja ma trzech i więcej autorów należy podać nazwisko pierwszego autora, skrót *et. al.* i rok wydania (nazwisko1 et al., rok), np. (Berners-Lee et al., 2001); gdy publikacja jest pracą zbiorową, należy podać nazwisko redaktora, skrót red. i rok wydania (nazwisko, red., rok), np. (Kocójowa, red., 2005). Jeśli w publikacji nie wskazano nazwiska autora lub redaktora, należy podać pierwszy wyraz tytułu i rok wydania (Wyraz, rok), np. (Biblioteki, 1976). Odwołania do określonych stron cytowanych tekstów należy podawać w formie: (Dembowska, 1991, 15), albo (Cisek & Sapa, 2007, 40–42), (Dervin & Nilan, 1986, 3) albo (Kocójowa, red., 2005, 18).

Opisy bibliograficzne wykorzystanych publikacji należy umieścić na końcu tekstu w układzie alfabetycznym, bez numeracji pozycji, pod nagłówkiem **Bibliografia**.

Opisy autorskich książek i artykułów umieszcza się pod nazwiskiem pierwszego autora. Opisy prac zbiorowych należy umieszczać pod nazwiskiem redaktora, po którym podaje się skrót *red.* lub *ed.* Jeśli w publikacji nie wskazano autora lub redaktora pracy zbiorowej, jej opis należy umieścić pod pierwszym wyrazem tytułu.

Tytuły książek i czasopism należy zapisać kursywą, tytuły artykułów w czasopismach i artykułów lub rozdziałów w książkach – czcionką prostą.

W opisach artykułów w pracach zbiorowych stosuje się oznaczenie skrótu „W” dla publikacji w języku polskim i „In” dla publikacji w językach obcych.

Opisy prac tego samego autora powinny być uporządkowane według chronologii wstępującej, a w każdym z nich należy powtórzyć nazwisko i inicjał (inicjały) autora. Prace tego samego autora opublikowane w tym samym roku należy uporządkować w kolejności alfabetycznej tytułów i oznaczać wg zasady:

Dembowska, M. (1976a) ...,

Dembowska, M. (1976b) ..., itd.

### 2.6.1 Przykłady redagowania opisów bibliograficznych

#### KSIĄŻKA

Breslin, J.G., Passant, A., Decker, S. (2009). *The Social Semantic Web*. Berlin: Heidelberg: Springer Verlag.

Dembowska, M. (1991). *Nauka o informacji naukowej: organizacja i problematyka badań w Polsce*. Warszawa: IINTE.

#### PRACA ZBIOROWA

Bellardo Hahn, T., Buckland, M., eds. (1998). *Historical Studies in Information Science*. Medford, NJ: Information Today.

Biblioteki (1976). *Biblioteki publiczne województwa toruńskiego: informator*. Toruń: Wojewódzka Biblioteka Publiczna i Książnica Miejska im. M. Kopernika.

Kocójowa, M., red. (2005). *Profesjonalna informacja w Internecie*. Kraków: Wydaw. UJ.

#### ARTYKUŁ W CZASOPIŚMIU

Dervin, B., Nilan, M. (1986). Information Needs. *Annual Review of Information Science and Technology*, 21, 3–31.

Osińska, V. (2010). Rozwój metod mapowania domen naukowych i potencjał analityczny w nim zawarty. *Zagadnienia Informatyki Naukowej*, 96(2), 41–51.

## ARTYKUŁ W PRACY ZBIOROWEJ

- Rayward, W.B. (1998). Visions of Xanadu: Paul Otlet (1868–1944) and Hypertext. In: T. Bellardo Hahn & M. Buckland (eds.). *Historical Studies in Information Science* (65–80). Medford, NJ: Information Today.
- Gawrysiak, P. (2000). W stronę inteligentnych systemów wyszukiwawczych. W: Cz. Daniłowicz (red.) *Multimedialne i sieciowe systemy informacyjne* (59–69). Wrocław: Oficyna PWR.

