# ZIN

## Studia Informacyjne
## Information Studies

VOL. 57 2019 NO. 1A(113A)
e-ISSN 2392-2648

WDIiB UW

SBP
STOWARZYSZENIE
BIBLIOTEKARZY
POLSKICH

Wersja elektroniczna jest wersją pierwotną czasopisma.
Zawartość czasopisma jest indeksowana w CEJSH, CSA, LISTA, Knowledge Organization Literature, PBN, PBB.

The electronic version is the original version of the journal.
The content of the journal is indexed in CEJSH, CSA, LISTA, Knowledge Organization Literature, PBN, PBB.

Pełne teksty roczników dostępne są w Archiwum SBP: www.sbp.pl/archiwumcyfrowe
oraz (do 2017 r.) w Mediotece WDIB UW: http://mediateka.uw.edu.pl/dlibra/

# ZAGADNIENIA INFORMACJI NAUKOWEJ

## Studia Informacyjne

# ISSUES IN INFORMATION SCIENCE

## Information Studies

VOL. 57 2019 NO. 1A(113A)
E-ISSN 2392–2648

WDIiB UW

SBP
STOWARZYSZENIE
BIBLIOTEKARZY
POLSKICH

WYDAWNICTWO SBP

Warszawa 2019

## ISSUES IN INFORMATION SCIENCE – INFORMATION STUDIES

The core purpose of *Issues in Information Science – Information Studies (Zagadnienia Informacji Naukowej – Studia Informacyjne,* ZIN – *Studia Informacyjne)* 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.

We publish papers in Polish or English. For each paper a set of matadata is provided: an abstract and keywords in both languages) as well as author's bio and contact information.

The subtitle of the journal – *Information Studies* – 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 – *Information Studies* 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 ZIN – *Information Studies* covers, among else, the issues of:

- information services in institutions of science, culture, business, education and administration,
- 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,
- information, digital and media 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 donators 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 the 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, Worldcat and Polish Scholarly Bibliography (PBN). The journal is registered in the European Reference Index for the Humanities (ERIH Plus).

## ZAGADNIENIA INFORMACJI NAUKOWEJ – STUDIA INFORMACYJNE

Głównym celem *Zagadnień Informacji Naukowej – Studiów Informacyjnych* (ZIN – *Studia Informacyjne*) 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.

Publikujemy artykuły w językach polskim i angielskim. Każdy artykuł posiada zestaw metadanych: abstrakt i słowa kluczowe (w obu językach) oraz nota biograficzna autora i dane do kontaktu z nim.

Czasopismo adresowane jest zarówno do czytelnika polskiego jak i zagranicznego, publikujemy artykuły zarówno w języku polskim jak i angielskim. Podtytuł czasopisma – *Studia Informacyjne* – 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:

- usług informacyjnych w instytucjach nauki, kultury, biznesu, edukacji i administracji,
- 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 różnymi aspektami funkcjonowania informacji we współczesnym świecie.

*Zagadnienia Informacji Naukowej* znajdują się na liście 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, Worldcat, Polska Bibliografia Naukowa (PBN). Czasopismo jest zarejestrowane w European Reference Index for the Humanities (ERIH Plus).

# INFORMATION SCIENCE
# TOWARDS OPEN SCIENCE

# NAUKA O INFORMACJI
# WOBEC OTWARTEJ NAUKI

# Contents | Spis treści

# Preface

Thanks to the financial support which the Ministry of Science and Higher Education provided under the program "Działalność upowszechniająca naukę: działalność wydawnicza (DUN)" (Actions to Promote Science: Publishing) to increase the national and international circulation of *ZIN. Issues in Information Science. Information Studies*, in 2019 and 2020 each year the editorial committee will release two additional thematic issues devoted to the newest topics in information science. We will publish them entirely in English, hoping that it will make their contents more accessible to the international audience.

The first thematic issue we deliver to the readers focuses on the open science movement, and the ways it challenges information science and information professionals' attitude to scientific practice and to the sharing of the results of scientific research. The term "open science" refers to a concept and to related organizational activities, supported by national and international programs to develop science, as well as by numerous scientific organizations, which aim to ensure that scientific research is available to a wide audience, including researchers from around the world, communities of academics interested in the results of scientific research, students, as well as amateurs passionate about science and curious about scientific research. The opening of science, or even more generally, the opening of knowledge, changes the way scientific research and scientific communication are conducted, as it increases the accessibility of scientific resources – scientific data, scientific publications, software, as well as academic teaching material and specialized online courses. Open science is founded on the idea of wide knowledge sharing, which follows the tradition of the first scientific revolution and first scientific journals. The goal of promoting wider knowledge sharing now is to guarantee the transparency of scientific research, especially government-funded research, and to increase the rate at which science develops. Put into practice, idea of open science creates opportunities for scientific collaboration; it accelerates the research of vital issues by facilitating the access to the newest research results for the benefit of the researchers from around the world; finally, it animates citizen science, allowing numerous science aficionados to realize their intellectual potential. If these changes are to take place, everyone interested in science must have a free access to scientific resources, the skills required to identify and use them properly, the ability to add the results of their own research to the existing resources, and the tools to actively share them with the scientific community. As far as information science is concerned, such phenomena have always been the central object of interest and research. The effective transfer of knowledge within the society remains the chief question for information science. Therefore, it comes as no surprise that the idea of open science and the practices it promotes constitute one of the most important areas of research in information science.

This issue gathers five articles which discuss various aspects of open science in relation to information science and to information management and provision of information services in academic communities. The articles are extended, and revised versions of the research papers presented at the 5th International Scientific Conference from the series "Information Science in the Age of Change", which took place in Warsaw on May 13–14,

2019. Our journal provided the conference with media patronage, while several members of its editorial committee presided over the conference's Scientific Committee and Organizational Committee. Organized biannually by the Department of Information Studies at the Faculty of Journalism, Information and Book Studies (before 2015, the Department of Information Systems at the Institute of Information and Book Studies) at the University of Warsaw, together with the Polish Chapter of the International Society for Knowledge Organization (ISKO-PL), the conference monitors the newest directions in information science and the changes in information services. This year, the University of Warsaw Library was the third co-organizer of the conference. The leading theme of this year's edition was "Digital Revolution – Today and Tomorrow. Infrastructures, Services, Users". The issues for the idea of open science and for the practices it promotes remain closely related to the technological and social changes referred to as the digital revolution. Accordingly, these issues attracted the attention of the conference speakers, whom the organizers asked to give keynote addresses on the challenges of open science for information science and information management.

The articles published in this issue of *ZIN. Issues in Information Science. Information Studies* discuss the challenges of open science on various levels of the involvement of information science and information management: they consider researchers' attitudes towards open science and their ready to take part in its development, the implications of open science for the design of systems for the research information management, as well as the effect of legal regulations and international programs on the process of the opening of science. The first two articles develop the keynote addresses, while the following three extend the research papers given during one of the two international sessions devoted to open science and access to research data, which took place on the second day of the conference.

The first article, *Challenges of Information Infrastructures for Open Science and Academic Libraries* was written by Professor Jela Steinerová, a prominent researcher of information behavior and information ecology from Department of Library and Information Science at Comenius University in Bratislava. The article discusses the concepts related to the infrastructure of information science and new models of scientific communication emerging in the connection with open science, as well as the results of the research Professor Steinerová conducted together with her team to study Slovak scientists' information behavior in the context of changes occurring in the academic community and in the contemporary scientific communication.

The author of second article is Professor Tibor Koltay from the Institute of Learning Technologies at Eszterházy Károly University, a well-known researcher of information literacy, who has recently begun to study a new current in the information activity of academic libraries i.e., the management of research data and the related trend of data literacy. In the article published in *ZIN*, *Curation in Academic Libraries as Part of the Digital Revolution*, Professor Koltay shows the importance of research data management for the information services targeting the academic community.

The new ecosystem of science, together with the information management systems serving it, is the subject of the article by Professor Katarzyna Materska from the Institute of Media Education and Journalism at Cardinal Stefan Wyszyński University in Warsaw, who has been studying information management in science for years. Her article, *Research Information Management in the Context of Open Science and Open Data*, discusses the

adjustment of Current Research Information Systems (CRIS) to accommodate collaboration between researchers promoted by open science, and to fulfill the requirements for the evaluation of scientific research set by particular institutions and countries.

The article by Zuzanna Wiorogórska from Department of Information Studies, Faculty of Journalism, Information and Book Studies at University of Warsaw (and an academic librarian at University of Warsaw Library), *Research Data: Management and Opening. Polish and European Perspectives* discusses the politics of opening and managing the research data in the light of European and Polish regulations and the policies of various international repositories. The analysis of these documents provides a foundation for a further consideration of the direction of such actions and their impact on the development of contemporary science.

The issue is closed by the article of Samia Takhtoukh from Laboratoire Geriico, University of Lille, *Exploring Humanities Research Data in Figshare*. It is devoted to the evaluation of the humanities scholars' practice of sharing their research data. The author presents the results of her empirical research of the resources of an online open access repository Figshare, which has also allowed her to analyze the affordances and limitations of the tools designed to deposit and store research data to be searched and browsed by other users. The research and the following analysis justify the author's call for a structured cultivation of an awareness of the research data management importance at universities and in research organizations, requiring an institutional encouragement of collaboration between the researchers, particularly in the areas of humanities and social science, and information professionals, whose close partnership is necessary to effectively promote the sharing of the research data, whether it be with fellow researchers, or with a wider audience.

Delivering this thematic issue of *ZIN – Information Studies* to the readers, I hope that it will serve them well, adding to their understanding of open science and of the opportunities it creates for the researchers of information science and to the information professionals.

Barbara Sosińska-Kalata
*Editor in-Chief*

*Warsaw, August 15, 2019*

# Challenges of Information Infrastructures for Open Science and Academic Libraries

Jela Steinerová
ORCID: 0000-0002-6360-0519
*Department of Library and Information Science,*
*Faculty of Arts, Comenius University in Bratislava, Slovakia*

**Abstract**

**Purpose/Thesis:** The paper reviews the main concepts of information infrastructure, information environment, and information behavior of researchers in the context of digital revolution. The concept of open science and new models of scholarly communication are considered.
**Approach/Methods:** Related studies of information infrastructure and information behavior of researchers are analyzed. We report on the results of a series of qualitative studies of information behavior and of information ecology of the academic environment in Slovakia based on interviews with doctoral students, expert researchers and information managers. We apply a previously developed original methodology of concept mapping.
**Results and conclusions:** The results allow us to re-consider relevance assessment in the digital environment and to present a model of the information ecology in the information environment. We present the analyzed data represented by the concept maps which show the attitudes of researchers to open science, economic models of science and values of researchers. The final model of academic information ecologies is explained, and a new interactive model of the academic library is presented.
**Practical implications:** We suggest the implications for the development of information science and for the education of information professionals.
**Originality/Value:** Original models emphasize values of academic research, values of information, value-added services of academic libraries, and value-sensitive design of digital libraries.

**Keywords**

Academic libraries. Information behavior of researchers. Information environment. Information infrastructure. Open science.

*Received: 31 May 2019. Reviewed: 8 June 2019. Revised: 13 July 2019. Accepted: 18 July 2019.*

## 1. Introduction

Information infrastructures are a part of information environment. The term "information infrastructure" refers to digital technologies, values, services, libraries, information professionals and users. New models of scholarly communication emerge with the development of information infrastructures. Digital revolution changes our information behavior in relation to science, education, health and workplaces. Following the trends of digital technologies, new questions emerge regarding information support of scholarly communication. That is why we ask the questions: Which values of information work are valid in digital environment? How should we re-design models of scholarly communication? What information

support is needed for open science and digital science? What changes should be reflected in new models of academic libraries?

In this paper we focus on the issues related to digital revolution, information infrastructures, open science, information environment and information behavior of researchers. We will briefly analyze related studies of information behavior of researchers and of open science. In the next section, we will discuss the results of three qualitative studies of doctoral students, information managers and researchers which we conducted in Slovakia.

We present characteristics of relevance in the digital environment and a model of information ecology in the academic information environment. Results of the study of information behavior of researchers are visualized on concept maps which represent the researchers´ responses to the survey. The selected concept maps present the perceptions of open science, economic models of science and values of research. The analysis of collected data serves us as a basis for a further consideration of the values of information as a resource, as understanding and knowledge, as relevance and as a product. Based on the latest study of the information environment and information behavior of researchers we propose a model of academic information ecologies and a model of an academic library. The framework of information ecologies helps us understand digital information infrastructures, develop a theory of information science and educate new information professionals.

## 2. Conceptual background and related studies

Information environment can be regarded as a complex system of information interactions which support the information process, namely the information lifecycle of creation, processing, communication and use of information (Steinerová, 2018a; Roos et al., 2008). It forms a framework of information processing and use. The related concept of information use environments (IUE) was introduced by Taylor and later re-conceptualized by Byström, Heinström and Ruthven (2019) in relation to the digital information. The phrase "digital environment" designates information interactions mediated by digital tools, digital resources and advanced information technologies in scholarly communication and information use. The increased prominence of the digital environment validates the concept of digital revolution and opens new avenues for the research on digital information (Floridi, 2014).

The information infrastructure consists of networks of people, objects, integrated sources, services and institutions, which include values, social interactions and knowledge (Bowker et al., 2015; Borgman, 2015). Information infrastructures enable information use. They can be described as a substrate upon which something else can operate. The concept of the ecology of information infrastructures implies a holistic perspective of interactions and adaptations of systems, tools, values, resources and people. Information infrastructures are connected to people in information practices, and information structures. Information infrastructures are transparent, invisible, fluid and interoperable. They are embedded in social contexts, and characterized by varying links with communities of practice, and spatial and temporal scope. In scholarly communication information infrastructures provide information, resources and services for researchers. Therefore, the library and information services of information institutions might be understood as a part of information infrastructures.

Digital scholarship can be explained as a transfer of scholarly communication into the digital environment. Open science emerged as a concept based on trends of open access movement and on access to digital resources and digital publishing, which changed scholarly communication and information use. The most important qualities of open science are digital access, transparency (of methodologies or procedures), and participation (in digital spaces). The emphasis on these qualities has made a significant impact on new models of scholarly communication and information services of academic libraries, and on work of information professionals.

At the conceptual level we will refer to the human information behavior (HIB) as an indicator of information needs of scholars, values of information and patterns of information use. Awareness of HIB can promote the designing of value-added information services, systems and products for researchers.

Human information behavior might be defined as a multilevel human activity related to information use. It is marked by relationships of people to sources and channels of information, and therefore it describes the different ways in which people need, manage, seek and use information (Fisher et al., 2005). HIB is composed of information-related needs, information behavior (searching, browsing, monitoring, seeking) and information use (effects of information) (Ford, 2015). Information practices are understood as a set of socially and culturally established ways to identify, seek, use and share information (Savolainen, 2008). HIB encompasses not only active information seeking, but also unintentional or passive behavior (Case & Given, 2016).

Empirical studies of information behavior of researchers found that scholars are gatekeepers who share information formally and informally. Many models of information behavior of scientists have been developed. One of the most influential models was developed by Ellis (2005) who identified the stages of information behavior of scholars, including starting, chaining, browsing, differentiating, monitoring, extracting, verifying, and ending. Another non-linear model of information behavior of scholars described the processes of opening, orientation, and consolidation, taking into account cognitive and external factors (Foster, 2004). Differences across various academic disciplines and modes of communication shape different patterns of information use (Talja, 2005; Brown, 2010; Fry, 2013). Recent studies point to new patterns of information behavior of researchers in digital environment, e.g. use of digital resources, data sharing, research data management (Tenopir et al., 2015) and use of social networks (Greifeneder et al., 2018). Xiang (2015) studied open science factors and found gaps in the use of open access sources. Harley et al. (2007) conducted interviews to study academic values which influence publishing, emphasizing the peer review process and weaker interest in electronic publishing. The authors proposed a deeper examination of scholars' needs and encouraged consideration of in-progress scholarly communication. Further studies presented limited use of social media (Bulger et al., 2011) and barriers between information needs of humanities scholars and information infrastructures (McGuiness, 2006). Based on interviews with 22 scholars, Scanlon (2014) found pragmatic patterns of information practices. The pragmatic approach emphasized the visibility and citations in the digital environment. Overall, the scholarship on information behavior of researchers shows that a scholar's discipline shapes his or her use of digital tools. For scholars of different disciplines to collaborate, common ground, readiness, management, planning and translations among disciplines are required (Given

& Willson, 2015; Olson & Olson, 2016; Palmer & Neumann, 2002). The awareness of these requirements gave rise to the concept of digital scholarship and big data in science emerged (Borgman, 2015; MacKenzie & Martin, 2016).

## 3. Open science and research information interactions

Open science might be defined as a relation of science to the public, while using of open access sources and tools, digital services, electronic resources and tools. Such an understanding of science is characterized by transparent scholarly communication in society and provision of access to results (Watson, 2015). Its main component is the digital environment which includes open data, open access, open methods and processes, open software, open review and open education. Digital science refers to the transfer of the research process into the digital environment, which includes new digital infrastructures, such as digital libraries, digital tools and services. Open science develops transparent information strategies, methods, and procedures with the emphasis on open access to publications and data. Several models of open science have been developed (e.g. Zuccala, 2009; Lyon, 2016; Open Science Framework, 2014), highlighting open access, transparency, participation, relations with public and promotion. Scholarly communication has been modelled by a number of authors (e.g. Björk, 2005; Hurd, 2000; Whitworth & Friedman, 2009). These models focused on rich research information interactions in digital environment and new actors (e.g. database providers, web services, digital libraries, aggregators, web editors). Research information interactions can be understood as multilayer relationships of researchers with the information environment. They are shaped by common factors determining information use of researchers, namely methodological background, domain expertise and open science factors (access, publicity) (Steinerová, 2016).

New value-added services for science are designed including sustainable digital information services (Chowdhury, 2014) and management of big research data (Borgman, 2015). For example, Open Science Framework (Centre for Open Science, 2015) have developed guidelines for support of open science, regarding citations, transparency of data, scientific documents, transparency of methods, design and analyses, registration of studies, plans, analyses, and replications of studies. Information professionals keep seeking new economic models of publishing and scholarly communication policies (Open Science, 2017).

One of the most important dimensions of open science is information infrastructure, composed of data and publications. There are many other complementary dimensions of open science, such as social, legislative, and technological aspects, summarized in several European initiatives. The open science models identified transparency, participation, communication, cooperation, and collaboration as contributing factors. Big research data management, electronic journals, digital tools and digital libraries, digital repositories, and social networks also shape the public's relation to science. Digital environment has changed the management of the research process. It gave rise to a need for new models of research assessment and need for new models of information and research ethics.

## 4. Qualitative studies of information behavior in the academic environment: a case of Slovakia

Our studies of information behavior of doctoral students, information managers and researchers demonstrate the effects of digital revolution and the challenges it poses to information infrastructures. Since 2005 we have run several qualitative studies of HIB with the use of phenomenography. All studies were conducted within the academic information environment as a part of larger projects on information use, information ecology and modelling of the information environment. We used methodologies of semi-structured interviews, qualitative analyses and an original methodology of concept mapping (Steinerová, 2018a). We also participated in an international study of research data literacy and data management (Steinerová & Ondrišová, 2018).

The first study focused on relevance assessment by doctoral students (2005–2007) (Steinerová et al., 2007). We conducted semi-structured interviews with 21 PhD students and a focus group discussion. Based on multiple qualitative analyses we represented results by many semantic models and several concept maps (perceptions of relevance, types of relevance, relevance in the electronic environment) (Steinerová, 2008; Steinerová, 2011). We re-defined relevance in the digital environment as socio-cognitive relationships supported by interactions, dynamics, construction of meaning. Several types of relevance were identified (10 faces of relevance) and we presented a final model of the collective discourse of relevance. Relevance of information was understood as importance, utility and value. In digital environment we explained three metaphoric faces of relevance as a construction, a pathway, and a pattern. In the digital environment the model of relevance 2.0 was marked by interactivity, dependence on contexts, participation, linking, multi-criterial access, mosaic-like construction of meaning, and non-linearity (Steinerová, 2011).

The second example comes from a study of the information ecology in the academic information environment based on semi-structured interviews with 17 information managers from Slovak universities (Steinerová et al., 2012). The project was realized in 2010–2011. Based on qualitative analyses, semantic representations and further experimental methods we developed a model of the information ecology in the academic information environment. The model showed three dimensions of the information ecology, namely the semantic, the behavioral and the visual dimension. The results had implications for academic libraries and universities, as well as for building digital spaces, digital libraries and repositories. The discourse favoured people, common values, creativity, communication and information sources and digital technologies as the main components of the information ecology. We also identified ecological dimensions of information literacy (Steinerová, 2010), ecological characteristics of digital libraries and information (Steinerová, 2014), information support of research information interactions and methodological literacy of doctoral students (Steinerová & Hrčková, 2014; Steinerová, 2013).

The third example comes from our latest project focused on modelling of digital scholarship (2015–2017). We conducted semi-structured interviews with 19 expert researchers, applied qualitative analyses and conceptual mapping. We developed 23 concept maps and final semantic models which identified common research information interactions and differences among disciplines (Steinerová, 2018b). The results were published in a monograph on information environment and scholarly communication (Steinerová, 2018a).

A part of the study was devoted to the attitudes of researchers to open science and digital publishing, economic models of science, and values of research. In the final chapters we proposed a model of academic information ecologies and an interactive model of the academic library.

## 5. Examples of concept maps

With regard to open access, open science and digital publishing, we asked researchers the questions: *Do you know the principles of OA sources, open science, do you use OA journals in your disciplines (electronic journals, data archives)?* The discourse, composed of all answers, was divided into the elements of support and the critique. In the supportive discourse benefits of open science were identified, such as increase of citations, speed of publishing, promotion of results to broader public, participation, transparency, open access, collaboration, peer networking, information sharing. Researches in STEM fields emphasized advanced technologies, while humanities scholars noted their need for building digital libraries and systems for cultural heritage (e.g. archival memory system of memorable sites, PamMap, atlas of Slavic languages, archaeological digital collections, Maya culture digital sources). In the critical discourse researchers expressed concerns regarding commercial influences and assessment of digital publications. The discourse of perceptions of open access and digital publishing is visualized in the concept map (Fig. 1).

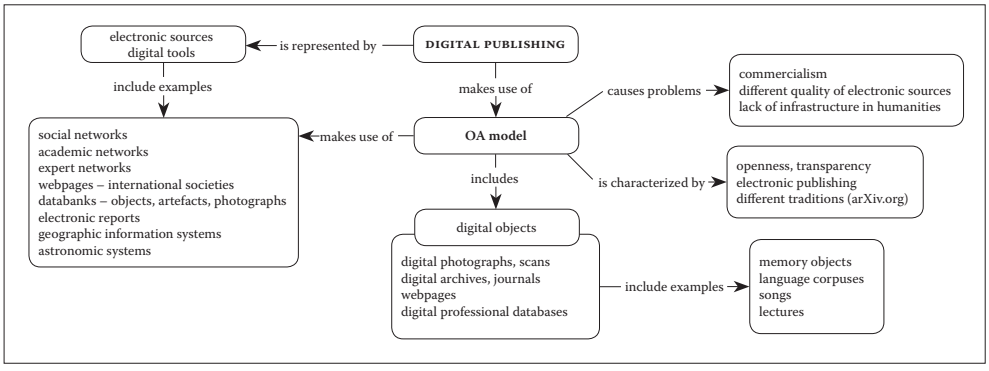


Fig. 1. Perceptions of digital publishing and the OA model (Steinerová, 2018a)

Researchers also strongly praised information policies, access to data and evaluation of results. Less important factors included social media used by researchers, information sharing and alternative metrics. Some scholars expressed agreement with the European trends and policies (Open Science, 2017), whereas others were concerned with lower quality of digital publishing. Several gaps in awareness and use of open access resources were identified. One respondent said:

> It is too early for a final conclusion, experiences are mixed. OA brought the invasion of journals...it is visible in the offer...speed of publishing ... (R6).

Researchers were particularly concerned with the excessive emphasis on quantity with the inappropriate evaluation of publications, and with the position of small disciplines and small countries. Benefits of OA sources were interpreted in context of interdisciplinary cooperation.

> OA has also a philosophical, conceptual problem, it is not only finances.... From public sources we support the private companies...I support the green model of OA in order to get rid of the commercial barriers (R5).

The respondents also discussed relations with public and information and research policies.

Another set of questions focused on the perception of the economic models of scholarly communication and open science: *What do you think about economic models of scholarly communication in your discipline (paid services versus OA movement, open scholarly communication versus publishing and information industry)? What is your opinion on the open science?* The discourse of researchers was mainly critical, but also constructive. Researchers used metaphors of academic market and academic culture in order to describe the co-existence of different models of scholarly communication. Some differences among policies of grant agencies, basic and applied research were considered. The volume of finances for science and research was deemed to be very low:

> We need finances for that infrastructure (R19).

Other respondents accepted some competition among projects and relationships with industry:

> ...we need to sell ourselves and persuade, I find it right and natural (R18).

Some participants recommended combined models of commercial and non-commercial scholarly communication. With regard to open science, researchers pointed to its conceptual problem in mixing the private and public sources and its elimination of differences among disciplines. Perceptions of OA sources and publishing were ambivalent, as concerns were expressed with regard to payments for publishing and predatory journals. Peer review process is regarded as the best way for the assessment of quality of publications. The open space of fast and commercial publishing can represent a threat to academic values. New partnerships with industry can be beneficial for interdisciplinary subjects. Some respondents emphasized the importance of freedom and independence of research. Generally, researchers agreed that academic information culture is a complex system with rich research information interactions and relationships among research, education, industry and publishing. The discourse of the economic models of science and open science is represented on Figure 2.

With regard to values of research we asked the question: *Which values are the most important for research work for you in your discipline?* The discourse of researchers was divided into the individual values and the social, collective values, while the most appreciated value was contribution to knowledge. The individual values were based on characteristics of the creative personality of a researcher, his or her motivation, interest, curiosity, aspiration to discovery of new perspectives and intellectual accomplishment:

> If I enjoy something and it is a challenge for me, I want to achieve something (R6).

The social values mentioned by the respondents were the advancement of knowledge, helping people, basic understanding of life, discovery of culture, service to society, and education of young scholars:

> The possibility to be ahead in knowledge and to invent something which has not been invented by anyone else before... (R16).

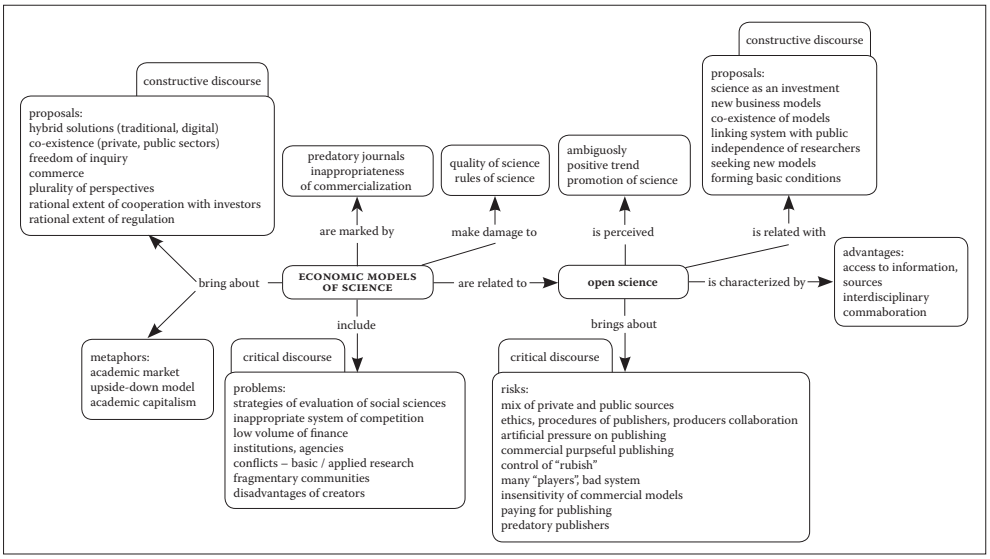


Fig. 2. Economic models of science and open science (Steinerová, 2018a)

The value of research is related to the position of science in society. Our respondents confirmed that there is some contrast between academic and commercial values. The analysis of the discourse is illustrated by the concept map on the Figure 3.

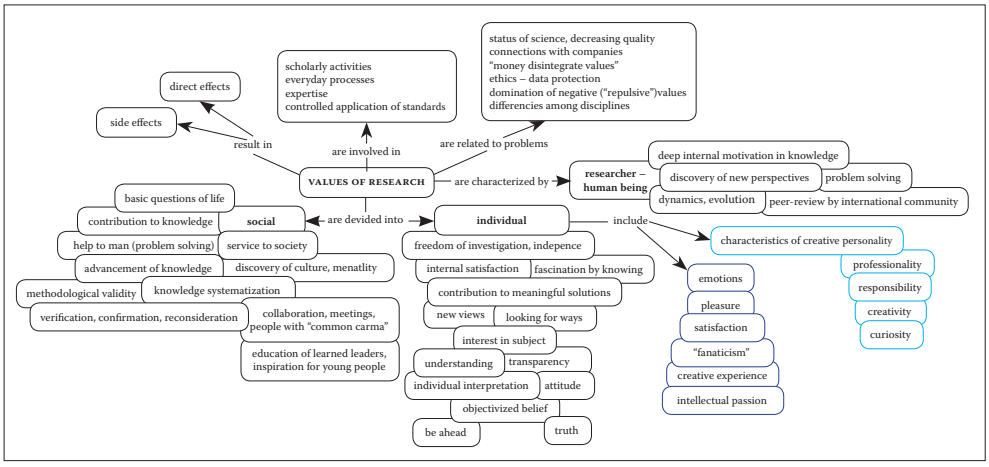


Fig. 3. Values of research (Steinerová, 2018a)

In further analyses we found that researchers were critical of the system of assessing the research results. Many of them called for reliance on basic academic values preserved in traditions of communication in academic communities:

> ...simply that value...also with the people with whom I collaborate, they are doing it for that value which it really has, not for that inappropriate counting in our system... (R18).

We found that the perceived value of information depends on the more general value attributed to research. Value of information might also be interpreted as worth, utility and desirability (Norton, 2010). Furthermore, values of information are integrated in the identity of information resources, relevance, information products, understanding and knowledge, and in information as social power.

The three concept maps represent the discourse based on the interviews with researchers. The results demonstrate the complexity of the information environment of scholarly communication and the need for further research on values attributed to information. The interpretation of values, and their transformation in the digital environment should be studied further. We need to explore the issues of building trust, truth, verification of resources, and freedom of expression in order to develop value-added services of academic libraries (Kelly & Bielby, 2016; Floridi, 2014).

## 6. A model of academic information ecologies

Based on our analyses and on the studies of information ecology, relevance and information behavior we developed a model of the academic information ecologies. It is a framework for understanding the information environment (Fig. 4).

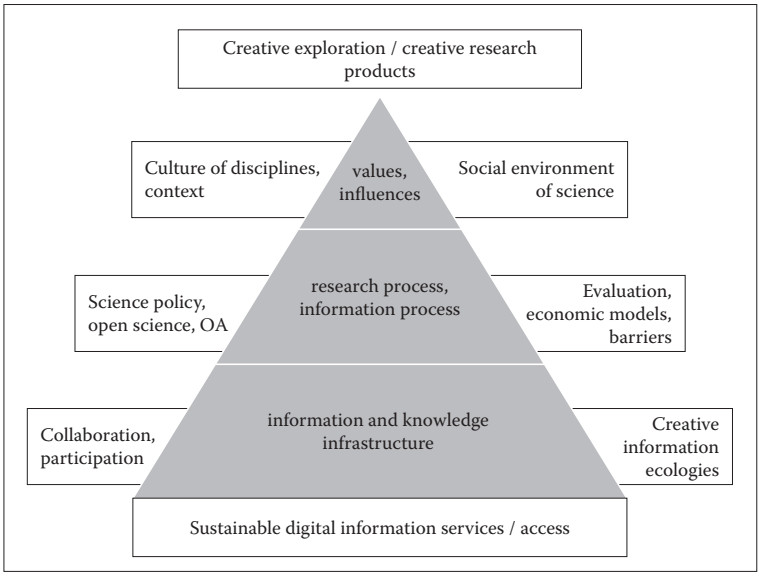


Fig. 4. A mode of academic information ecologies (Steinerová, 2018a)

This model illustrates the basic infrastructure of digital information access and the following strata of the research process, values and influences of academic cultures and social environment. The model is aimed at a better understanding of the information environment as it emphasizes ecological features of adaptations, interactions, socio-technological evolution and information re-use. Information ecologies are complex information spaces, places, interactions. The model can be applied to information and research policies and further research of values of information and creative exploration. It can also be used for design of value-added services of academic libraries.

## 7. An interactive model of the academic library

Following the results of our studies we designed a model of an interactive academic library. We used proposals of the constructive discourse of researchers with regard to the improvement of library services and information infrastructure. The model is depicted on the Figure 5.

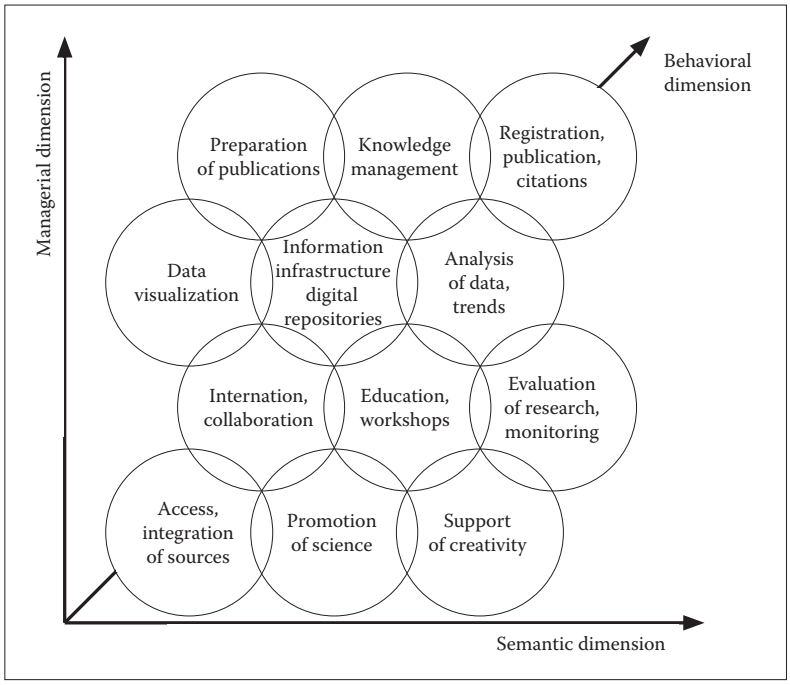


Fig. 5 An interactive model of the academic library (Steinerová, 2018a)

The model is designed as an interactive space based on interactions and dialogue of academic communities. The space is composed of the semantic dimension, the managerial dimension and the behavioral dimension. The inner intertwined circles represent the required value-added services in the three dimensions. In the model we can discover several pathways of transformation of data to knowledge, new ways of information interactions,

the shift from access to interactive and collaborative information behavior and support of research creativity in digital spaces. This model can help build digital services for academic communities, which would facilitate social networking, open science, project management, publishing strategies and digital ethics. Its features might include interactive digital repositories, research data management and multiple uses of data and information. The perceived value of information and information ethics should be taken into consideration when developing integrative value-added services of academic libraries. The model includes researchers´ concerns regarding the issues of data protection, privacy, and intellectual property. The academic library should play a crucial role in scholarly collaboration, networking and innovative teams. Researchers are also interested in support of research data management and digital publishing. Value-added services and value-sensitive design of digital libraries and spaces for researchers should start with an awareness of values of research and values of information for researchers.

## 8. Conclusions

We have reviewed challenges of information infrastructures for open science and academic libraries based on examples of our studies of relevance, information ecology and information behavior of researchers, doctoral students and information managers. Related studies presented the challenges posed by digital environment to development of new, value-added services of information infrastructures and academic libraries. We proposed a theoretical and methodological framework of academic information ecologies and a new model of an interactive academic library.

Information ecology can be regarded as a common background of our models and studies. We emphasize adaptations to changes of the information environment, sociotechnological evolution, re-use of digital information and value-added services focused on creativity, interactivity, social media, information literacy and information ethics. In the digital environment, researchers and information science should pay more attention to concepts of digital literacy and digital ethics. The role of information science will be crucial for understanding the changing digital information environment, as well as risks of digital information and changing services of information infrastructures.

Information science should respond to the challenges of information infrastructures and digital revolution with an enhanced theoretical understanding of scholars' changing pathways in the digital environment and with proposals of new value-added services. Therefore, information research should emphasize values of information and values of research. Emergent services require that new information professionals (e.g. digital librarians and data curators) should be educated in and focused on digital library services, research data management, digital publishing and data curation. Gaps in information infrastructures need to be filled with funded projects and value-added information services supporting scholarly communication in digital environment.

### Acknowledgements

This paper was developed as part of the research project APVV 15-0508 HIBER.

## References


Björk, B. Ch. (2005). A Life-Cycle Model of the Scientific Communication Process. *Learned Publishing*, 18 (3), 165–176.

Borgman, Ch. L. (2015). *Big Data, Little Data, No Data. Scholarship in the Networked World*. Cambridge: MIT Press.

Byström, K., Heinström, J., Ruthven, I., eds. (2019). *Information at Work. Information Management in the Workplace*. London: Facet.

Bourg, C., Coleman, R., Erway, R. (2009). *Support for the Research Process: An Academic Library Manifesto*. Report produced by OCLC Research [online]. OCLC Research [15.07.2019], www.oclc.org/research/publications/library/2009/2009-07.pdf

Bowker, G., Timmermans, S., Clarke, A.E., Balka, E., eds. (2015). *Boundary Objects and Beyond. Working with Leigh Star*. Cambridge, MA: MIT Press.

Brown, C. (2010). Communication in the Sciences. *Annual Review of Information Science and Technology*, 44, 287–316.

Bulger, M. E., Meyer, E. T., De la Flor, G., Terras, M., Wyatt, S., Jirotka, M., Eccles, K., Madsen, C. (2011). *Reinventing Research? Information Practices in the Humanities. A Research Information Network Report*, April 2011 [online]. SSRN. DOI: http://dx.doi.org/10.2139/ssrn.1859267

Case, D.O., Given, L. (2016). *Looking for Information: A Survey of Research on Information Seeking, Needs and Behavior* (4th ed.). Bingley, UK: Emerald.

Center for Open Science. (2016). [online]. Open Science Framework [15.07.2019], https://cos.io/our-products/open-science-framework/

Chowdhury, G. G. (2014). *Sustainability of Scholarly Information*. London: Facet.

Ellis, D. (2005). Ellis´s Model of Information-Seeking Behavior. In: K. E. Fisher, S. Erdelez, L. McKechnie (eds.). *Theories of information behavior* (138–142). Medford: Information Today.

Erdelez, S., Means, T. (2005). Measuring Changes in Information Sharing among Life Science Researchers. In: S. Al-Hawamdeh (ed.). *Knowledge Management: Nurturing Culture, Innovation and Technology* (29–40). New Jersey: World Scientific.

Edwards, P. N., Jackson, S. J., Bowker, G. C., Knobel, C. (2007). *Understanding Infrastructure: Dynamics, Tensions and Design. Report of a Workshop on History and Theory of Infrastructure: Lessons for New Scientific Cyberinfrastructures* [online]. Washington: NSF [09.08.2008], http://deepblue.lib.umich.edu/bitstream/2027.42/49353/3/UnderstandingInfrastructure2007.pdf

Fidel, R. (2012). *Human Information Interaction: An Ecological Approach to Information Behavior*. Cambridge, Mass: MIT Press.

Fisher, K.E., Erdelez, S., McKechnie, L., eds. (2005). *Theories of Information Behavior*. Medford, MA: Information Today.