## ARTYKUŁ W CZASOPISIMIE ELEKTRONICZNYM

- Berners-Lee, T., Hendler, J., Lassila, O. (2001). The Semantic Web. *Scientific American* [online], May, [30.06.2013], <http://www.scientificamerican.com/article.cfm?id=the-semantic-web>
- Hollender, H. (2013). SYNAT: dziesiątki dużych i małych pomysłów na informację naukową. *Biuletyn EBIB* [online], 135(8), [15.07.2013], [http://www.ebib.pl/?page\\_id=413#art6](http://www.ebib.pl/?page_id=413#art6)
- Miller, H. (2013). Big-Data in Cloud Computing: A Taxonomy of Risks. *Information Research* [online], 18(1), [15.07.2013], <http://informationr.net/ir/18-1/paper571.html>

## HASŁA ENCYKLOPEDYCZNE

- Psychology of Culture Contact (1926). *Encyclopaedia Britannica*, Vol. 1, 13th ed. (765–771). London and New York, NY: Encyclopaedia Britannica.
- Iluminatorstwo (1971). *Encyklopedia Wiedzy o Książce* (911–952). Wrocław – Warszawa – Kraków: Zakł. Narod. im. Ossolińskich.
- Big Data (2013, November 12). *Wikipedia, The Free Encyclopedia* [online] [12.11.2013], [http://en.wikipedia.org/w/index.php?title=Big\\_data&oldid=581347727](http://en.wikipedia.org/w/index.php?title=Big_data&oldid=581347727)

Autorskie artykuły encyklopedyczne należy opisywać tak jak artykuły w pracach zbiorowych.

## DOKUMENT Z WITRYNY INSTYTUCJI, ORGANIZACJI LUB OSOBY PRYWATNEJ

- Aristotle (2009). *Organon*. From 1a to 164 a according to Bekker numbers [online]. Translated under the editorship of W.D. Ross. Internet archive [29.10.2013], [http://archive.org/stream/AristotleOrganon/AristotleOrganon-collectedWorks\\_djvu.txt](http://archive.org/stream/AristotleOrganon/AristotleOrganon-collectedWorks_djvu.txt)
- MNiSW (2011). *Narodowe Centrum Nauki w Krakowie. Nadchodzi czas nauki* [online]. Ministerstwo Nauki i Szkolnictwa Wyższego, [15.07.2013], <http://www.nauka.gov.pl/?id=2268>
- Smith, B. (2004). *Ontology and Information Systems* [online]. The Buffalo University, Department of Philosophy, [15.07.2013], <http://ontology.buffalo.edu/ontology.doc>
- US NLM (2004). *Medical Subject Headings* [online]. US National Library of Medicine. National Institutes of Health, [15.07.2013], <http://www.nlm.nih.gov/mesh/meshhome.html>

## Guidelines for Authors

ZIN – *Studia Informacyjne* (ZIN – *Information Studies*) accepts only manuscripts that have not been published before and are not under consideration for publication anywhere else. Following types of paper may be submitted for publication: original papers, book reviews, conference (and other events) reports.

Each manuscript is reviewed under a double-blind peer review process. In order to ensure the anonymity of the review process, please do not place any information in the text that could be used to identify the author.

Each manuscript is reviewed by two referees, selected on the basis of necessary expertise in the subject area under review. The review report is based on standard form containing a statement whether the manuscript is recommended for publication. Criteria for acceptance include appropriateness to the field of the Journal, scientific merit, proper text organization and correct language use.

The final decision about publication of manuscript will be sent to Author within 10 weeks after text submission. Manuscript should be formatted according to guidelines listed below and submitted via e-mail: [zin@uw.edu.pl](mailto:zin@uw.edu.pl)

### 1. General guidelines

#### 1.1. Format

All files should be submitted in RTF (Rich Text Format) files, including text and illustrative content. All pages must be typed and 1.5 spaced using 12-point Times New Roman font. The title of the manuscript should be typed 14-point font. Please do not use any preformatted styles.

Illustrative content inserted in the article, should be send also in JPG format. Attachments should be numbered in order of occurrence and include the title, for example: *1. Tab. 1. List...* or *3. Fig. 1. System...*

#### 1.2. Extent

Manuscript should be no longer than 40,000 characters (including spaces), review and report no longer than 14,000 characters.

#### 1.3. Title page

Authors should prepare **separate title page**, which include:

- **title of the paper,**
- **the name(s) of the author(s) with appropriate affiliations,**
- **he e-mail address of the corresponding author,**
- **address for correspondence,**
- **biographic note (see below),**
- **structured abstract (see below),**
- **keywords (see below),**
- **statement of originality (see below).**

According to the Journal policy against *ghostwriting* and *guest authorship*, authors are requested to list on title page names and affiliations of each person that contributed to the text (author of the idea, methods, etc. used in the submitted manuscript; percentage of contribution to the research process and text compilation). Authors are also requested to describe sources of founding that have supported the work and the financial involvement of research institutes, associations and other entities (*financial disclosure*).

#### 1.4. Author(s) biographic note

Title page should include concise biographic notes (about 70 words) of each author : academic degree or professional position, current place of work and position, area of interest, the most important publications (max. 3).

#### 1.5. Structured abstract

An abstract (about 100 words or 1000 characters) should be included with each submission and placed on the title page. Abstract should be formatted according to categories listed below. Author should identify at least four mandatory sections:

- **Purpose/Thesis** (*mandatory*)
- **Approach/Methods** (*mandatory*)
- **Results and conclusions** (*mandatory*)
- **Research limitations** (*optional*)
- **Practical implications** (*optional*)
- **Originality/Value** (*mandatory*)

### **1.6. Keywords**

Title page should include keywords (4 to 10) as a noun phrases in singular form, where first element is capitalized. Keywords in alphabetical order should be delimited by full stop.

### **1.7. Statement of originality**

Author(s) should include on title page statement that submitted text has not been published before and is not under consideration for publication anywhere else. If the paper was presented at a scientific meeting, provide detailed information about the event and the conference proceedings. If the paper will be the part of the author's book, provide its details and planned publishing date.

## **2. Manuscript Format and Preparation**

### **2.1. Body of the Paper**

The text should be organized into entitled sections and subsections. Text should start with **Introduction**, giving an overview and stating the purpose and end with **Conclusion**, giving the summary of the author contributions to the study.

Author may use three levels of headings. Each heading should have its own title and number according to the following pattern:

#### **1. First-level heading**

##### **1.1. Second-level heading**

###### **1.1.1 Third-level heading**

### **2.2. References**

Bibliographic citations are not allowed in footnotes. The reference list should be prepared according to APA 6-th Edition citation style (see below). Footnotes can be used only to give additional information or commentary. Footnotes to the text are numbered consecutively with Arabic numerals. It is recommended to limit the amount of footnotes per page.

### **2.3. Titles in the body of the text**

Titles of exhibitions, conferences, programmes, etc should be written within double quotation marks. Use italics for publication titles (books, journals, papers, etc.).

### **2.4. Emphasis**

Bold face should be used to emphasize certain words or passages.

### **2.5. Illustrative content**

All illustrations (tables, charts, figures etc.) should be converted to greyscale. All illustrations should be cited in the text properly to their form (Table, Figure, Photograph, etc.) and have title and consecutive number (e.g. Tab. 1. Metadata levels). Use abbreviation in the text when refereeing to the illustrative content (e.g. see Tab. 1, see Fig. 5).

### **2.6. Citations and reference list**

Use APA 6-th Edition as a citation and reference list format. The references list should only include works that are cited in the text.

Cite references in the text by name of the author(s) and year of publication in parentheses: (Name, Year of publication), eg. (Dembowska, 1991). If there are two authors, put their names with ampersand (&) mark

between: (Name & Name, Year of publication), eg. (Cisek & Sapa, 2007). If there are more than two authors, put the name of the first one followed by abbreviation *et al.*: (Name et al., Year of publication), eg. (Berners-Lee et al., 2001). Edited books are cited by the name(s) of the editor(s) followed by abbreviation *ed(s)*: (Name, ed., Year of publication), eg. (Bellardo Hahn & Buckland, eds., 1998). If there is no author or editor information, put the first word from the title and the year of publication: (Word, Year of publication), eg. (Biblioteki, 1976). Use the following pattern when referring to specific pages in the cited publications: (Dembowska, 1991, 15) or (Cisek & Sapa, 2007, 40–42) or (Bellardo Hahn & Buckland, eds., 1998, 18).

Place the reference list at the end of the text under the heading **References**. Reference list should be in alphabetical order without numbering.

List the references (books and journal articles) in alphabetical order by authors' last names. Citations of edited books list under the name of editor followed by abbreviation *Ed.*. If there is no author or editor information, list the publication under the first word from the title.

Use italics for book titles and regular font for titles of papers and book chapters. Use abbreviation *In*: when referring to book chapters in citations.

If there are two or more items by the same author(s), list them in order of year of publication (reverse date order). If two or more works are by the same author(s) within the same year, list them in alphabetical order by title and distinguish them by adding the letters a, b, c, ... to the year of publication:

Dembowska, M. (1976a) ....

Dembowska, M. (1976b) ....., etc.