Floridi, L. (2014). *The Fourth Revolution. How the Infosphere is Reshaping Human Reality*. Oxford: Oxford University Press.

Foster, A. (2004). A Nonlinear Model of Information-Seeking Behaviour. *Journal of the American Society for Information Science and Technology*, 55 (3), 228–237, DOI: https://doi.org/10.1002/asi.10359

Ford, N. (2015). *Introduction to Information Behaviour*. London: Facet.

Fry, J. (2013). Considerations in Adopting a Disciplinary Analysis of Scholarly Communication and Information Behaviours. In: I. Huvila (ed.). *ASIST European Workshop* (63–78). Abo: Abo University.

Given L.M., Willson R. (2015). Collaboration, Information Seeking, and Technology Use: A Critical Examination of Humanities Scholars' Research Practices [online]. In: P. Hansen, C. Shah, C.P. Klas (eds.). *Collaborative Information Seeking. Computer Supported Cooperative Work* (139–164). Cham, Switzerland: Springer. DOI:10.1007/978-3-319-18988-8_8

Greifeneder, E. (2014). Trends in Information Behaviour Research. *Information Research* [online], 19 (4) [16.07.2019], http://www.informationr.net/ir/19-4/isic/isic13.html

Greifeneder, E., Pontis, S., Blandford, A., Attalla, H., Neal, D., Schlebbe, K. (2018). Researchers' Attitudes Towards the Use of Social Networking Sites. *Journal of Documentation*, 74 (1), 119–136. DOI: https://doi.org/10.1108/JD-04-2017-0051

Harley, D., Earl-Novell, S., Arter, J., Lawrence, S., King, C.J. (2007). The Influence of Academic Values on Scholarly Publication and Communication Practices. *Journal of Electronic Publishing*, 10 (2), DOI: http://dx.doi.org/10.3998/3336451.0010.204

Hurd, J. (2000). The Transformation of Scientific Communication: A Model for 2020. *Journal of the American Society for Information Science and Technology*, 51 (14), 1279–1283.

Kelly, M., Bielby, J., eds. (2016). *Information Cultures in the Digital Age*. A Festschrift in Honor of Rafael Capurro. Wiesbaden: Springer VS.

Lyon, E. (2016). Transparency: The Emerging Third Dimension of Open Science and Open Data. LIBER Quarterly, 25 (4), 153–171. DOI: http://doi.org/10.18352/lq.10113

Kirchner, J., Diaz, J., Henry, G., Fliss, S., Culshaw, J., Gendron, H., Cawthorne, J. (2015). *The Centre of Excellence Model for Information Services* [online]. Washington: CLIR. CLIR [16.07.2019], http://www.clir.org/pubs/reports/pub163

MacKenzie, A. & Martin, L., eds. (2016). *Developing Digital Scholarship. Emerging Practices in Academic Libraries*. London: Facet.

Malpas, C., Proffitt. M. (2017). *The Transformation of Academic Library Collecting: A Synthesis of the Harvard Library's Hazen Memorial Symposium* [online]. Dublin, OH: OCLC Research. DOI:10.25333/C3J04Z.

McGuiness, C. (2006). What Faculty Think-Exploring the Barriers to Information Literacy Development in Undergraduate Education. *Journal of Academic Librarianship*, 32 (6), 573–582.

Norton, M. (2010). *Introductory Concepts in Information Science*. 2nd ed. Medford: Information Today.

Olson, G.M., Olson, J. (2016). Converging on Theory from Four Sides. In: D. Sonnenwald (ed.). *Theory Development in the Information Sciences* (87–100). Austin: Univ. of Texas.

Open Science Research and Innovation. (2017). European Commission. Open Science Policy Platform. [online]. [16.07.2019], https://ec.europa.eu/research/openscience/index.cfm?pg=open-science--policy-platform

Open Science and Research Initiative. (2014). *The Open Science and Research Handbook* [online]. [23.03.2017], http://openscience.fi

Palmer, C. L., Teffeau, L.C., Pirmann, C. M. (2009). Scholarly Information Practices in the Online Environment: Themes from the Literature and Implications for Library Service Development [online]. OCLC Research [01.12.2017], http:// www.oclc.org/programs/publications/reports/2009-02.pdf

Palmer, C.L., Neumann, L.K. (2002). The Information Work of Interdisciplinary Humanities Scholars: Exploration and Translation. *The Library Quarterly*, 72 (1), 85–117. DOI: https://doi.org/10.1086/603337

Roos, A., Kumpulainen, S., Järvelin, K., Hedlund, T. (2008). The Information Environment of Researchers in Molecular Medicine. *Information Research* [online], 13 (3), paper 353 [17.07.2019], http://InformationR.net/ir/13-3/paper353.html

Researcher Development Framework (2011). [online]. Vitae Researcher Development Framework. Vitae. Realizing the Potential of Researchers. Vitae, Careers Research and Advisory Center. 2019. Research Council UK [17.07.2019], https://www.vitae.ac.uk/vitae-publications/rdf-related/researcher-development-framework-rdf-vitae.pdf/view

Savolainen, R. (2008). *Everyday Information Practices. A Social Phenomenological Perspective*. Lanham: Scarecrow.

Scanlon, E. (2014). Scholarship in the Digital Age: Open Educational Resources, Publication and Public Engagement. *British Journal of Educational Technology*, 45(1), 12–23.

Steinerová, J. (2008). Seeking Relevance in the Academic Information Use. *Information Research* [online], 13 (4), 1–11. [01.12.2017], http://InformationR.net/ir/13-4/paper380.html

Steinerová, J. (2010). Ecological Dimensions of Information Literacy. *Information Research* [online], 15 (4) [17.07.2019], http://InformationR.net/ir/15-4/colis719.html

Steinerová, J. (2011). Premeny relevancie v informačnej vede a informačná ekológia. *Knihovna: knihovnická revue* [online]. Praha: Národní knihovna ČR, 22 (2), 59–70 [17.07.2019], http://oldknihovna.nkp.cz/knihovna112/11259.htm

Steinerová, J. (2012). Information Ecology – Emerging Framework for Digital Scholarship. LIDA 2012 [online], 12 [17.07.2019], http://ozk.unizd.hr/proceedings/index.php/index/index, http://ozk.unizd.hr/proceedings/index.php/lida2012/article/view/66/37

Steinerová, J., Ilavská, J., Ondrišová, M., Lichnerová, L., Ondriašová, H., Haršányiová, M., Prágerová, L., eds. (2012). *Informačná ekológia akademického informačného prostredia. Záverečná správa z výskumu VEGA 1/0429/10. (Information Ecology of the Academic Information Environment. Final Report)*. Bratislava: Univerzita Komenského.

Steinerová, J. (2013). Methodological Literacy of Doctoral Students – An Emerging Model. In: S. Kurbanoglu, E. Grassian, D. Mizrachi, R. Catts, S. Špiranec (eds.). *Worldwide Commonalities and Challenges in Information Literacy Research and Practice* (148–154). Heilderberg: Springer.

Steinerová, J. (2014). Ecological Information Interactions for Digital Scholarship. In: P. Návrat, M. Bieliková, M. Laclavík, J, Paralič, J. Steinerová (eds.). *Cognitive Traveling in Digital Space of the Web and Digital Libraries. Interdisciplinary Multi-Partner Project TraDiCe* (158–169). Bratislava: STU.

Steinerová, J. (2016). Open Science and the Research Information Literacy Framework. In: S. Kubanoglu et al. (eds). *Information Literacy: Key to an Inclusive Society* (277–285). Cham: Springer, CCIS 676. DOI:10.1007/978-3-319-52162-6_27

Steinerová, J. (2018a). *Informačné prostredie a vedecká komunikácia: informačné ekológie*. Bratislava: UK.

Steinerová, J. (2018b). Perceptions of the Information Environment by Researchers: A Qualitative Study. *Information Research* [online], 23(4), paper isic1812. Proceedings of ISIC, The Information Behaviour Conference, Krakow, Poland, 9–11 October 2018: Part 1 [17.07.2019], http://www.informationr.net/ir/23-4/isic2018/isic1812.html

Steinerová, J.(2018c). Informačné prostredie a vedecká komunikácia: aspekty vedeckej dátovej gramotnosti. *ProInFlow* [online], 10 (2), [17.07.2019], http://www.phil.muni.cz/journals/index.php/proinflow/article/view/2018-2-2, https://doi.org/10.5817/ProIn2018-2-2.

Steinerová J. (2019). Vedecká dátová gramotnosť: konceptuálne východiská a situácia v SR. In: *Knižničná a informačná veda*. 28. Zborník FiFUK. Bratislava: UK, (in print).

Steinerová, J., Grešková, M., Šušol, J. (2007). *Prieskum relevancie informácií. Výsledky analýz rozhovorov s doktorandmi FiF UK. (Survey of Relevance of Information. Results of Analyses of Interviews with Doctoral Students in Faculty fo Arts)*. Bratislava: Centrum VTI SR.

Steinerová, J., Grešková, M., Ilavská, J. (2010). *Informačné stratégie v elektronickom prostredí. (Information Strategies in the Electronic Environment)*. Bratislava: Univerzita Komenského v Bratislave.

Steinerová, J., Hrčková, A. (2014). Information Support of Research Information Interactions of PhD. Students in Slovakia. *The Grey Journal*, 10 (2), 79–85.

Steinerová, J., Ondrišová, M. (2019). Research Data Literacy Perception and Practices in the Information Environment [online]. In: S. Kurbanoglu, S. Špiranec, Y. Ünal, J. Boustany, M.-L. Huotari, E. Grassian, D. Mizrachi, L. Roy (eds.). *Information Literacy in Everyday Life*. ECIL 2018. 6th European Conference, ECIL 2018. Revised selected papers. Oulu, Sept., 24–27, 2018. Cham: Springer Nature Switzerland AG, CCIS 989. DOI: 10.1007/978-3-030-13472-3_51

Talja, S. (2005). The Domain-Analytic Approach to Scholars´ Information Practices. In: K. Fisher, S. Erdelez, L. McKechnie (eds.). *Theories of Information Behavior* (123–127). Medford: ASIST-Information Today.

Tenopir, C., King, D.W., Christian, L., Volentine, R. (2015). Scholarly Article Seeking, Reading, and Use: A Continuing Evolution from Print to Electronic in the Sciences and Social Sciences, *Learned Publishing*, 28 (2), 93–105.

Ünal, Y., Chowdhury, G., Kurbanoğlu, K., Boustany J., Walton G. (2019). Research Data Management and Data Sharing Behaviour of University Researchers. In: *Proceedings of ISIC, The Information Behaviour Conference, Krakow, Poland, 9–11 October 2018: Part 2. Information Research* [online], 24(1), paper isic1818 [17.07.2019], http://InformationR.net/ir/24-1/isic2018/isic1818.html. WebCite®: http://www.webcitation.org/74U6bBMqk)

Watson, M. M. (2015). When Will "Open Science" Become Simply "Science"? *Genome Biology* [online], 16, 1[17.07.2019], https://genomebiology.biomedcentral.com/articles/10.1186/s13059-015-0669-2

Whitworth, B., Friedman, R. (2009). Reinventing Academic Publishing Online Part II: A Sociotechnical Vision. *First Monday* [online], 14 (9) [17.07.2019], http://firstmonday.org/ojs/index.php/fm/article/view/2642, DOI: https://doi.org/10.5210/fm.v14i9.2642

Wilson, T. D. (2016). A General Theory of Human Information Behavior. In: *Proceedings of ISIC, the Information Behaviour Conference, Zadar, Croatia, 20–23 September, 2016: Part 1. Information Research* [online], 21(4), paper isic1601 [17.07.2019], http://InformationR.net/ir/21-4/isic/isic1601.html

Wilson, T. D., ed. (2013). *Theory in Information Behaviour Research* (e-book). Eiconics Ltd.

Xiang, R. (2015). The Quandary Between Communication and Certification. *Online Information Review* [online], 39 (5), 682–697. DOI: https://doi.org/10.1108/OIR-04-2015-0129

Zuccala, A, A. (2009). The Layperson and Open Access. *Annual Review of Information Science and Technology*, 43, 359–396.

# Wyzwania dla infrastruktur informacyjnych otwartej nauki i bibliotek akademickich

**Abstrakt**

**Cel/Teza:** W artykule dokonano przeglądu głównych pojęć dotyczących infrastruktury informacyjnej, środowiska informacyjnego i zachowania informacyjnego badaczy w kontekście rewolucji cyfrowej. Rozważana jest koncepcja otwartej nauki i nowe modele komunikacji naukowej.

**Koncepcja/Metody badań:** Analizie poddano powiązane ze sobą badania infrastruktury informacyjnej i zachowań informacyjnych naukowców. Przedstawiono wyniki serii badań jakościowych dotyczących zachowań informacyjnych i ekologii informacji w środowisku akademickim na Słowacji. Badania te oparte były na wywiadach z doktorantami, badaczami i menedżerami informacji. Wykorzystano opracowaną wcześniej autorską metodologię mapowania pojęć.

**Wyniki i wnioski:** Uzyskane wyniki pozwalają ponownie rozważyć ocenę relewancji w środowisku cyfrowym i przedstawić model ekologii informacji w środowisku informacyjnym. Analizowane dane przedstawiono za pomocą map pojęć, które pokazują postawy badaczy wobec otwartej nauki, ekonomicznych modeli nauki i wartości badaczy. Objaśniono utworzony na podstawie badań autorski model akademickiej ekologii informacji i przedstawiono nowy interaktywny model biblioteki akademickiej.

**Zastosowania praktyczne:** Sugerujemy zastosowania dla rozwoju nauki o informacji oraz edukacji specjalistów informacji.

**Oryginalność/Wartość poznawcza:** Przedstawione autorskie modele podkreślają wartości badań akademickich, wartości informacji, usługi o wartości dodanej w bibliotekach akademickich oraz projektowanie bibliotek cyfrowych uwzględniające wartości.

**Słowa kluczowe**

Biblioteki akademickie. Infrastruktura informacyjna. Otwarta nauka. Środowisko informacyjne. Zachowania informacyjne badaczy.

*Professor JELA STEINEROVÁ is a professor of Library and Information Science at the Comenius University in Bratislava, Faculty of Arts, and the head of the Department of Library and Information Science. In her research and teaching she focuses on the topics of human information behavior, theory and methodology of information science, relevance, information ecology and information products. She has published in home and foreign journals, presented at international conferences. She is a member of boards of international journals and conference committees and projects (e.g. ECIL, ISIC, CoLIS, ENWI). Shas has published several monographs, chapters in books and papers in journals. Her latest monograph is* Information Environment and Scholarly Communication: Information Ecologies *(2018). She organizes annual international conferences "Information Interactions" in Bratislava.*

*Contact to the Author:*
*jela.steinerova@uniba.sk*
*Katedra knižničnej a informačnej vedy*
*Filozoficka fakulta Univerzity Komenskeho v Bratislave*
*Gondova 2, 814 99 Bratislava, Slovakia*

# Data Curation in Academic Libraries as Part of the Digital Revolution[1]

Tibor Koltay
ORCID 0000-0003-0548-7527
*Institute of Learning Technologies, Eszterházy Károly University, Hungary*

**Abstract**

**Purpose/Thesis:** This paper outlines the role of data curation in the context of Research 2.0 and Research Data Management.
**Approach/Methods:** The argument is based on a non-exhaustive review of the literature.
**Results and conclusions:** Despite the relative vagueness and variety of definitions of data curation, academic libraries should engage in it.
**Research limitations:** The study focused mainly on theoretical writings.
**Practical implications:** The worldwide challenge associated with Research Data Management and data curation. Several countries and institutions have already answered the challenge, but the overall level of its recognition is low, and thus there is a need to raise awareness of its importance.
**Originality/Value:** The premise of the argument is based on the assumption that views on data are changing.

**Keywords:**

Data curation. Digital curation. Research 2.0. Research data management.

*Received:26 May 2019. Reviewed: 14 June 2019. Accepted: 15 July 2019.*

## 1. Introduction

As asserted by Michael Buckland,

> the potentially useful record of science comes in the form of (mainly non-textual) digital data sets, but the infrastructure behind it is not yet comparable to the one of print-on-paper materials and their digital versions (Buckland, 2011, 35).

The expression *Research 2.0* indicates this growing importance of data, especially in regard to research data that requires new approaches to scholarly research that influences the natural sciences, the social sciences and the humanities. It is clearly associated with the digital revolution, as without easy access to a wealth of information and data, enabled by a well-developed information infrastructure, openness of science would not be possible. Neither would have new disciplines, such as the digital humanities, been born.

If we agree that Research 2.0 is characterized by the recognition that sharing and re-using research data is inevitable, we also agree with the views, promulgated by the proponents of

[1] Preparation of this text was supported by the EFOP-3.6.1-16-2016-00001 project "Complex Development of Research Capacities and Services at Eszterházy Károly University".

information ecology, who are committed to providing a new framework for interpretation of relationships between social actors, information technologies, and information objects (Steinerová, 2012).

This review paper is in many regards a continuation of this author's earlier paper in this journal (Koltay, 2017). It focuses on data curation, which is part of Research Data Management (RDM) that needs to be cared for by researchers, as well as by academic and special libraries across the world In order to contribute to a better understanding of the nature of data curation, the conceptual analysis covers the effect of new views on data, some features of the changing academic library environment and varied approaches to data curation itself.

## 2. The environment

Today, we can clearly see an increase in the attention towards data literacy, which shares several features with other literacies and is especially closely connected to information literacy (Koltay, 2015). The nature of this interest is influenced by the changing ways of seeing data, compared to information. As a result of this and owing to the importance of research data, never seen before, working climate of academic library services has begun to change.

### *2.1. New views on data*

Describing the aforementioned development would be impossible without contemplating the changes in the ways of seeing data. Library and Information Science (LIS) has traditionally perceived data as somewhat secondary (or even inferior) to information, because it was regarded as occupying the bottom of the data–information–knowledge–wisdom pyramid (Rowley, 2007). However, the growing importance of data may change this view (Cox, 2018), inducing views that are far from seeing the relationship between these fundamental concepts as simple as it has been presented by earlier views (Makani, 2015), especially as data is not only a concept that appears with much higher frequency than ever before in the history of LIS but is "heavily loaded with epistemological problems" (Hjørland, 2019, 686). Nonetheless, some definitions clearly emphasize the closeness of data and information. For instance, Liangzhi (2015) affirms that from the ontological point of view, data and information are close to each other as both exist as signs. Stating that data is anything recordable in a semantically and pragmatically sound way, Frické (2008) indicates the potential of comprehensive approaches to defining data.

As this paper focuses on academic libraries' role mission, it is necessary to address the more specific idea of research data, i.e. data collected as part of a research project that might be contrasted to data collected or digitized for curation and preservation, without forgetting that data collected for similar reasons can become research data (Robinson, 2016). Research data can be seen not just as the result of empirical work or the raw material for statistical analysis, but also as a research object in its own right (Golub & Hansson, 2017). Furthermore, in the digital humanities there seems to be an implicit agreement that data can be interpreted as texts, and – conversely – that texts can be interpreted as data (Koltay, 2016).

Even new and utterly simple views on data can find their legitimate use and appropriate place. For instance, the Digital Curation Centre (2018) defines data as "any information in

binary digital form". To this definition, we can add that documents not born digital may become digital at some point (Robinson, 2016). If we take another, principally practice-oriented view, we can say that data constitutes a primary intellectual asset that can be subjected to peer review and other forms of quality assessment, and is open to reuse (Heidorn, 2011).

Approaches that conceive data differently are also influenced by a new definition of information literacy (IL), which states that IL is related not only to print, but to data, images and the spoken word (CILIP, 2018). This definition recognizes the abovementioned close relationship between data and information, and indicates information literacy's overlap with other literacies.

### *2.2. The changing academic library environment*

There is a host of varied and often synonymous or vaguely defined data-related terms, such as research data management, data curation, preservation, and data stewardship, used in library environments. Having chosen one of these in its policy document, the European Open Science Cloud sets the requirement for European science to be

> grounded in a common culture of data stewardship, so that research data is recognized as a significant output of research and is appropriately curated throughout and after the period conducting the research (EOSC, 2017,1).

This statement reminds us of the fact that – differing form the case of Open Access to journal articles – the worldwide challenge to give the appropriate attention to research data is far from being properly recognized in some countries, and that its recognition has not been evenly spread yet. However, when acknowledged, this challenge is often answered by academic libraries in many countries by providing a wide array of RDM services. Notwithstanding, there is still a need to raise awareness of its importance.

Another issue that we should not forget about is that our information environment is not only dominated by the importance of data, underlined in the above document. We must also see that the quantity of data available is overwhelming, and therefore we cannot be mistaken if accepting that – beside information overload – we witness the phenomenon of data overload or data deluge (Little, 2012). This means that we are overloaded by different kinds of data. Such a situation calls for efficient data curation and a high level of data literacy education (Robinson, 2016).

Data curation has been identified as both a challenge and an opportunity for

> finding new ways to communicate the value of the skills librarians already possess and in developing roles that were previously not associated with librarians (ACRL 2012, 312).

The issue becomes especially important if we consider that one of the academic librarians' roles is to support researchers (Mierzecka, 2019).

Data curation activities should be guided by the understanding that researchers do not need more data, but the right data (Borgman, 2015) and that data is of no use if not is not analyzed. Accordingly, librarians have to acquire analytical skills and abilities (Kirkwood, 2016). They also should set up appropriate and innovative new services, recognized by several stakeholders, including research funders, research organizations, the communities of researchers, as well as academic and special libraries. The latter two are crucial actors,

representing the human element in data curation, declared of uttermost importance by Poole (2015). Librarians' participation is of real value as libraries are integral part of the necessary institutional structure and only a few librarians are qualified to perform the related tasks (Griffin, 2013). Obviously, data management initiatives must acquire backing from the whole staff, but the role of the libraries' managements should not be underestimated either (Burton & Lyon, 2017).

## 3. Approaches to data curation

The place of data curation in academic libraries' workflow is not defined precisely. Nonetheless, it is one of the potential RDM activities, which ideally begin with data literacy instruction, followed by RDM itself, then by data curation and data preservation (Thomas & Urban, 2018). Digital curation can be seen as an umbrella term for data curation and digital preservation. Its main goals are preventing data loss, and adding value to trusted data assets for current and future use (Poole, 2016).

By subsuming digital preservation and supplying different kinds of documentation to provide context, as well as including metadata, data curation aims to make selected data accessible, usable, and useful throughout its lifecycle (Giarlo, 2013). As Burgi, Blumer, and Makhlouf-Shabou (2017) put it, the preservation of digital information is a complex and costly process that cannot be circumvented. Consequently, we can say – in general terms – that curation involves planning, acquiring, preparing, analyzing, preserving and discovering the data (Pouchard, 2016).

Curation in analogue or digital form is not only a domain of librarians, who manage libraries as organized collections within a certain place and space. Curators work in a variety of institutions, who care for a host of artefacts. Somewhat differently, archivists care for the permanency of records and their historical value. The definitions of digital curation and data curation converge as they emphasize the processes related to taking care of resources, including organization and description (Kouper, 2016). Therefore, if we regard data (as we did above) as any information in digital form (sic!), then data curation is identical with digital curation, i.e. they both consist of the management and preservation of digital material (Robinson, 2016).

While researchers inevitably need to curate their own data, most of them do not receive formal training in data management practices, and thus they are learning data management and curation on the job. Moreover they do it in an ad hoc fashion, because it is secondary to their main occupation, i.e. doing research (Jahnke et al., 2012). In a broader context, mandating open access to data may pressure researchers and research teams to provide data, even if intellectual property or data protection rights would specifically and explicitly limit their ability to do so, or it may aggravate ethical concerns about sharing research data in general (RECODE, 2014). Helping researchers to solve such dilemmas matches the librarians' skill portfolio that traditionally contains open access and copyright (Southall & Scutt, 2017).

Furthermore, data curation can be seen from different angles. If we emphasize the role of data repositories, data curation can be defined as the

> encompassing work and actions taken by curators of a data repository, in order to provide meaningful and enduring access to data (Johnston et al., 2018, 5).

As said above, data curation can also be seen as a part of RDM, which might be defined as a set of general activities, not specifically attached to the library, but potentially performed by it, involving organizing, collecting, describing, and storing data for communication among scholars. Accordingly, the emphasis of data curation is not on internal storage, but gives priority to extracting data for general purposes (public use) is stress, in which it is similar to curating and exhibiting collections in museums (Thomas & Urban, 2018). In general, knowledge and skills that may prove useful for data curation are among the ones that librarians are qualified to have and can be required to make use of (Burgi et al., 2017).

In an online survey of professionals, who identify themselves as digital curators Kouper (2016) found preservation as the most frequent theme. Slightly less frequently mentioned terms were access and dissemination, and the lifecycle approach. If we subscribe to the idea that digital curation and data curation are closely related to each other, we can suppose that the tasks of both types of curation are identical to a considerable degree.

While data curation, similarly to data literacy and RDM has to focus on data quality (Ridsdale et al., 2018), it goes beyond RDM, because it additionally comprises of preservation and enables reuse (Partlo et al., 2015).

In order to be successful, data curation requires management of digital data over its entire lifecycle, including the use of varied services, tools, and infrastructure (Schmidt & Shearer, 2016) to

> ensure the maintenance of authenticity, reliability, integrity and usability of digital material, which in turn ensures maximization of the investment in their creation (Higgins, 2008, 135.)

There is a close connection between technical aspects of RDM, such as preparing datasets for deposit into a repository and deaccessioning them from depositories, providing persistent storage, assign unique identifiers and data curation (Cox & Pinfield, 2014). However, the latter activities appear less frequently than informational (consultative) services, such as offering guidance on copyright, helping researchers or students in setting up Data Management Plans, or consulting on metadata standards (Tenopir et al., 2016; Tenopir et al., 2017).

An analysis of several data curation and digital curation curricula in the United States shows the presence of varied topics and broad scope of issues that include data types, standards, lifecycles, provenance, metadata, collection development, as well as digital preservation. Digital scholarship in the humanities also appears in the list of subjects (Harris-Pierce & Quan Liu, 2012). This reminds us of the decisive impact of data-intensive research on the digital humanities, mentioned at the beginning of this paper.

## 4. Conclusion

Data curation can be seen as a new label, conceived to name already existing activities that have been practiced in libraries, museums, and archives (Ray, 2012). In libraries, this can be observed if we understand that the roles of librarians in general and the duties of subject librarians in particular already entail activities designated as data curation (Bracke, 2011). No need to say that this is also true for data librarians. Still, it can be expected that data curation will become more than a label and will catch up with research data management

in meeting the demand of serving researchers, teaching staff members and students, as soon the need for such services arises.

Offering data-related services, including data curation, may prove useful for transcending the – otherwise respectable – role of the library in relieving researchers' technical and administrative burdens (LERU, 2013). In other words, data curation is not only one of the professional tasks for libraries on the to-do list, but it might serve to demonstrate that they play an essential role in research processes (Scaramozzino et al., 2012).

We should know that supporting researchers involves much more than data curation and RDM. Such activity might be, for instance Data Science Support, which is an emerging field of activities, provided by academic libraries (Oliver et al., 2019).

In order to keep the views and library practice related to research data up to date, librarians should constantly look out for new approaches and methods brought in to managing research data by data science, which – despite of its different origin and goals – may turn out to be applicable to data librarianship (Cox, 2018).

If we said that academic libraries' involvement in curating research data may be looked at and appreciated differently, it is worth to add that already in 2012 a research team at Colorado State University Libraries found that researchers require quality assistance and

> do not necessarily have preconceived perceptions that the libraries are poorly equipped to offer assistance (McLure et al., 2014, 157).

This shows that data-related activities may involve not only new tasks, but enhanced recognition of their work as well.

Even though it may sound strange, data curation-as-publishing is a promising new way for libraries to cooperate with digital humanists as it consists of curation not only in the sense of registering and making it public, but making it available and suitable for re-use (Muñoz, 2013).

Last, but not least, librarians should not forget that researchers often – and in many cases rightfully – may perceive themselves as owners of research data. Therefore librarians have to ensure them that libraries will be responsible stewards of their information, without taking ownership of this data, while original owners will be able to access their access content or take it out of library systems at any point and in-perpetuity (Lucky & Harkema, 2018).

## References

ACRL (2012). 2012 Top Ten Trends in Academic Libraries: A Review of the Trends and Issues Affecting Academic Libraries in Higher Education. *College & Research Libraries News*, 73(6), 311–320.

Borgman, Ch. L. (2015). *Big Data, Little Data, No Data: Scholarship in the Networked World*. Cambridge, MA: MIT Press.

Bracke, M. S. (2011). Emerging Data Curation Roles for Librarians: A Case Study of Agricultural Data. *Journal of Agricultural & Food Information*, 12(1), 65–74.

Buckland, M. (2011). Data Management as Bibliography. *Bulletin of the American Society for Information Science and Technology*, 37(6), 34–37.

Burgi, P. Y., Blumer, E., Makhlouf-Shabou, B. (2017). Research Data Management in Switzerland: National Efforts to Guarantee the Sustainability of Research Outputs. *IFLA Journal*, 43(1), 5–21.

Burton, M., Lyon, L. (2017). Data Science in Libraries. *Bulletin of the Association for Information Science and Technology*, 43(4), 33–35.

CILIP (2018). *CILIP Definition of Information Literacy 2018* [online]. CILIP. The Library and Information Association, [16.05.2019], https://infolit.org.uk/ILdefinitionCILIP2018.pdf
Cox, A. M. (2018). Academic Librarianship as a Data Profession: The Familiar and Unfamiliar in the Data Role Spectrum. *eLucidate*, 15(1/2), 7–10.
Cox, A. M., Pinfield, S. (2014). Research Data Management and Libraries: Current Activities and Future Priorities. *Journal of Librarianship and Information Science*, 46(4), 299–316.
DCC (2018). DCC Curation Lifecycle Model [online]. Digital Curation Centre, [16.05.2019] http://www.dcc.ac.uk/resources/curation-lifecycle-model
EOSC (2017). *EOSC Declaration* [online]. European Open Science Cloud, [16.05.2019], http://ec.europa.eu/research/openscience/pdf/eosc_declaration.pdf#view=fit&pagemode=none
Frické, M. (2008). The Knowledge Pyramid: a Critique of the DIKW Hierarchy. *Journal of Information Science*, 35(2), 131–142.
Giarlo, M. (2013). Academic Libraries as Quality Hubs. *Journal of Librarianship and Scholarly Communication*, 1(3): 1–10.
Golub, K., Hansson, J. (2017). (Big) Data in Library and Information Science: a Brief Overview of Some Important Problem Areas. *Journal of Universal Computer Science*, 23(11), 1098–1108.
Griffin, S. (2013). New Roles for Libraries in Supporting Data-Intensive Research and Advancing Scholarly Communication. *International Journal of Humanities and Arts Computing*, 7(Suppl.), 59–71.
Harris-Pierce, R. L., Quan Liu, Y. (2012). Is Data Curation Education at Library and Information Science Schools in North America Adequate? *New Library World*, 113(11/12), 598–613.
Heidorn, P. B. (2011). The Emerging Role of Libraries in Data Curation and e-Science. *Journal of Library Administration*, 51(7–8), 662–672.
Higgins, S. (2008). The DCC Curation Lifecycle Model. *International Journal of Digital Curation*, 3(1), 134–140.
Hjørland, B. (2019). Data (with Big Data and Database Semantics). *Knowledge Organization,* 45(8), 685–708.
Jahnke, L., Asher, A., Keralis, S. D. (2012). *The Problem of Data*. Washington, DC: Council on Library and Information Resources.
Johnston, L. R., Carlson, J., Hudson-Vitale, C., Imker, H., Kozlowski, W., Olendorf, R., Stewart, C. (2018). How Important Are Data Curation Activities to Researchers? Gaps and Opportunities for Academic Libraries. *Journal of Librarianship and Scholarly Communication*, 6(1), p.eP2198.
Kirkwood, R. J. (2016). Collection Development or Data-Driven Content Curation? *Library Management*, 37(4–5), 275–284.
Koltay, T. (2015). Data Literacy: in Search of a Name and Identity. *Journal of Documentation*, 71(2), 401–415.
Koltay, T. (2016). Library and Information Science and the Digital Humanities: Perceived and Real Strengths and Weaknesses. *Journal of Documentation*, 72(4), 781–792.
Koltay, T. (2017). Beyond Literacies: The Evolving Landscape of Library Support to Research 2.0. *Zagadnienia Informacji Naukowej. Studia Informacyjne*, 55(1), 7–18.
Kouper, I. (2016). Professional Participation in Digital Curation. *Library & Information Science Research*, 38(3), 212–223.
LERU (2013). *LERU Roadmap for Research Data* [online]. League of European Research Universities Research Data Working Group, [16.05.2019], https://www.leru.org/publications/leru-roadmap-for-research-data
Liangzhi, Y. (2015). Back to the Fundamentals Again. *Journal of Documentation*, 71(4), 795–816.
Little, G. (2012). Managing the Data Deluge. *Journal of Academic Librarianship*, 38(5), 263–264.
Lucky, S., Harkema, C. (2018). Back to Basics: Supporting Digital Humanities and Community Collaboration Using the Core Strength of the Academic Library. *Digital Library Perspectives*, 34(3), 188–199.

Makani, J. (2015). Knowledge Management, Research Data Management, and University Scholarship: Towards an Integrated Institutional Research Data Management Support-System Framework. *VINE*, 45(3), 344–359.

McLure, M., Level, A. V., Cranston, C. L., Oehlerts, B., Culbertson, M. (2014). Data Curation: A Study of Researcher Practices and Needs. *Portal: Libraries and the Academy*, 14(2), 139–164.

Mierzecka, A. (2019). The Role of Academic Libraries in Scholarly Communication. A Meta-Analysis of Research. *Studia Medioznawcze*, 19(1), 42–55.

Muñoz, T. (2013). Data Curation as Publishing for the Digital Humanities. *Journal of Digital Humanities*, 2(3), 14–22.

Oliver, J. C., Kollen, C., Hickson, B., Rios, F. (2019). Data Science Support at the Academic Library. *Journal of Library Administration*, 59(3), 241–257.

Partlo, K., Symons, D., Carlson, J.D. (2015). Revolutionary or Evolutionary? Making Research Data Management Manageable. In: B.L. Eden (ed.). *Creating Research Infrastructures in the 21st Century Academic Library: Conceiving, Funding, and Building New Facilities and Staff* (pp. 175–201). Lanham, MD: Rowman and Littlefield.

Poole, A. H. (2015). How Has Your Science Data Grown? Digital Curation and the Human Factor: a Critical Literature Review. *Archival Science*, 15(2), 101–139.

Poole, A. H. (2016). The Conceptual Landscape of Digital Curation. *Journal of Documentation*, 72(5), 961–986.

Pouchard, L. (2016). Revisiting the Data Lifecycle with Big Data Curation. *International Journal of Digital Curation*, 10(2), 176–192.

Ray, J. (2012). The Rise of Digital Curation and Cyberinfrastructure: From Experimentation to Implementation and Maybe Integration. *Library Hi Tech*, 30(4), 604–622.

RECODE (2014) *RECODE Policy Recommendations for Open Access to Research Data* [online]. RECODE Project Consortium, [16.05.2019], https://trilateralresearch.co.uk/wp-content/uploads/2018/09/RECODE-D5.1-POLICY-RECOMMENDATIONS-_FINAL.pdf.

Ridsdale, C., Rothwell, J., Smit, M., Ali-Hassan, H., Bliemel, M., Irvine, D. (2015). *Strategies and Best Practices for Data Literacy Education. Knowledge Synthesis Report*. Halifax, NS: Dalhousie University.

Robinson, L. (2016). Between the Deluge and the Dark Age: Perspectives on Data Curation. *Alexandria*, 26(2), 73–76.

Rowley, J. (2007). The Wisdom Hierarchy: Representations of the DIKW Hierarchy. *Journal of Information Science*, 33(2), 163–180.

Scaramozzino, J. M., Ramírez, M. L., McGaughey, K. J. (2012). A Study of Faculty Data Curation Behaviors and Attitudes at a Teaching-Centered University. *College & Research Libraries*, 73(4), 349–365.

Schmidt, B., Shearer, K. (2016). *Librarians' Competencies Profile for Research Data Management* [online]. Joint Task Force on Librarians' Competencies in Support of e-Research and Scholarly Communication, [15.07.2019], https://www.coar-repositories.org/files/Competencies-for-RDM_June-2016.pdf

Southall, J., Scutt, C. (2017). Training for Research Data Management at the Bodleian Libraries: National Contexts and Local Implementation for Researchers and Librarians. *New Review of Academic Librarianship*, 23(2–3), 303–322.

Steinerová, J. (2012). Information Ecology – Emerging Framework for Digital Scholarship. *Libraries in the Digital Age (LIDA) Proceedings* [online], 12, [16.05.2019] http://ozk.unizd.hr/proceedings/index.php/lida/article/view/66/37

Tenopir, C., Pollock, D., Allard, S., Hughes, D. (2016). Research Data Services in European and North American Libraries: Current Offerings and Plans for the Future. *Proceedings of the Association for Information Science and Technology*, 53(1), 1–6.

Tenopir, C., Talja, S., Horstmann, W., Late, E. Hughes, D., Pollock, D. Schmidt, B., Baird, L., Sandusky, R. J., Allard, S. (2017). Research Data Services in European Academic Research Libraries. *LIBER Quarterly*. 27(1), 23–44.

Thomas, C. V., Urban, R. J. (2018). What Do Data Librarians Think of the MLIS? Professionals' Perceptions of Knowledge Transfer, Trends, and Challenges. *College & Research Libraries*, 79(3), 401–423.

# *Data curation* w bibliotekach akademickich jako część rewolucji cyfrowej

**Abstrakt**

**Cel/Teza:** W artykule przedstawiono rolę *data curation* w kontekście Nauki 2.0 i zarządzania danymi badawczymi.
**Koncepcja/Metody badań:** Wykorzystano krytyczną analizę wybranego piśmiennictwa przedmiotu.
**Wyniki i wnioski:** Pomimo względnej niejasności i wielości definicji terminu "data curation", biblioteki akademickie powinny zaangażować się w związane z nim działania.
**Ograniczenia badań:** W artykule skupiono uwagę głównie na publikacjach opisujących podejście teoretyczne.
**Zastosowanie praktyczne:** Obserwowane na całym świecie przedsięwzięcia związane z zarządzaniem danymi badawczymi i *data curation*. Wiele krajów i instytucji podjęło już działania na tym polu, ale ogólny poziom rozpoznania tematu nadal pozostaje niewysoki, dlatego potrzebne jest zwiększenie świadomości dotyczącej znaczenia tego zagadnienia.
**Oryginalność/Wartość poznawcza:** Przesłanki przedstawione w artykule oparte są na założeniu, że zmienia się podejście w postrzeganiu danych.

**Słowa kluczowe:**

Data curation. Digital curation. Nauka 2.0. Zarządzanie danymi badawczymi.

*Dr. habil. TIBOR KOLTAY, PhD is Professor, Chair of the Institute of Learning Technologies and Principal Researcher of the Information Society research group at Eszterházy Károly University, Hungary. His research ~~focuses on information literacy and other literacies, as well as Research Data Management. His most recent~~ publication is* Accepted and Emerging Roles of Academic Libraries in Supporting Research 2.0 *(2019, Journal of Academic Librarianship, https://doi.org/10.1016/j.acalib.2019.01.001).*

*Contact to the Author:*
*koltay.tibor@uni-eszterhazy.hu*
*Institute of Learning Technologies*
*Eszterházy Károly University, Hungary*
*Rákóczi út 53,*
*Jászberény,*
*H-5100, Hungary*

# Research Information Management in the Context of Open Science

Katarzyna Materska
ORCID 0000-00002-9194-380X
*Institute of Media Education and Journalism,*
*Cardinal Stefan Wyszynski University in Warsaw, Poland*

**Abstract**

**Purpose/Thesis:** The aim of this paper is to discuss the mutual relationship between open science and theory and practice in RIM (Research Information Management), especially apparent in CRISs (Current Research Information Systems), with an emphasis on the context of research assessment requirements, open science policies and CRIS interoperability.
**Approach/Methods:** The study has been based on the critical review of the newest literature (2015–2019) presenting international studies of chosen aspects. Institution-centric and researcher-centric approaches has been presented.
**Results and conclusions:** The emerging new ecosystem of open science changes the way research is done and modifies the monitoring and evaluating of it. CRIS must take into account the new transformation processes and research evaluation measures. This allows more transparency and more interaction with the individual researcher, as well as with the institutional, national and international stakeholders. Full interoperability and open standards are desired to improve the discoverability and reusability of research outputs and metadata for different purposes.
**Originality/Value:** The study shows the significance of new tendencies in RIM/CRIS for researchers and research-performing organizations at institutional, national and international level.