### 2.6.1 References List Examples

#### BOOK

Breslin, J.G., Passant, A., Decker, S. (2009). *The Social Semantic Web*. Berlin: Heidelberg: Springer Verlag.

Dembowska, M. (1991). *Nauka o informacji naukowej: organizacja i problematyka badań w Polsce*. Warszawa: IINTE.

#### BOOK (EDITED)

Bellardo Hahn, T., Buckland, M., eds. (1998). *Historical Studies in Information Science*. Medford, NJ: Information Today.

Biblioteki (1976). *Biblioteki publiczne województwa toruńskiego: informator*. Toruń: Wojewódzka Biblioteka Publiczna i Książnica Miejska im. M. Kopernika.

#### JOURNAL ARTICLE

Osińska, V. (2010). Rozwój metod mapowania domen naukowych i potencjał analityczny w nim zawarty. *Zagadnienia Informatyki Naukowej*, 96(2), 41–51.

Dervin, B., Nilan, M. (1986). Information Needs. *Annual Review of Information Science and Technology*, 21, 3–31.

#### BOOK CHAPTER

Rayward, W.B. (1998). Visions of Xanadu: Paul Otlet (1868-1944) and Hypertext. In: T. Bellardo Hahn & M. Buckland (eds.). *Historical Studies in Information Science* (65–80). Medford, NJ: Information Today.

#### ELECTRONIC JOURNAL ARTICLE

Berners-Lee, T., Hendler, J., Lassila, O. (2001). The Semantic Web. *Scientific American* [online], May, [30.06.2013], <http://www.scientificamerican.com/article.cfm?id=the-semantic-web>

Hollender, H. (2013). SYNAT: dziesiątki dużych i małych pomysłów na informację naukową. *Biuletyn EBIB* [online], 135 (8), [15.07.2013], [http://www.ebib.pl/?page\\_id=413#art6](http://www.ebib.pl/?page_id=413#art6)

Miller, H. (2013). Big-Data in Cloud Computing: A Taxonomy of Risks. *Information Research* [online], 18(1), [15.07.2013], <http://informationr.net/ir/18-1/paper571.html>

#### ARTICLE IN ENCYCLOPEDIA

Psychology of Culture Contact (1926). *Encyclopaedia Britannica*, Vol. 1, 13th ed. (765–771). London and New York, NY: Encyclopaedia Britannica.



Iluminatorstwo (1971). *Encyklopedia Wiedzy o Książce* (911–952). Wrocław – Warszawa – Kraków: Zakł. Narod. im. Ossolińskich.

Big Data (2013, November 12). *Wikipedia, The Free Encyclopedia* [online] [12.11.2013], [http://en.wikipedia.org/w/index.php?title=Big\\_data&oldid=581347727](http://en.wikipedia.org/w/index.php?title=Big_data&oldid=581347727)

Article in encyclopedia with author information describe as book chapter.

**ELECTRONIC DOCUMENT FROM WEBSITE**

MNiSW (2011). *Narodowe Centrum Nauki w Krakowie. Nadchodzi czas nauki* [online]. Ministerstwo Nauki i Szkolnictwa Wyższego, [15.07.2013], <http://www.nauka.gov.pl/?id=2268>

Smith, B. (2004). *Ontology and Information Systems* [online]. The Buffalo University, Department of Philosophy, [15.07.2013], <http://ontology.buffalo.edu/ontology.doc>

US NLM (2004). *Medical Subject Headings* [online]. US National Library of Medicine. National Institutes of Health, [15.07.2013], <http://www.nlm.nih.gov/mesh/meshhome.html>

**Adres Wydawnictwa**

ul. Konopczyńskiego 5/7

00-335 Warszawa, tel. 22 827 52 96

**Prenumerata i sprzedaż**

Dział Promocji i Kolportażu SBP

Al. Niepodległości 213, 02-086 Warszawa, tel. 22 825 50 24

**Cena prenumeraty na 2017 r. – 118 zł**

Wydawnictwo SBP – Warszawa 2017. Nakład 350 egz.

Ark. wyd. 16,08. Ark. druk. 13,25.

Skład i łamanie: Justyna Grzymała-Łuszcz

Druk i oprawa: Mazowieckie Centrum Poligrafii

ul. Piłsudskiego 2A, 05-270 Marki, [www.c-p.com.pl](http://www.c-p.com.pl)

e-mail: [biuro@c-p.com.pl](mailto:biuro@c-p.com.pl), tel. 22 497 66 55