**Keywords**

Current research information system (CRIS). Open Science. Research information. Research information management. Scholarly communication.

*Received: 6 June 2019. Reviewed: 23 June 2019. Revised: 29 July 2019. Accepted: 30 July 2019.*

## 1. Introduction

Researchers, managers, funders, publishers, libraries, and other stakeholders in scholarly communications seek to foster greater research access, transparency, collaboration, use, and innovation. Analysis from practical study emphasizes that:

> High quality data about research activities and processes, so called research information (RI), are of strategic relevance and vital importance for both science communication and for research governance and policy (Biesenbender et.al, 2019, 143).

Decision makers and managers from Research Performing Organisations (RPOs) and Research Funding Organisations (RFOs) alike increasingly depend on indicators, reports and studies that draw data about research activities from research information systems

(Science Europe, 2016). Research information systems (RIS) are used to support the collection, integration, processing, storage and presentation of research information. In the light of the increased significance of RIS, it is pertinent to ask how the new ways of managing and reporting RI adopted in CRIS can foster the development of Open Science, and how the idea of Open Science shapes the CRIS infrastructure.

A German study observes that these questions have not received the attention they merit:

> an explicit reference to CRIS forming part of the Open Science movement is often lacking (Biesenbender et al., 2019, 143).

Until now the impact of open science on the development of Research Information Management and implementation of CRISs has been described with an emphasis on the role academic (research) libraries and information professionals play in supporting open access, copyright, metadata entry and validation, training and support with research data management (Clements & Proven, 2015; Bryant et al. 2017a; Bryant et al., 2018; Brennan, 2018).

## 2. Research Information Management and development of the basic RIM infrastructure

One of the most important components of information management in science is research information management (RIM). RIM refers to the integrated management of information about the research life-cycle, and about the entities which constitute it (e.g. researchers, research outputs, organizations, grants, facilities etc.)[1].

When thinking about RIM, we usually consider information regarding research activities and research results associated with institutions and their scientists, gathered from different units of the university (different HEIs[2]) or other research institutions. The information consists of continually updated data about researchers and their affiliations, research outputs (publications, datasets, and patents), grants and projects, academic service and honors, media reports and statements of impact. RIM is defined as information about research activities, not research data generated by researchers. Research information can be collected, curated and processed by research institutions for different internal and external purposes, and for various recipients (Biesenbender & Herwig, 2019; Stvilia et al., 2019).

In the new digital era RIM is defined as:

> the aggregation, curation and utilization of metadata about research activities (Bryant et.al, 2017a, 6).

Usually two main database components of the Research Information Management are indicated – Institutional Repositories (IRs) and Current Research Information Systems (CRISs), with the increasing attention to the latter (Ribeiro et al., 2016). The rising strategic importance of CRISs has been confirmed in several international studies, surveys and

[1] It is worth noticing that RIM is distinct from research data management (RDM), a similar-sounding term that is used to describe the processes researchers and institutions use for organizing, securing, archiving, and sharing research data throughout the research lifecycle. A data management plan (DMP) provides information on the data the research will generate, i.e., how to ensure its curation, preservation and sustainability and what parts of that data will be open (and how).

[2] Higher Education Institutions.

statements during the last four years (e.g. Ribeiro et al., 2016; Science Europe, 2016; Bryant et al., 2017b, Bryant et al., 2018) and the national projects like "Research Core Dataset" (RCD)[3] for the German science system (Riechert et al., 2017).

The term CRIS indicates a software system or solution to manage research information. Current Research Information Systems, now sometimes shortened to Research Information System (RIS), have different names, especially in North America where they are described as Research Networking System (RNS), Research Profiling System (RPS)[4], or Faculty Activity Reporting (FAR) (Bryant et al., 2018, 12, 19). Europeans are most familiar with the term "CRIS" or "RIS"[5]. The origins of CRIS systems in Europe date to the beginning of the 90s of the twentieth century, when they were principally used by the research offices (Bryant et al., 2018, 9). Gradually more and more university units were reported to the system. But for many years no regulations or standards for research information were implemented, which made the comparison of RI across different institutions, disciplines and countries impossible..

New challenges and technological possibilities have led to the creation of the new generation of CRIS (with more efficient and flexible infrastructure) to showcase institutional and national research (as well as research potential) (Biesenbender et al., 2019). Gathered in one place (one platform), RI does not only provide greater visibility and discoverability of institutional research activity, synchronise these data, and reduce burden to all involved of collecting and managing data about the research process, but it also facilitates internal and external data exchange, reporting and reuse. CRIS collects information by automatically synchronizing existing data volumes with various data sources (with different data formats and structures), so that this qualified RI is easily available to the end users (management) and gives them a better basis for decision making.

Among institutional CRIS software widely used in Europe in XXI century there are Pure (Elsevier), Converis (Clarivate Analytics), Elements (Symplectic), DSpace-CRIS (open source), VIVO (open source). Polish (local) system is Omega-Psir.

Institutional repository (IR) is the second main component of RIM. It "is a digital collection of research outputs (mainly publications and datasets) aiming to collect, preserve and disseminate the intellectual output of a higher education or research institution" (Ribeiro et al., 2016, 7). The analysis of the results from the CRIS/IR survey conducted by EUNIS[6] and euroCRIS[7] in 2015 revealed the complementarity of these two systems with repositories for managing research publications, and with CRISs for managing the institutional research information as a whole (Ribeiro et al., 2016, 5). In turn, *Practices and Patterns in Research Information Management. Findings from a Global Survey* report (Bryant et al., 2018) documents a tendency to merge functionalities of RIM systems and institutional repositories.

[3] The RCD is a standard for the collection, provision and exchange of research information.

[4] As RIM is treated here as an institutional information management (rather than personal), independent researcher profile systems (social networking platforms) like Research Gate or Academia.edu are not considered here. It is worth noting that some functionalities of RPS are taken into account by the creators of Omega-Psir, as an element of researcher-centric approach (Rybinski et al., 2018).

[5] Both terms will be used in this article interchangeably.

[6] EUNIS – the European University Information Systems Organization.

[7] EuroCRIS – an international not-for-profit association founded in 2002 to bring together experts on research information in general and research information systems (CRIS) in particular.

Rybinski et al. (2017) describe connecting IR with CRIS as a novel researcher-centric (not research-cenric) approach integrating the conflicting functionalities of IR and CRIS.

As we can observe an increasing tendency towards the adoption of wider CRIS functions, the following chapters of this paper are devoted to these systems.

## 3. Research Information Management in the context of new research assessment requirements and open science policies

New interest in RIM arose around 2010 when research institutions had to respond to national-level assessment policies, new policies on open access (open access mandates) and the demands of research funders. Therefore, it is obvious that CRIS had to be adopted to the emerging new ecosystem of open science, which changes the way research is done and modifies the way it is evaluated (Ribeiro, 2016, 11).

Open Science encompasses numerous components of the research life cycle, including open access to publications, research data, peer review, source software, standards, collaboration, educational resources, citizen science and more. It is based on the idea that scientific knowledge of all kinds should be openly shared as early as is practical in the discovery process (Nielsen, 2012). This systemic change is strongly emphasised in European Commission documents:

> Open Science represents new approach to the scientific process based on cooperative work and new ways of diffusing knowledge by using digital technologies and new collaborative tools (European Commission, 2015, 33).
> Open Science goes hand in hand with research integrity, and requires legal and ethical awareness on the part of researchers (O'Carroll et al., 2017, 5).

Open Science idea has a significant impact on scholarly communication models and new methodology of research assessment[8], in which it breaks with an assessment dependent on the place of publication (and "principle of inheritance of prestige"). National research assessment requirements and the resulting need for reporting of institutional research activity are indicated as the most significant driver of CRIS adoption.

Open Science policy impacts scientific practice on different levels – the individual (career assessment, grant awarding), local (institutional), organisational, national and international. CRIS must be able to monitor, evaluate and otherwise react to these new features to fulfil demands open science creates on all these levels. More recently, open-access mandates are also beginning to directly influence research output and publication management priorities, as supporting institutional compliance[9] with open science policies is another important incentive for pursuing RIM activities (Bryant et al., 2018, 31–32).

[8] A report written by the EC *Working Group on Rewards under Open Science* provides a matrix for the evaluation of research careers fully acknowledging open science practices. It proposes a large number of possible evaluation criteria for the assessment of six domains and 24 open science activities (European Commission, 2017).

[9] "Institutional compliance may mean different things from one country or institution to another, and may relate to satisfying mandates requiring research assessment reporting, open access, or research data management (...) [It] might also refer to individual funder requirements or to local institutional policies" (Bryant et al., 2018, 40).

## 4. CRIS interoperability – protocols, standards, identifiers

Research information workflows increasingly demand greater interoperability between internal and external systems. As research information systems proliferate, the issue becomes highly relevant, regarding interoperability both between CRISs themselves and with other, complementary systems, such as institutional repositories, systems at a national level and with other external stakeholders such as OpenAIRE.

> The growing need for improved interoperability between managing open access workflows and the curation of institutional research outputs metadata is giving rise to the increasing functional merging of RIM systems and institutional repositories [into hybrid platforms] (Bryant et al., 2018, 9–10).

To enhance the interoperability of RISs, Science Europe provided in November 2016 a set of common principles to guide their development. It invited all research organisations to adopt the following core principles for research information systems: flexibility, openness, FAIRness (to be Findable, Accessible, Interoperate, Reusable) and data entry minimisation. FAIR principles should always be implemented with respect to legal and ethical standards (Science Europe, 2016).

To ensure the standardized collection and interchangeability of RI and to be able to integrate as many decentralised stocks of RI as possible, there have been established (inter) national standards for supporting RIS. EUNIS – EuroCRIS survey (Ribeiro et al., 2015) indicated identified three most frequently adopted technologies and standards (in order of popularity): OAI-PMH, CERIF[10] and ORCID. Findings from the global survey *Practices and Patterns* (Bryant et al., 2019) confirmed this claim. This result could be explained by the emphasis placed on Open Access policies, interoperability and data exchange among different systems, and the unique identification of researchers. These three areas are all related not only to technological decisions, but to political ones as well, both at individual institution and at governmental level.

The latest release of the *OpenAIRE Guidelines for CRIS Managers* (Dvořák et al., ed., 2017) is a milestone to achieve interoperability between CRISs and OpenAIRE. Integration of CRISs in OpenAIRE is of mutual benefit. As a pan-European technical infrastructure with a strong support of Open Science experts network, OpenAIRE is harvesting metadata about research outputs from data sources across Europe and beyond. Defining interoperability guidelines it enables CRIS managers and other stakeholders to make the open scholarly communication environment much more efficient, innovative and creative.

In January 2018, the fourth project phase of OpenAIRE, *OpenAIRE-Advance*, started. Its aim is:

> making Open Science the default in Europe, reshaping the scholarly communication system towards openness and transparency serving as a trusted pillar of the European Open Science Cloud (EOSC) (OpenAire, 2018).

In this sense the OpenAIRE infrastructure can be itself considered a global CRIS system.

[10] The Common European Research Information Format (CERIF) has been developed as a standard to facilitate interoperability of CRIS systems within Europe. Today CERIF is being maintained by euroCRIS, its use is recommended across the EU.

## 5. Conclusion

Diversified research landscape with the diverting goals of the stakeholders makes RIM a serious problem in information management. RIM has received more attention in the recent years and various initiatives have been introduced in several countries to address issues related to research information.

The growing interest in research information systems (CRISs and institutional repositories) and increase in their strategic importance for higher education and research institutions come from their wide functionality and applicability in research governance and policy (e.g. sharing and monitoring institutional and national research potential, ensuring the quality of research information, fostering research and innovation, discovering and evaluating research, facilitating performance-based funding and generally augmenting excellence in research).

The national research assessment framework and open science policies are the key drivers strongly shaping priorities of RIM activities (in those countries and regions where they exist). CRISs are evolving to incorporate activities beyond research administration, and are increasingly used to support the FAIR communication, sharing and profiling of research through open access linkages and compliance tracking for publications, and to a lesser degree, research datasets (Bryant et al., 2018). Pablo de Castro highlights that the value of the institutional Open Science implementation strategy lies in the fact that CRIS is able to provide the required links between publications, projects, persons and affiliations (de Castro, 2018). The awareness that an institutional repository can be a tool supporting the implementation of Open Acces and Open Science is growing (see for instance de Castro, 2018; Rybinski et al., 2018).

RIM represents growing resource allocations by research institutions worldwide. We can observe evolving roles of CRIS – disambiguation, de-fragmentation, de-duplication; interoperability; links between publications, projects, persons and affiliations; new collaborative tools. All this shows that:

> a closed RIS in an environment of open science makes no sense (Azeroual at al., 2018, 36).

Given the richness of interlinked (meta)data on research present in CRISs, these systems could substantially contribute to the FAIRness of research and its products and as such become important building blocks for an open science infrastructure (Bryant et.al., 2018, 9). CRISs aggregate many types of data, harvest publications from a growing number of external sources, and serve as an important node interoperating within a large, complex scholarly communications landscape. Systems interoperability, and adoption of the standards, protocols and identifiers, which facilitate interoperability, lies at the heart of RIM (Bryant et al., 2018, 80).

Three main factors contribute to creating the open science ecosystem in which CRIS is developed and implemented: the governance structure and policies at the institutional level, at the national level, and at the international level[11]. The interplay between all these

[11] Governance instruments such as setting of research priorities, research funding, quality control procedures, performance monitoring etc. are parts of mechanisms and strategies of coordination of independent actors and organizations.

levels is crucial for the satisfying integration of CRIS and fulfilment of Open Science goals. National and institutional levels of cooperation suggested by Biesenbender et al. (2019) do not seem to be sufficient. To truly leverage the wealth of information available in CRISs it is of importance that they do not remain isolated resources on a local or national level, but become interconnected on an international scale. It is especially important in the increasingly networked research information, global competitions and rankings, external mandates. Confirmation of this international direction of developing CRISs is the topic of euroCRIS Spring 2019 membership meeting at CSC in Espoo/Helsinki: *Taking steps towards international CRIS systems*. But we should remember that RIM development, practices, incentives, priorities, maturity, scope, and nomenclature vary broadly by region (Bryant et al., 2018).

> Open science is an example of how quickly the context, needs and objectives related to research systems can evolve. The pace, direction and nature of such changes are unpredictable (Science Europe, 2016, 3).

The dynamism of the CRIS/RIM ecosystem creates the necessity of building global research information community.

## References

Azeroual, O., Saake, G., Abuosba, M., Schöpfel, J. (2018). Text Data Mining and Data Quality Management for Research Information Systems in the Context of Open Data and Open Science [online]. In: N. Ibnlkhayat, A.A. Bachr, A. Benchakroun, O. Roudiés (eds.). *Third International Colloquium on Open Access – Open Access to Science Foundations, Issues and Dynamics. ESI – CNRTS, 28–30 November 2018, Rabat, Morocco*, [30.07.2019], https://icoa2018.sciencesconf.org/data/actes_icoa18_fr.pdf

Biesenbender, S., Herwig, S. (2019). Support Structures to Facilitate the Dissemination and Implementation of a National Standard for Research Information – the German Case of the Research Core Dataset. *Procedia Computer Science*, 146, 131–141.

Biesenbender, S., Petersohn, S., Thiedig, C. (2019). Using Current Research Information Systems (CRIS) to Showcase National and Institutional Research (Potential): Research Information Systems in the Context of Open Science. *Procedia Computer Science*, 146, 142–155.

Brennan, N. (2018). CRIS Systems as Key e-Infrastructure Elements to Support Open Science Implementation within the European Research Area [online], In: *CRIS2018: 14th International Conference on Current Research Information Systems, Umeĺ, June 13–16, 2018*, [1.06.2019], https://dspacecris.eurocris.org/handle/11366/695

Bryant, R., Clements, A., Feltes, C., Groenewegen, D., Huggard, S., Mercer, H., Missingham, R., Oxnam, M., Rauh, A., Wright, J. (2017a). *Research Information Management: Defining RIM and the Library's Role* [online], Dublin, OH: OCLC Research, [12.04.2019], https://doi:10.25333/C3NK88

Bryant, R., Dortmund, A., Malpas, C. (2017b). *Convenience and Compliance: Case Studies on Persistent Identifiers in European Research Information*, 8 [online], Dublin, Ohio: OCLC Research [12.04.2019], https://doi:10.25333/C32K7

Bryant, R, Clements, A., de Castro P., Cantrell, J., Dortmund, A., Fransen J., Gallagher P., Mennielli M. (2018). *Practices and Patterns in Research Information Management: Findings from a Global Survey* [online], Dublin, OH: OCLC Research, [12.04.2019], https://doi.org/10.25333/BGFG-D241

Castro, P. de (2018). The Role of Current Research Information Systems (CRIS) in Supporting Open Science Implementation: the Case of Strathclyde. *ITlib.Informačné technológie a knižnice* [online], Special Issue, 21–30, [30.05.2019], https://dx.doi.org/10.25610/itlib-2018-0003

Clements, A., Proven, J. (2015). *The Emerging Role of Institutional CRIS in Facilitating Open Scholarship* [online]. St Andrews Research Repository, [30.07.2019], https://research-repository.st-andrews.ac.uk/bitstream/handle/10023/6867/LIBER2015-Abstract-Emerging_Role_CRIS_Open_Scholarship.pdf?sequence=1&isAllowed=y

Dvořák, J., Bollini, A., Rémy, L., Schirrwagen, J. (eds.) (2017). *OpenAIRE Guidelines for CRIS Managers* v. 1.1.1. [online]. OpenAIRE, [30.07.2019], https://openaire-guidelines-for-cris-managers.readthedocs.io/en/v1.1.1/

European Commission (2015). *Open Innovation, Open Science, Open to the World – a Vision for Europe* [online]. Publication Office of the European Union [30.07.2019], http://publications.europa.eu/resource/cellar/3213b335-1cbc-11e6-ba9a-01aa75ed71a1.0001.02/DOC_2

European Commission (2017). *Evaluation of Research Careers Fully Acknowledging Open Science Practices. Rewards, Incentives and/or Recognition for Researchers Practicing Open Science* [online]. European Commission, Working Group on Rewards under Open Science. Publication Office of the European Union [12.04.2019], https://ec.europa.eu/research/openscience/pdf/os_rewards_wgreport_final.pdf

Nielsen, M. (2012). *Reinventing Discovery: The New Era of Networked Science*. Princeton: Princeton University Press.

O'Carroll, C., Kamerlin, C.L., Brennan, N., Hyllseth, B., Kohl, U., O'Neil, G., Van Den Berg, R. (eds.) (2017). *Providing Researchers with the Skills and Competencies They Need to Practise Open Science. Open Science Skills Working Group Report* [online]. Publication Office of the European Union [30.07.2019], https://ec.europa.eu/research/openscience/pdf/os_skills_wgreport_final.pdf

OpenAire (2018). *Empowering Open Science: Kick Off of the OpenAIRE-Advance H2020 project* [online]. OpenAire Advance [30.07.2019], https://www.openaire.eu/empowering-open-science-kick-off-of-the-openaire-advance-h2020-project

Ribeiro L., de Castro P., Mennielli M. (2016). Final Report: EUNIS – EuroCRIS Joint Survey on CRIS and IR. *Copyright, Fair Use, Scholarly Communication, etc.*, [online], 5, [30.07.2019], https://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1004&context=scholcom

Riechert M., Biesenbender, S., Quix, Ch. (2017). Developing and Standardising Definitions for Research Information: Framework and Methods of Successful Process Documentation. *Procedia Computer Science*, 106, 260–267.

Rybinski H., Skonieczny L., Koperwas J., Struk W., Stepniak J., Kubrak W. (2017). Integrating IR with CRIS – a Novel Researcher-Centric Approach. *Program*, 51(3), 298–321.

Rybinski, H., Kubrak, W., Koperwas, J., Skonieczny, L., Struk, W. (2018). Omega-Psir – Institutional CRIS at Polish Universities. *ITlib. Informačné technológie a knižnice* [online], Special Issue, 36–44, [30.05.2019], https://dx.doi.org/10.25610/itlib-2018-0005

Science Europe (2016). *Science Europe Position Statement on Research Information Systems* [online]. Science Europe [30.07.2019], https://www.scienceeurope.org/our-resources/position-statement-on-research-information-systems/

Stvilia, B., Wu, S., Lee, D. J. (2019). A Framework for Researcher Participation in Research Information Management Systems. *The Journal of Academic Librarianship*, 45(3), 195–202.

# Zarządzanie informacją o nauce w kontekście otwartej nauki

**Abstrakt**

**Cel/Teza:** Celem artykułu jest omówienie wzajemnych relacji między otwartą nauką a teorią i praktyką zarządzania informacją o nauce, szczególnie widoczną w systemach typu CRIS (Current Research Information System), ze szczególnym uwzględnieniem kontekstu wymagań oceny w nauce, polityki open science oraz interoperacyjności systemów CRIS.
**Koncepcja/Metody badań:** Temat został opisany z wykorzystaniem krytycznego przeglądu najnowszej literatury (2015–2019) prezentującej badania międzynarodowe w wybranych aspektach. Przedstawiono podejścia skoncentrowane na instytucjach i naukowcach.
**Wyniki i wnioski:** Wyłaniający się nowy ekosystem otwartej nauki zmienia sposób prowadzenia badań i modyfikuje sposób ich monitorowania i oceny. CRIS musi uwzględniać nowe procesy transformacji i miary oceny badań. Oznacza to większą przejrzystość i większą interakcję z indywidualnym badaczem, interesariuszami instytucjonalnymi, krajowymi i międzynarodowymi. Pełna interoperacyjność i otwarte standardy są pożądane, aby poprawić wykrywalność i możliwość ponownego wykorzystania wyników badań i metadanych do różnych celów.
**Oryginalność/Wartość poznawcza:** Pokazanie znaczenia nowych tendencji w zarządzaniu informacją o nauce w systemach CRIS dla naukowców i organizacji prowadzących badania na poziomie instytucjonalnym, krajowym i międzynarodowym.

**Słowa kluczowe:**

Current research information system (CRIS). Informacja o nauce. Komunikacja naukowa. Otwarta nauka. Zarządzanie informacją o nauce.

*Dr. habil. KATARZYNA MATERSKA, PhD, is Professor at the Institute of Media Education and Journalism (Cardinal Stefan Wyszyński University in Warsaw). She is a researcher in the area of information and knowledge management, communication (including openness) in science, changes in the area of scientific libraries. Selected publications: Informacja w organizacjach społeczeństwa wiedzy (Warszawa, Wydaw. SBP, 2007); Information heuristics of information literate people (2014), published in: S. Kurbanoğlu, S. Špiranec, E. Grassian, D. Mizrachi, R. Catts (eds), Information Literacy. Lifelong Learning and Digital Citizenship in the 21st Century, ECIL 2014, Cham: Springer. https://doi.org/10.1007/978–3-319–14136-7_7; Polskie czasopisma naukowe z obszaru nauk przyrodniczo-leśnych w światowych bazach danych oraz ich ocena bibliometryczna (co-authors: I. Skrzecz, J. Szewczykiewicz, Sękocin Stary, 2019).*

*Contact to the Author:*
*katarzyna.materska@gmail.com*
*Cardinal Stefan Wyszynski University in Warsaw*
*Institute of Media Education and Journalism*
*Dewajtis 5*
*01-815 Warsaw, Poland*

# Research Data: Management and Opening. Polish and European Perspectives

Zuzanna Wiorogórska
ORCID 0000-0003-1015-2000
*Department of Information Studies, Faculty of Journalism, Information and Book Studies*
*University of Warsaw, Poland*

**Abstract**

**Purpose/Thesis:** This paper attempts to present the trends in management and opening of research data in Poland and the European Union, based on the analysis of the recently published Polish and European acts and documents as well as of other international initiatives which might influence scholarly publishing and scholarly communication.
**Approach/Methods:** An in-depth review of the latest documents was applied.
Results and conclusions: I focused on highlighting the key elements of the reviewed documents and initiatives, highlighting the directions for managing and opening of research data they set and the implications they might have for Polish and European science. I also sketched the possible inconsistencies between the European and Polish policies related to research data and scholarly communication.
**Research limitations:** The documents investigated for the purpose of this paper were either Polish or provided by the European Union (EU). I have not analyzed the national documents issued by the individual member states of the EU other than Poland. Hence, it is probable that some solutions on research data management and opening already taken on the level of individual member states have not been included in this paper.
**Practical implications:** This paper may encourage a reflection on the relationship between the regulations issued at the European (EU) or at the national (in this case, Polish), and the practices and requirements of scholarly communication which often contradict those regulations.
**Originality/Value:** This is the first analysis of the latest Polish and European documents and initiatives as related to data management and open data (open science).

**Keywords**

Open research data. Research data. Research data management.

*Received: 11 May 2019. Reviewed: 24 June 2019. Revised: 30 June 2019. Accepted: 10 July 2019.*

## 1. Introduction

We are living in the period of dynamic changes in the environment of research on the higher education. Some of these changes pertain to research data management and the so-called "opening of science". Within only six months, at the turn of 2018 and 2019, four important initiatives were launched; each of them may create new issues for research data. And yet, we may assume that there will be more transitions.

In European Union (EU), on 4 September 2018, Plan S was launched to accelerate the transition to open access up to 2020 (Science Europe, 2018a). Signatories of cOAlition S,

13 countries, agreed that all data and results of research financed by the research funding bodies must be published in the open access. On 31 May 2019 the revisions of Plan S' principles and implementation guidance were published[1], the timeline for implementation was extended to 2021 and the criteria of transformation towards open access were broadened, among others.

In turn, on 1 October 2018, The Constitution for Science – a new Act on Higher Education – came into force (Ministerstwo Nauki i Szkolnictwa Wyższego, 2018) in Poland. According to its authors, it gave Polish science an opportunity for pro-quality development and visibility in the world. It drew new challenges, particularly for Humanities and Social Sciences (HSS), as well as offered new outlooks for Polish scholarly communication.

On 30 January 2019 the European Commission published a report of the expert group titled *Future of Scholarly Publishing and Scholarly Communication* (European Commission, 2019). The expert group did propose a set of directions which scholarly communication should follow to be more effective, accessible and maximally usable; in general, the report had a pessimistic view of evaluation – and rankings-driven research and scholarly communication. According to its authors, scholarly publishing and communication can be meaningfully changed only if the agencies funding research initiate and lead such a change.

Finally in May, European University Association (EUA) published its latest report on the so-called *Big Deals* contracts[2] (Morais et al., 2019). For 31 national consortia representing 30 EU countries the total subscription fee for periodicals was reported to be 726.350.945 EUR (with average yearly increase of 3.6%). Although 68% of those countries had a national open access policy, 55% of contracts did not include any specific provision for open access in these countries' big deal contracts. It was reported that 56% of subscriptions costs were spent on Elsevier publications; Wiley came second at 18%.

Research environment and scholarly communication operate therefore in a specific "duality". On one hand, there is a call for actions aimed at science opening, while on the other hand, the researchers are "prisoners" of high-impact journals, offered by the leading publishers where choosing gold open access way is costly and not affordable to many. Between 2012 and 2018 we witnessed a movement called "The Cost of Knowledge boycott"[3], targeted against Elsevier's business practices (such as excessively high prices or/and lobbying in support of policies aimed at restricting the free exchange and access to information). It was an American grassroots initiative without a larger, international impact. However, as Tenent (2018, 39) pointed out, at least it worked as an expression of the academics' continued dissatisfaction with Elsevier – no other publisher has received this kind of negative attention – but then no other publisher fell into disfavor as much as Elsevier did.

[1] https://www.coalition-s.org/rationale-for-the-revisions/

[2] *Big deals* term is used to describe subscription agreements for electronic resources with the biggest publishers, like for ex. Elsevier, Springer-Nature, Taylor-Francis, Wiley, or American Chemical Association.

[3] http://thecostofknowledge.com

## 2. Research Data and Their Management

Horizon 2020, the EU Framework Programme for Research and Innovation (European Commission, n.d.), defines research data as information, particularly facts or numbers, collected to be examined and considered and as a basis for reasoning, discussion or calculation.

Research Data Management (RDM) has emerged as a topic of the 21st century science, related to what is the so-called fourth paradigm (Hey, 2009). The common use of ICT tools made research more intensive, technology – and data-driven and allowed handling huge volumes of data. Although there has been considerable investment in services, resources, and infrastructure to support researchers' data management needs, the level of researchers' awareness and skills regarding their own data management is still rather low; and RDM depends on institutional strategies and research habits in specific disciplines (Bryant et al., 2017; Burgi et al., 2017; Johnson et al., 2014).

Opening research data should be supported by increasing the awareness and skills required by RDM. Some studies have already confirmed low level of comfort and expert self-assessment with the life cycle of research data – and RDM-related topics (see e.g. Burgi et al, 2017; Conrad et al, 2017). Furthermore, one of OCLC's research reports highlighted the efficacy of education services in promoting RDM recognition with curation and expertise as the most important (Bryant et al., 2017). In this context, in 2017 and 2018 the Information Literacy Association coordinated an international research project ReDaM aimed to collect data about data literacy of academics and research students in higher education institutions. A part of this project was a study conducted in Poland (Wiorogórska et al., 2018). The purpose of the study was to explore the types of RDM of academic staff and research students and to assess the RDM awareness level of both target groups. The objective was to investigate the RDM practices and to find whether there were any differences between the practices of academic staff and of research students. The results of the Polish part of the study revealed that although a significant number of respondents was familiar with the basic concepts related to RDM, they had not used institutional solutions, including the Data Management Plan (DMP) in particular provided by their parent institutions. Most frequently, the respondents were not aware that such solutions were available. The study also noticed significant differences between academic staff and research students where their opinion on the usefulness of DMP (academic staff more often than research students perceived this tool as useful) and on the usefulness of formal training on metadata (research students more often than academic staff perceived this kind of training as useful) was concerned.

Undoubtedly, there is a need for enhancement of education on RDM for academic staff both to raise awareness and to present the benefits of data opening. So far in Poland, the Open Science Platform (*Platforma Otwartej Nauki*, http://pon.edu.pl) has been organizing regular workshops on RDM since 2015, approximately twice a year. However, they are not widely accessible since they take place on-site in Warsaw and the number of available places is limited. In Europe, academic libraries offer RDM education – it is enough to mention the solutions implemented by Switzerland, Estonia, or France (Burgi et al., 2017; Heidelberg University, 2018; Tarkpea & Seiler, 2016). However, system solutions, founded by government or research agencies, like those established in Australia or Great Britain, such as national data centers or data services support RDM more successfully.

In 2004, the British Data Curation Centre (DCC)[4] was established to provide expertise in digital information curation and hence to support researchers in the United Kingdom. This big multi-structured organization offers trainings, guidelines, events, and support in data management planning, data preservation, copyright issues, or creating institutional policies. Recently, it has been actively involved in open science advocacy in the UK.

Four years later, in 2008, the Australian National Data Service (ANDS)[5] was founded by the Australian Government. ANDS offers support in managing research data, but it also offers services that allow researchers to share their data, making them more visible and reusable. For example, ANDS hosts Research Data Australia (RDA)[6] discovery service that enables researchers to access and reuse datasets created by Australian researcher organizations, government agencies, and cultural institutions. Although their main mission is to: "make Australia's research data assets more valuable for researchers, research institutions and the nation", by opening Australian data, ANDS helps to promote national research and makes it visible worldwide. Thus, Australia has already opened its data and made them FAIR (for details on FAIR see section 3.1).

## 3. Data Opening

The Open Data Handbook defines open data as

> data that can be freely used, re-used and redistributed by anyone – subject only, at most, to the requirement to attribute and sharealike (Open Knowledge International, n.d.).

Open data is a crucial component of the so-called "Open Science", a concept defined by OECD as

> unhindered access to scientific articles, access to data from public research, and collaborative research enabled by ICT tools and incentives (OECD, 2019),

and by the European Commission as

> a new approach to the scientific process based on cooperative work and new ways of disseminating knowledge, improving accessibility to and re-usability of research outputs by using digital technologies and new collaborative tools (European Commission, 2018a, 12).

### *3.1. European Union and Data Opening*

The European Union has been encouraging the opening of data and open science initiatives through several projects aimed at supporting policy makers, stakeholders, or researchers. Open Science[7] portal hosted by the European Commission fulfils one of the goals of the EU research and innovation policy. It was created as an online hub to share European Commission's news, events, and publications related to Open Science.

[4] http://www.dcc.ac.uk/
[5] https://www.ands.org.au/
[6] https://researchdata.ands.org.au/
[7] http://ec.europa.eu/research/openscience/index.cfm

In October 2017 European Open Science Cloud (EOSC)[8] Declaration was signed. EOSC's aim is to establish a trustworthy environment for processing and hosting research data to support European science. One of EOSC's strategic goals is to implement FAIR Research Data Principles.

The FAIR Principles (the abbreviation "FAIR" stands for: Findable, Accessible, Interoperable, Reusable) were drafted in 2015 at a Lorentz Center workshop in Leiden, The Netherlands. Since then, they have received worldwide recognition by various organizations, including the European Commission, as a useful framework for thinking about sharing data in a way that will maximize use and re-use. The authors of a Final Report and Action Plan on FAIR Data admitted that introduction of the FAIR principles requires significant resources at the disciplinary level so that the data-sharing framework might be developed, i.e. principles and practices, community-agreed data formats, metadata standards, tools, data infrastructures, etc. (European Commission, 2018b, 11).

To support open scholarly communication and foster open science in EU, OpenAIRE[9] was established. This non-profit civil partnership is an European Commission-related project financed from Horizon 2020 program. It works in EU member states through the network of National Open Access Desks (NOADs) whose task is to connect researchers, research institutions, policy makers, citizen scientists, educators, industry, and the general public at a national level on the one end, and the OpenAIRE services on the other (OpenAIRE, 2018, 4). In Poland, NOAD is held by Interdisciplinary Centre for Mathematical and Computational Modelling (ICM UW), the leader of the abovementioned Open Science Platform.

Facilitating Open Science Training for European Research (FOSTER)[10] project was launched on the similar principles to OpenAIRE. Founded by the EU's Seventh Framework Programme, a partnership of 11 EU universities and organizations (LIBER, DCC, and ICM UW among others). It aimed to provide a European-wide training program, targeted particularly at young researchers, so as to help them understand and implement open access policies in Horizon 2020 (described below). Therefore, it focused more on cultural change, the modification of previous practices in scholarly communication and the researchers' behavior to ensure that open science becomes a standard. The duration of FOSTER's two phases was 4.5 years (between February 2014 and May 2019). It resulted in a number of events, onsite and online trainings, videos, podcasts, and Open Science Toolkit (a set of ten online courses).

Nevertheless, all those documents, initiatives, and projects described above are merely EU's recommendations, declarations, or supporting actions. They are not regulations, and thus they do not have a binding legal force. In practice, it means that the research governing agencies or research funding bodies in EU member states are not obliged to follow these suggestions. It also means that the researchers are not obliged to adhere to open science principles as they will not be evaluated according to them. Therefore, they do not have to be afraid that there will be any negative consequences to actions contradicting these principles.

The opposite is the case with the Horizon 2020 Program. In fact, it was the first EU-funded research program that imposed the rule of open access on its beneficiaries. Under Horizon

[8] http://ec.europa.eu/research/openscience/index.cfm?pg=open-science-cloud

[9] https://www.openaire.eu/

[10] https://www.fosteropenscience.eu/

2020 the European Commission offers prestigious ERC grants for conducting pioneering, innovative studies to ensure research excellence in all fields of science. Previously, ERC grants were founded by the EU's Seventh Framework Programme. Now, under Horizon 2020-financed projects

> beneficiaries of ERC grants must ensure open access (free of charge, online access for any user) to all peer-reviewed scientific publications relating to its results" (European Research Council, 2017, 3).

The grantees choose between a green and a gold open access route; they must also deposit an electronic copy of their publication(s) in a repository and ensure the open access to the deposited version. Horizon 2020 also offers a Pilot on Open Research Data. This means that grantees may, but are not obliged to, facilitate the access, re-use and preservation of research data generated during their research work. These regulations are model-based solutions for opening access to research data. However, it still concerns only a minority of scholarly input in the EU. For instance, in Poland there are only 16 on-going ERC grants (February 2019 status)[11], although a new call for ERC Advanced Grants applications was opened in May 2019.

### *3.2. National and International Undertakings*

When talking about the opening of research data and, more broadly, the open science, it is important not only to describe the legal foundations, but also to highlight the obstacles to the process of opening.

In November 2018, a report on Elsevier titled *Democratising Knowledge* (Tennant, 2018) was published. It provides an in-depth exploration of the business model and publication practices of this biggest scholarly publisher, often called the monopolist in the scholarly publication environment. These practices not only limit the open access to research results. By promoting different commercial bibliometric and evaluation tools (e.g. Scopus or SciVal) specific to itself, Elsevier forces the higher education institutions into a position of dependence and costly cooperation.

The second obstacle might be national regulations and policies' reliance on bibliometrics and preference for high-impact, internationally indexed publications, preferably in English and in internationally recognized journals. Such a scholarly communication model is not only very expensive, but it also limits the scope of possible places of publication to the most recognized ones, i.e. those managed by the biggest worldwide publishers (like Elsevier, IEEE, Nature, etc.).

There were, however, some higher education institutions, consortia, or science governing agencies who dared to show their resistance against the practices limiting the academic and research freedom. In February 2019, University of California (UC)[12] announced the termination of their subscription to Elsevier. To justify this decision, they said that being a leader in the global movement toward open access to publicly funded research is contradictory

[11] Source: National Contact Point for Research Programmes of the European Union (http://www.kpk.gov.pl/?page_id=10227).

[12] University of California encompasses 10 campuses (for. ex. Berkeley, Los Angeles, San Diego) and three national research laboratories for the US Department of Energy (Los Alamos, Lawrence Berkeley, and Lawrence Livermore). See https://www.universityofcalifornia.edu/

with spending large amounts of money on access to knowledge[13]. While negotiating, UC wanted Elsevier to ensure universal open access at least to UC research; the publisher did not comply to meet this term.

One month later, Norway also decided to terminate its institutions' subscriptions to Elsevier. And what did not work in the US, seems to have worked in Norway. Elsevier quickly sat at the negotiating table and in April 2019 they agreed with the Norwegian consortium for higher education and research on a two-year pilot program aimed, on one hand, at providing open access to research for a Norwegian research community and, on the other hand, at allowing Norwegian researchers to publish in open access. In result, seven universities and 39 research institutions across the country have full read access to the whole Science Direct Freedom Collection. Starting from January 2019, articles with Norwegian corresponding authors are published in open access with a CC-BY-license. Authors who have already published with Elsevier in an eligible journal will be contacted by Elsevier and offered to make the article in open access at no extra cost. This pilot program will ensure that about 90% of the article output from Norwegian institutions in Elsevier journals will be published with an open license in Gold Open Access and in hybrid journal titles.

These two cases may prove that it is not the law that regulates scholarly communication patters at the national level, but rather persistence and local initiatives. What occurred at the University of California did not have an impact on the whole California state, not mentioning the whole country. The Norwegian case shows that an active consortium in a smaller (as a reminder: having seven universities only), but well-governed country may achieve more than a big federation.

### *3.3. Poland*

In 2015, The Ministry of Science and Higher Education published a 20-page document titled *Directions of the development of open access to research publications and research results in Poland*[14], where it presented the principles of open access policy and proposed recommendations for open access to publications and to the results of research funded by financing agencies as well as by research units, higher education institutions, and publishers. The Ministry recommended national research financing bodies that they apply and publicize the rules of open access. This document was based on the EU Commission Recommendation of 17 July 2012 on access to and preservation of scientific information[15], but it was published in Poland only three years after its publication in the Official Journal of the European Union!

Apparently, nothing has changed in this field in Poland since then, so the Ministry somehow managed to put pressure on one of the biggest research financing agencies, supervised by the Ministry of Science and Higher Education, the National Science Centre (NCN), to enforce these recommendations. This way, in March 2019 the NCN Director

[13] The full text of statement is available at https://www.universityofcalifornia.edu/press-room/uc-terminates-subscriptions-worlds-largest-scientific-publisher-push-open-access-publicly

[14] https://www.gov.pl/documents/1068557/1069061/20180413_Kierunki_rozwoju_OD_wersja_ostateczna.pdf [own translation of the title].

[15] https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:32012H0417

issued an open letter[16] where, having referred to the same EU Recommendation of 2012 (seven years later!), he reminded about the guidelines intended to make research data management more uniform across various member states (Science Europe, 2018b) and the project of incorporating a requirement of presenting data management plan (DMP) into the applications for a NCN research grant. According to this letter, DMP will be an element of the first stage of implementing open access to research data in Poland.

For those who are aware what DMP serves, this statement sounds simply ridiculous. Yes, DMP is a very important element of the research cycle and it may help researcher to describe what s/he is going to do with data during and after her/his research project. Obviously, it may also facilitate setting a standard of data openness (and thus, the possibility of sharing) or confidentiality, and to define the period and conditions of data preservation after the research is completed (see e.g. Buddenbohm et al., 2016; Carlson et al., 2011; Higman & Pinfield, 2015). However, without a solid legal national policy DMP itself will not ensure open science. This process should be first managed by the state: firm decisions about open access roads should be taken and then, elements of more far-reaching policies might be implemented in the research grants application requirements.

The researchers in Poland must be first assured that when opening their data, they will not lose their evaluation points, that publishing in open access will give them the same career benefits as publishing at the biggest publishers, or that they will be given money for choosing the golden open access route.

## 4. Conclusion

Even though the analysis of several documents and initiatives showed how important openness seems to be for stakeholders and EU governance, my vision of the short-term perspectives for the nearest future of data opening is not optimistic. In everyday research and communication practice, the non-profit ideas meet a strong pro-profit movement (the so-called paywalls). Proponents of open access and open science policies use catchy slogans, but as long as they will not be supported by strong legal acts, there will not be a definitive change. It is important to remember that regulations and directives are the sole binding legislative acts on the European level. Recommendations, opinions, or decisions are presenting a view and preparing a line of action[17]. And to inspire a change, also a change of mentality, on the one hand the law is needed, and on the other hand, more sustainable and open alternatives for researchers. Currently, the majority of EU state members rely on the contents offered by one or two biggest publishers, not only for subscriptions to e-resources, but also for evaluation of science. For example in Poland, since 2018, InCites by Web of Science and SciVal by Elsevier are purchased under the national license and ultimately, it will be one of these tools that is used for the evaluation of researchers under the new Act on Higher Education, which will only strengthen the position of the big publishers' on the market and in the research environment. In 2017, the Association of European Research

[16] https://www.ncn.gov.pl/sites/default/files/pliki/2019_04_03_pismo_dyrektora_NCN_zarzadzanie_danymi_naukowymi.pdf

[17] https://europa.eu/european-union/eu-law/legal-acts_en

Libraries (LIBER) presented Five Principles for libraries to use when conducting Open Access negotiations with publishers[18] (since LIBER joins academic libraries which mostly take care of e-resources subscriptions in higher education institutions). The report on *big deals* (Morais et al., 2019) cited above revealed that two years later the majority of EU universities has not implemented those principles, so they are still paying double: the researchers pay for being published and the universities pay for access to their researchers' publications. I agree with the authors of Future of Scholarly Communication report. No matter how much will be written about opening data and how many institutions will be encouraged to promote this movement,

> (t)he evolution of open access and open science is tied to the ways in which these actors will cooperate with each other, or struggle against each other, and for this reason, their futures remain unclear (European Commission, 2019, 23).

## References

Bryant, R., Lavoie, B., Malpas, C. (2017). *A Tour of the Research Data Management (RDM) Service Space. Part 1: The Realities of Research Data Management*. Dublin, Ohio [online]. OCLC Online Computer Library Center, Inc. [9.07.2019], http://www.oclc.org/content/dam/research/publications/2017/oclcresearch-research-data-management-service-space-tour-2017-a4.pdf

Buddenbohm, S., Cretin, N., Dijk, E., Gaiffe, B., De Jong, M., Le Teiller-Becquart, N., Minel, J.-L. (2016). *State of the Art Report on Open Access Publishing of Research Data in the Humanities* [online]. Archive ouverte HAL [9.07.2019], https://halshs.archives-ouvertes.fr/halshs-01357208/document

Burgi, P.-Y., Blumer, E., Makhlouf-Shabou, B. (2017). Research Data Management in Switzerland: National Efforts to Guarantee the Sustainability of Research Outputs. *IFLA Journal* [online], 43(1), 5–21, https://doi.org/10.1177/0340035216678238

Carlson, J., Fosmire, M., Miller, C. C., Sapp Nelson, M. (2011). Determining Data Information Literacy Needs: A Study of Students and Research Faculty. *Library and the Academy*, 11(2), 629–657.

Conrad, S., Shorish, Y., Whitmire, A. L., Hswe, P. (2017). Building Professional Development Opportunities in Data Services for Academic Librarians. *IFLA Journal* [online], 43(1), 65–80. https://doi.org/10.1177/0340035216678237

European Commission (n.d.). Research & Innovation. Participant Portal H2020 Online Manual [online], Research and Innovations. European Commision [9.07.2019], http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/open-access_en.htm

European Commission (2018a). Commission Recommendation (EU) 2018/790 of 25 April 2018 on Access to and Preservation of Scientific Information. *Official Journal of the European Union*, 134, 12–18.

European Commission (2018b). *Turning FAIR into Reality. Final Report and Action Plan on FAIR Data*. Brussels: European Commission, Directorate General for Research and Innovation.

European Commission (2019). *Future of Scholarly Publishing and Scholarly Communication. Report of the Expert Group to the European Commission* [online]. Publications Office of the European Union [9.07.2019], https://publications.europa.eu/en/publication-detail/-/publication/464477b3-2559-11e9-8d04-01aa75ed71a1

[18] https://libereurope.eu/blog/2017/09/07/open-access-five-principles-for-negotiations-with-publishers/

European Research Council (2017). *Guidelines on Implementation of Open Access to Scientific Publications and Research Data in Projects Supported by the European Research Council under Horizon 2020* [online]. European Commission [9.07.2019], http://ec.europa.eu/research/participants/data/ref/h2020/other/hi/oa-pilot/h2020-hi-erc-oa-guide_en.pdf

Heidelberg University (2018). Competence Centre for Research Data [online]. Competence Centre for Research Data [9.07.2019], https://data.uni-heidelberg.de/index.en.html

Hey, A. J. G. (2009). *The fourth paradigm. Data-intensive scientific discovery*. Redmond, Washington: Microsoft Research.

Higman, R., Pinfield, S. (2015). Research Data Management and Openness: The Role of Data Sharing in Developing Institutional Policies and Practices. *Program: Electronic Library and Information Systems* [online], 49(4), 364–381, https://doi.org/10.1108/PROG-01-2015-0005

Johnson, L., Adams Becker, S., Estrada, V., Freeman, A. (2014). *NMC Horizon Report: 2014 Higher Education Edition*. Austin, Texas.

MNiSW (2018). *Konstytucja dla Nauki* [online]. Ministerstwo Nauki i Szkolnictwa Wyższego [9.07.2019], https://konstytucjadlanauki.gov.pl/

Morais, R., Stoy, L., Borrell-Damián, L. (2019). *2019 Big Deals Survey Report*. European University Association.

OECD (2019). Open Science [online]. Organisation for Economic Co-operation and Development [9.07.2019], http://www.oecd.org/science/inno/open-science.htm

OKI (n.d.). Open Data Handbook [online]. Open Knowledge International [9.07.2019], http://opendatahandbook.org/

OpenAIRE. (2018). *Statutory Documents for the Non-Profit Civil Partnership "OpenAIRE"* [online]. OpenAIRE [9.07.2019], https://www.openaire.eu/organization

Science Europe (2018a). Plan S [online]. Coalition S [9.07.2019], https://www.coalition-s.org/

Science Europe (2018b). *Practical Guide to the International Alignment of Research Data Management* [online]. Science Europe [9.07.2019], https://www.scienceeurope.org/wp-content/uploads/2018/12/SE_RDM_Practical_Guide_Final.pdf

Tarkpea, T., Seiler, V. (2016). Integrating Data Literacy into Information Literacy e-Course for PhD Students [online]. In: S. Spriranec, S. Kurbanoglu, H. Landova, E. S. Grassian, D. Mizrachi, L. Roy, & D. Kos (eds.). *The Fourth European Conference on Information Literacy : (ECIL) October 10th-13th, 2016, Prague, Czech Republic : abstracts*. Prague: Association of Libraries of Czech Universities. European Conference on Information Literacy (ECIL) [9.07.2019], http://ecil2016.ilconf.org/wp-content/uploads/2016/10/ECIL2016_BoA.pdf

Tennant, J. (2018). *Democratising Knowledge: A Report on the Scholarly Publisher, Elsevier* [online]. Education International [9.07.2019], https://issuu.com/educationinternational/docs/2018_eiresearch_elsevier_final_en

Wiorogórska, Z., Leśniewski, J., Rozkosz, E. (2018). Data Literacy and Research Data Management in Two Top Universities in Poland. Raising Awareness. In: S. Kurbanoğlu, J. Boustany, S. Špiranec, E. Grassian, D. Mizrachi (eds.). *Information Literacy in the Workplace: 5th European Conference, ECIL 2017, Saint Malo, France, September 18–21, 2017, Revised Selected Papers* (pp. 205–214). Cham: Springer. Springer Link, https://doi.org/10.1007/978-3-319-74334-9_22

# Dane badawcze: zarządzanie i otwieranie. Perspektywy polskie i europejskie

**Abstrakt**

**Cel/Teza:** Celem artykułu jest nakreślenie kierunków zarządzania i otwierania danych badawczych w Polsce i w Unii Europejskiej, na podstawie analizy ostatnio opublikowanych polskich i europejskich aktów prawnych i dokumentów innego typu, a także różnych międzynarodowych przedsięwzięć, które mogą mieć wpływ na komunikację i publikowanie naukowe.
**Koncepcja/Metody badań:** Wykorzystano krytyczną analizę dokumentów prawnych i piśmiennictwa przedmiotu.
**Wyniki i wnioski:** Skupiono się na uwypukleniu kluczowych elementów omawianych dokumentów i inicjatyw, podkreślając, jakie kierunki wyznaczają one w zarządzaniu i otwieraniu danych badawczych i jaki wpływ mogą wywrzeć na polską i europejską działalność naukową. Nakreślono także możliwe sprzeczności pomiędzy europejskimi (unijnymi) a polskimi politykami dotyczące danych badawczych i komunikacji naukowej.
**Ograniczenia badań:** W artykule skupiono uwagę na dokumentach i inicjatywach polskich i wydanych (zainicjowanych) przez Unię Europejską (UE). Nie poddano analizie dokumentów poszczególnych państw – członków UE. Jest zatem możliwe, że pewne rozwiązania w zakresie zarządzania i otwierania danych badawczych zostały podjęte na szczeblu krajowym w niektórych z tych państw, jednak nie zostały one uwzględnione w artykule.
**Zastosowanie praktyczne:** Artykuł może stanowić podstawę do refleksji nad powiązaniami pomiędzy regulacjami wydanymi na szczeblu europejskim (unijnym) i krajowym (w tym przypadku polskim) a dotychczasowymi praktykami i wymaganiami stawianymi obecnie w komunikacji naukowej, często stojącymi w sprzeczności z tymi regulacjami.
**Oryginalność/Wartość poznawcza:** Zgodnie z wiedzą autorki, artykuł jest pierwszą próbą analizy najnowszych polskich i europejskich dokumentów i inicjatyw związanych z zarządzaniem danymi badawczymi i otwieraniem danych (otwartą nauką).

**Słowa kluczowe**

Dane badawcze. Otwarte dane badawcze. Zarządzanie danymi badawczymi.

*ZUZANNA WIOROGÓRSKA, PhD, is Assistant Professor at the Faculty of Journalism, Information and Book Studies, University of Warsaw, Academic Librarian at The University of Warsaw Library and Associate Member of Laboratoire GERiiCO (University of Lille, France). Her main areas of interest include research data management, information literacy in multicultural context, users of information, and 21st century skills. Her most important recent publications are:* Data Literacy and Research Data Management in Two Top Universities in Poland. Raising Awareness *(2018, co-authors: J. Leśniewski, E. Rozkosz, E.; published in: S. Kurbanoğlu, J. Boustany, S. Špiranec, E. Grassian, D. Mizrachi (eds.),* Information Literacy in the Workplace: 5th European Conference, ECIL 2017, Cham: Springer. *https://doi.org/10.1007/978-3-319-74334-9_22);* The Importance of Information Literacy for Asian Students at European Universities: Outlines. *(2018, Paideia, https://doi.org/10.33034/PAIDEIA.2018.6.1.103);* Nauka o informacji w okresie zmian. Innowacyjne usługi informacyjne *(co-editors: B. Sosińska-Kalata, P. Tafiłowski, Warszawa, 2017).*

*Contact to the Author:*
*z.d.wiorogorska@uw.edu.pl*
*Department of Information Studies*
*Faculty of Journalism, Information and Book Studies*
*University of Warsaw*
*Nowy Świat 69*
*00-927 Warszawa*

# Exploring Humanities Research Data in Figshare

Samia Takhtoukh
ORCID 0000-0001-8543-5537
*GERiiCO Laboratory*
*University of Lille, France*

**Abstract**

**Purpose/Thesis:** The purpose of this study is to explore the uptake and use of Figshare by humanities community and to know if opening of the research data is enough to make them visible and findable.
**Approach/Methods:** The overall research design was to gather a sample of humanities artifacts found on Figshare through its API, and then to analyze such indicators as the number of articles by discipline, the evolution of contribution per year, and the number of views and downloads by affiliation to an institution.
**Results and conclusions:** The humanities community engagement with Figshare varies depending on the discipline; in addition, even when research data is published, 50% is rarely viewed. This means that making research data open is not enough to make it visible and to ensure their reuse. The final conclusion is that articles affiliated to an institution have more chance to be visible than unaffiliated ones.
**Research limitations:** The limit of our study is that it can only analyze data present in Figshare and not the data that is missing, so we cannot know the reasons for the small visibility of humanities research data; the study of other data warehouses is necessary to give more explanation, but the results conform to the pattern visible in the data presented in several studies of research management practices.
**Originality/Value:** The first study that explores the humanities contents the multidisciplinary data repository Figshare.

**Keywords**

Figshare. Humanities. Open data. Research data.

*Received: 14 June 2019. Reviewed: 4 July 2019. Revised: 22 July 2019. Accepted: 12 August 2019.*

## 1. Introduction

Since the sharing of research data is of great scientific, economic and social importance, the movement of Open Access is expanding more and more to research data. Indeed, the sharing of research data has several advantages in terms of efficiency, power and rigor. To take advantage of these benefits, many funding agencies put policies in place to promote the optimal use and reuse of data in which their funds were invested. They encourage good data practices, investing in data infrastructure and raising data awareness (Swan & Brown, 2008). In addition, several formalized data management services and tools (Figshare, Zenodo, Nakala, Dryad) have emerged as a result of funding agency's requirement that the grant applicants explicitly document their methods of storing research data and make them available for future use.

Specific studies of research data management practices were conducted at different universities, in different countries and in different years. They have shown that the management

of research data is still not common among researchers in the humanities (Chowdhury et al., 2018; Neylon, 2017; Prost & Schöpfel, 2015; Stuart et al., 2018). A struggle to organize the data in a presentable and useful way, the uncertainty as to the copyright and the licenses, the ignorance of the warehouse to use, the lack of time for the deposit of the data and the costs of sharing the data (Stuart et al., 2018) are the main open data challenges in the social sciences and humanities. With all these obstacles, it seems that the researchers are very far from embodying the ideal of open science in their practice.

But it is also important to conduct studies of general multidisciplinary repositories in order to know if the willingness to make the research data more open depends on the discpiline. It is also good to know if opening research data is enough to make it visible and findable and if the constraints of funding agencies and institutions policies have an impact on making research data available to the public.

In this article, we will focus on Figshare, which is a multidisciplinary research data repository. Established in 2011, it allows researchers from all disciplines to upload online any type of electronic information that can be used by others. Technically, Figshare users can make all their search results available in a way that is citable, shareable and discoverable. Data may be shared privately with collaborators or made public in the name of open research, or to comply with the mandates of funders and publishers.

Figshare also offers a Figshare for Institutions service which allows institutions to easily aggregate research at the departmental and institutional levels, automatically providing an institutional repository with reporting functions. They can make use of indicators about all the research they generate, which is not possible with the silo systems that exist in many research organizations today. Figshare for Institutions ensures that their data management requirements are made clear for the researcher.

As Figshare does not target a specific discipline, it allows various types of resources to be uploaded, making it the main universal scientific repository of this type. It will allow us to take account of the disciplines in our comparative study. In addition, Figshare publishes views, shares and downloads counts. We consider these counts informative for digital resources (Konkiel & Scherer, 2013), therefore we will use this alternative online metrics to measure the visibility of data. The following research questions drive this of Figshare: Is there any difference in community engagement depending on the discipline? Does institution affiliation have any impact on the visibility of research data on Figshare?

To answer these questions, we have created a web application to automatically query the Figshare API; the source code is available as open source in GitHub[1].

## 2. Conceptual background

### *2.1. Potential benefits of open research data*

The movement of Open Access is expanding more and more to research data, which is defined as following:

[1] The source code is accessible at: https://github.com/aallou/figshare-vue-app

> Research data refers to information, in particular facts or numbers collected to be examined and considered and as a basis for reasoning, discussion, or calculation. In a research context, examples of data include statistics, results of experiments, measurements, observations resulting from fieldwork, survey results, interview recordings and images (Open Access, n.d.).

The sharing of research data is of great importance; in particular, it helps to accelerate innovation through the exploitation, which can stimulate innovation and increase the collective knowledge. This knowledge will be translated into products and services with significant social, economic and scientific value (Ali-Khan et al., 2018). On the other hand, the sharing of research data guarantees an increase in quality. In fact, openly sharing research data will lead to more controllable and reproducible research, which will make it possible to easily validate the results presented in scientific publications. In addition, making research data open by default will help ensure greater transparency and confidence, and help to prevent massive data loss (Jacobson et al., 2014).

To take advantage of these benefits, many funding agencies put policies in place to promote the optimal use and reuse of data in which funds were invested. For example, in 2013 the European Commission launched a pilot project of open access to data from publicly funded research only (Research Infrastructures, n.d.); the pilot expanded to all disciplines in 2016. All stakeholders involved in scientific research, including people from academia, industry, funding agencies and scientific publishers, have come together to jointly develop and agree on a set of concise and measurable principles known by the acronym FAIR, suggesting that the data should be Findable, Accessible, Interoperable and Reusable. The intention is that these principles will serve as guidelines for those wishing to improve the reuse of research data. It should be noted that the FAIR principles place particular emphasis on improving the ability of machines to search and use data automatically, as well as facilitating their reuse by individuals. Funding agencies increasingly require the research data (and publications) resulting from funded research projects to be published in open access. However, as numerous as the potential opportunities are, there remain obstacles to managing and opening up research data.

### *2.2. Issues around research data for institutions and researcher*

As funding agency requires that the grant applicants explicitly document how research data will be stored and made available for future use, many institutions have adopted models and developed tools to help grant applicants submit competing data management plans (Weber, 2013). Nevertheless, the development of these policies (HLEG, 2010) is not necessarily followed by their implementation in organizations; moreover, studies have shown that there is a gap between funding bodies and best practices for institutions supporting research activities (Weber, 2013). Indeed, the openness of the data represents a profound change, in particular in human and social sciences, and poses several challenges:

> To collect, curate, preserve and make available ever-increasing amounts of scientific data, new types of infrastructures will be needed. The potential benefits are enormous but the same is true for the costs. We therefore need to lay the right foundations and the sooner we start the better (HLEG, 2010).

Data sharing can cause practical problems (Borgman, 2012), especially in terms of interoperability, standardization, collaboration, curation, decentralization of content integrity

(Girard, 2017). In addition, researchers who are the major producers of data can be very cautious and their current practices do not necessarily fit into this movement (Prost & Schöpfel, 2015).

The questions for the researcher are not simple either; the openness of the data implies that researchers are particularly concerned about the quality of the data they offer, as well as the clarity of the documentation attached to them. Should the researcher develop new practices and skills to enable others to replicate or reuse her/his data? Is it necessary for him to verify the integrity and interoperability of the data, to decide which data sets will be shared, considering in particular their potential for re-use and the costs generated, respecting the terms of the legal, economic, ethical and technical sharing (Carbou, 2017)?

In addition to depositing a dataset in a repository, the researcher should also provide the appropriate information about the dataset, which is known as metadata. Metadata describes the dataset and makes it possible for others to find, understand, and reuse the data. Besides standard information such as the creator and contributors of the dataset, the title, year of publication, and access rights, it can be necessary to add documentation such as codebooks, lab journals, informed consent forms and used software. There are various metadata standards for different disciplines, according to which the researcher must supply relevant additional information, necessary to make specific datasets comprehendible for the other users. For example, archaeological datasets require metadata about the spatial coverage area, while linguistics datasets require information about the language. Thus, as sharing is not an end in and of itself, the researcher should also develop new practices and skills to make the reusing of her/his data possible.

## 3. Methodology

The overall research design was to gather a sample of humanities artifacts found on Figshare and then to analyze some indicators like the number of articles by discipline, the evolution of contribution per year, and the number of views and downloads by affiliation to an institution (or lack thereof). In order to answer our research questions, we used the API provided by Figshare[2], which exposes several services to execute queries on its content. All communications are via https and all data is encoded in JSON format.

### *3.1. Figshare API*

To collect data, it is essential that our first step in exploring this API is to understand how content is structured in Figshare, so that we might pose relevant questions. To accomplish this, we made use of the documentation provided (Figshare, 2018). Any resource uploaded in Figshare is an "article"; the article may be public or private. A private article is inaccessible and is beyond the scope of the study as we are interested in open data only. We found that Figshare users do not declare a specialty domain but are asked to enter one or more categories for each downloaded article. We therefore based our queries on the category to deduce the field of the resource. We used queries applying a filter by "category", to limit

[2] https://api.Figshare.com/v2

ourselves to articles in the humanities. Next, we relied on the "publication date" of the articles to restrict our study to the period between January 2015 and July 2019. We also used the concept of a "group" to differentiate articles deposit by an institution account and those deposited by a researcher without a known institutional membership.

Figshare makes it possible to manage several types of data:

- Book
- Chapter
- Composition
- Conference contribution
- Data management plan
- Dataset
- Educational resource
- Event
- Figure
- Funding
- Journal contribution
- Media
- Monograph
- Online resource
- Peer review
- Performance
- Physical object
- Poster
- Preprint
- Presentation
- Report
- Software
- Standard
- Thesis
- Workflow

Figshare API makes it possible to run out queries by filtering the results by the "type of item". However, we did not apply a particular filter for the type because we wanted to analyze all types of data.

We encountered difficulties in determining the geographical origin of the data from research conducted in the area of human sciences present in Figshare. The API does not provide this information. To work around this problem we had to recover the ORCID of the author, and then query the ORCID database through its API. Unfortunately, the majority of authors on Figshare do not have expose their ORCID ID. We therefore could not know if this warehouse is used by researchers in some countries more than others; thus our identification of the country of origin of the resource and the design of Figshare and its API was limited.

The set of subcategories of Humanities does not necessarily correspond to disciplines, so we have selected only six, which were the most generic and which corresponded best to universally known disciplines. These were Linguistics, Law, History, Philosophy, Art and Literature.

To measure community engagement, it is possible to use the overall user acceptance taking into account the quantitative (number of users) and qualitative (best practices) aspects (Koureas et al., 2016). The number of users is not significant in the case of Figshare, because an author may not be active (having never deposited resources). In this case her/his resource was filed by a third party, and the author of the resource is a user of Figshare himself. For these reasons, we chose the number of articles per sub-category as an indicator of a community's overall membership. We began by comparing the number of open access articles in Figshare between January 2015 and May 2019 in the six sub-categories chosen.

The Figshare API also provides access to online metrics, such as the number of views and downloads; we have used these metrics to measure the visibility of articles by affiliation to an institution (or lack thereof).

### *3.2. Web application for data processing*

The next step of our study was the creation of a web application that allowed us to automatically execute various queries and to perform the necessary processing to restore the

information that interests us. The creation of such application[3] was very beneficial and allowed us to reproduce the results of this research at any time without having to store them in personal databases.

The application retrieved data from Figshare and performed the necessary calculations to generate tables with all expected statistics. The goal of having a reusable application was to eventually expand the scope of this study in the future, and to make any researcher able to reproduce the results at any time. Indeed, our web application allowed restoring the data for a category and a period of the researcher's choosing. For this study, our parameters were the category Humanities and the period between January 1, 2015 and July 1, 2019, but it was possible to change the scope of the study without additional effort to collect and process the data. Note that the application could also download the data generated in CSV format for use by other tools for analysis.

Figshare presents a set of categories on main navigation page[4] including the category Humanities, but when downloading an article, it allows authors to select additional sub-categories from a longer list. In our case study, we looked at 163 subcategories. The application allows restoring the information for all subcategories of Humanities and groups the results by subcategory; they are then displayed in a web page with a paging system to browse (Fig.1).

By pressing the Details link for a given sub-category, a list of results is displayed. Our web application makes it possible to browse all the articles of a given subcategory and to recover the following information for every article: the total number of articles by category for the time span, the number of articles by item type, the number of articles published each year for each sub-category and the number of articles with group ID and without group ID. The group ID identifier is not empty if the article is affiliated with an institution.

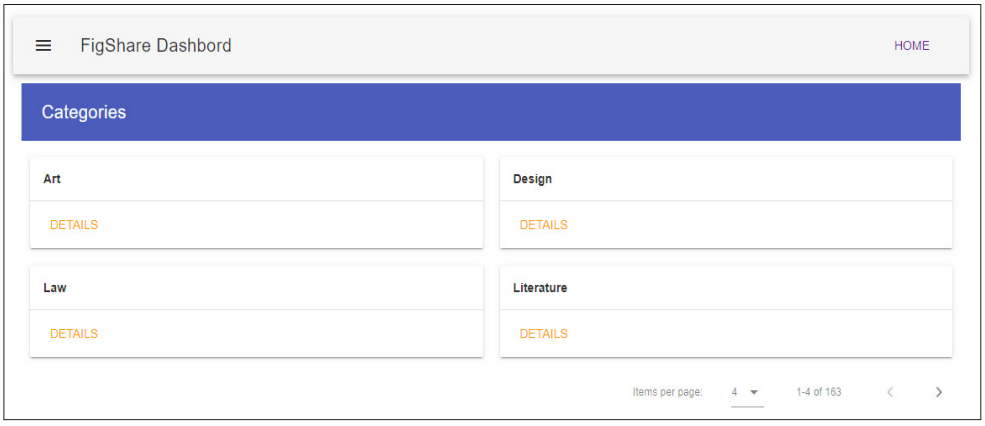


Fig.1. The list of results by subcategory

In addition, we were able to browse all the articles for a given sub-category and retrieve the number of views, downloads and sharing for every article. Although these metrics are

[3] The application is accessible at: https://vigorous-cori-9b6eb9.netlify.com/categories
[4] https://figshare.com/browse

widely used to indicate the impact of academic articles, they are to be considered with care (Bornmann, 2015).

Altmetrics refers to data sources, tools, and metrics (other than citations) that provide potentially relevant information on the impact of scientific outputs (e.g., the number of times a publication has been tweeted, shared on Facebook, or read in Mendeley). It also allows a broader interpretation of the concept of impact and to more diverse forms of impact analysis (Waltman & Costas, 2014). According to Taylor and Plume

> altmetrics hold great promise as a source of data, indicators and insights about online attention, usage and impact of published research outputs (Taylor & Plume, 2014).

Since Figshare tracks the views and downloads statistics for the published research data, these statistics indicators were used in this study as a source for altmetrics to measure the visibility of articles.

The data was imported to Microsoft Excel and prepared for quantitative analysis.

## 4. Results

### *4.1. Humanities participation in Figshare bydiscipline*

We've noticed that there is indeed a difference of participation between the sub-categories chosen. In fact, over the same period, the number of articles found in Figshare was 1361 for History, 1342 for Linguistics, 1280 for Law, 681 for Philosophy, 504 for Art and 404 for Literature (Fig. 2).

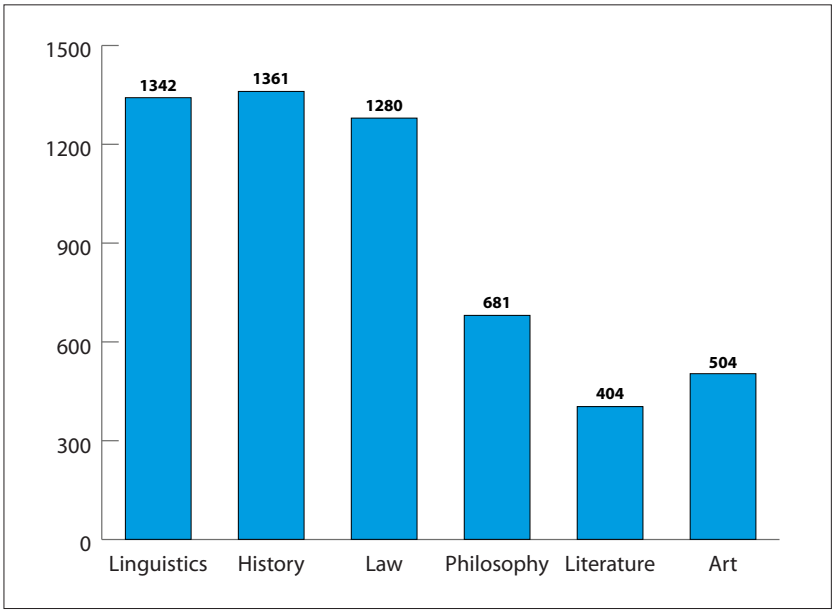


Fig. 2. The distribution of articles per discipline

### *4.2. Humanities participation evolution by field per year*

Next we wanted to show how this participation has evolved between 2015 and 2018. We have excluded the articles from 2019. As the year is not yet finished, the comparison will not be significant, and the indicator will not be relevant.

The participation curve grows at different rates depending on the discipline, for example, the number of articles in Linguistics was three times greater in 2018 than in 2015, while the number of articles in Philosophy was almost the same for four years. A relative increase was visible in 2018 for Law, History and Art (Fig. 3).

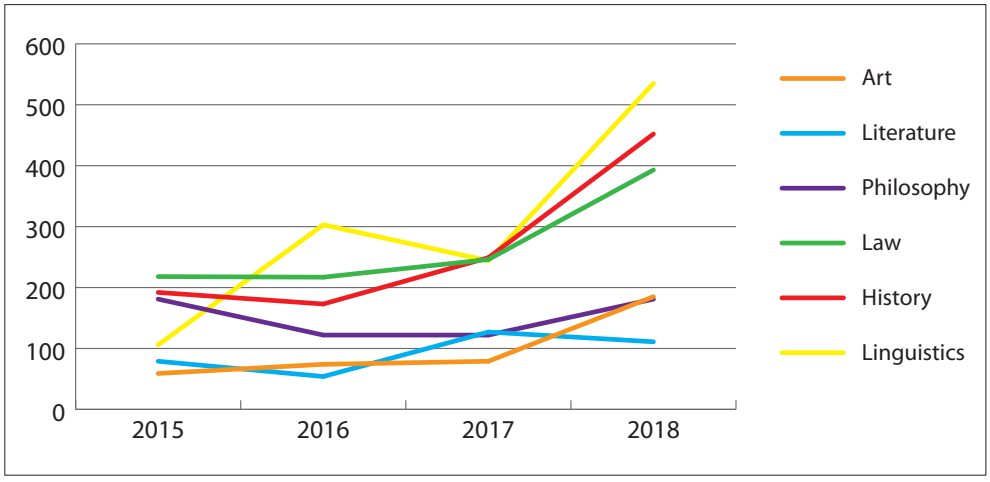


Fig. 3. The growth of the contributions across 4 years

### *4.3. Affiliated vs non-affiliated datasets*

Figshare offers a service for institutions that ensures that their data management requirements are made transparently for the researcher. This feature has attracted some universities' attention which have chosen Figshare for Institutions for management of their research data. For the six chosen fields, the majority of articles were not affiliated to any institution, and only 30% of articles had metadata records about that issue. But there was a small difference between the fields. 60%of articles in Philosophy were affiliated to an institution, followed by 53% for Art. The percentage of affiliated articles for the four other fields was lower: 35% for History, 33% for Literature, 22% for Law and only 12% for Linguistics (Fig. 4).

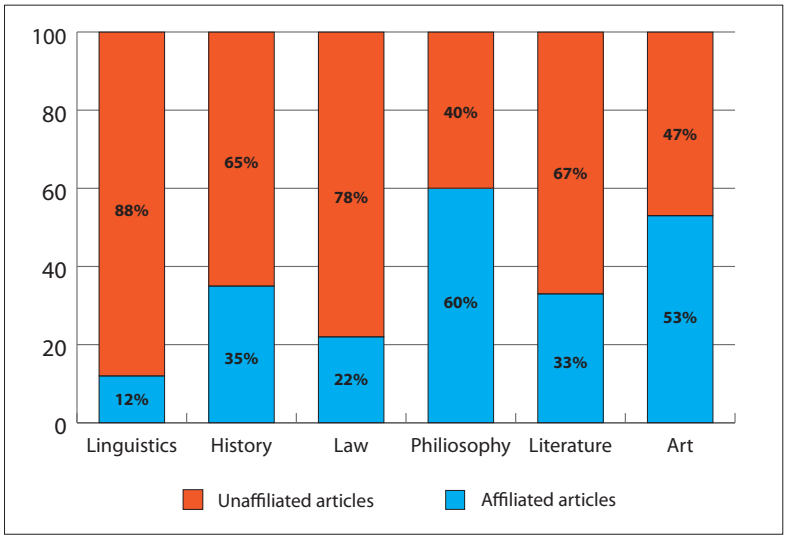


Fig. 4. Articles published with institutional affiliation

## *4.4. Data types by disciplines*

The results show that for the six fields, the most used data types were "Journal contribution", followed by the "Dataset" type which was in second position for almost all disciplines.

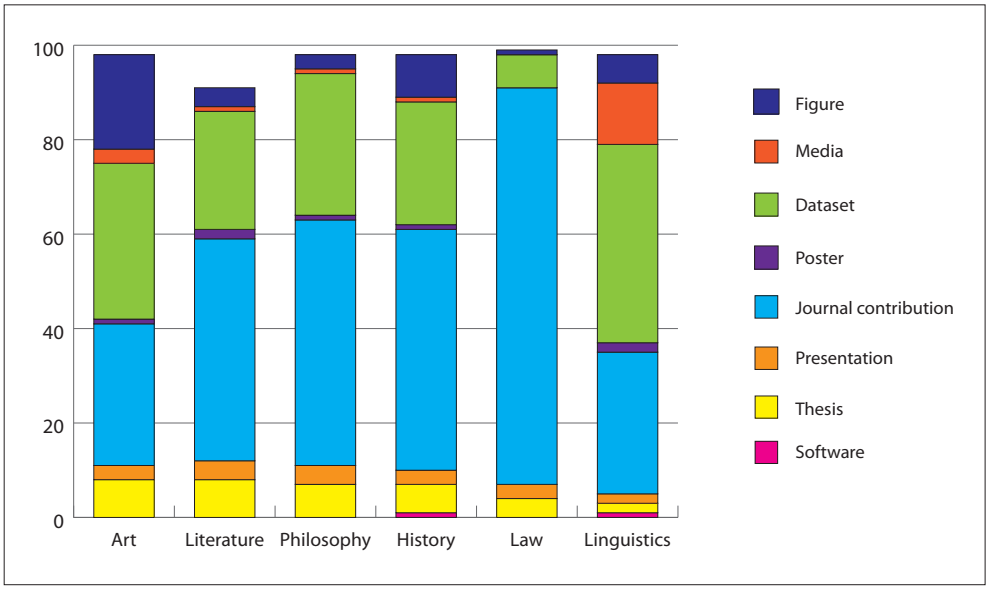


Fig. 5. Data types by disciplines

### *4.5. Visibility of articles*

Figshare provides access to online metrics, such as the number of views that we used for measuring the visibility of open data. The number of views was very uneven. We used the quartiles for each field to calculate the interquartile range, which is a measure of variability around the median. The quartile breaks down the data into quarters so that 25% of the measurements are less than the lower quartile, 50% are less than the mean, and 75% are less than the upper quartile (Tab. 2).

Tab. 2. Quartiles for number of viewed articles

| | Art | Law | History | Linguistics | Literature | Philosophy | All fields |
|---|---|---|---|---|---|---|---|
| **Min** | 0 | 0 | 0 | 0 | 0 | 2 | 0 |
| **Lower quartile** | 15 | 6 | 10 | 13 | 56.75 | 14 | 11 |
| **Median** | 139 | 15 | 30 | 97.5 | 154 | 99.5 | 52 |
| **Upper quartile** | 327 | 41.5 | 94.5 | 252 | 331.5 | 245.25 | 197 |
| **Max** | 7677 | 13 275 | 2093 | 8378 | 3916 | 2584 | 13 275 |

Overall, 50% of articles has been viewed less than 52 times and 75% less than 197 times. Nevertheless, the variability around the median changes from one discipline to another. For example, 75% of Law and History articles has been viewed less than 95 times. But in the fields of Philosophy, Literature, Linguistics and Art, 50% of articles has been viewed more than 95 times. Literature is the field with the more viewed open research data, with 50% more than 154 times and 25% more than 331 times.

But it seems that some articles from other disciplines have found an audience: the most viewed article is from Law with 13275 views (Ruiz, 2018), followed by a Linguistics article with 8378 (Styles, 2017), then an Art article with 7677 (Snider, 2015).

Tab. 3. Quartiles for number of downloaded articles

| | Literature | Art | Law | Linguistics | Philosophy | History | All fields |
|---|---|---|---|---|---|---|---|
| **Lower quartile** | 4 | 2 | 0.25 | 3 | 2 | 1 | 1 |
| **Median** | 23 | 2 | 3 | 18 | 15 | 5 | 10 |
| **Upper quartile** | 72.5 | 2 | 12 | 65 | 48.5 | 13 | 44.25 |
| **Max** | 388 102 | 2 | 5510 | 40 927 | 11 434 | 494 | 388 102 |

Concerning the number of downloads (Tab 3.), more than 75% of articles has been downloaded less than 45 times. There were no differences concerning the discipline. But some of the uploaded resources have generated much interest (more than 388 000 downloads for a Literature article).

We also compared the mean for the views of articles by publication date, which was very low for both affiliated and unaffiliated articles published after 2018. But the mean view of affiliated articles was generally higher than the mean view of unaffiliated articles for articles published in the same period (Fig. 6).

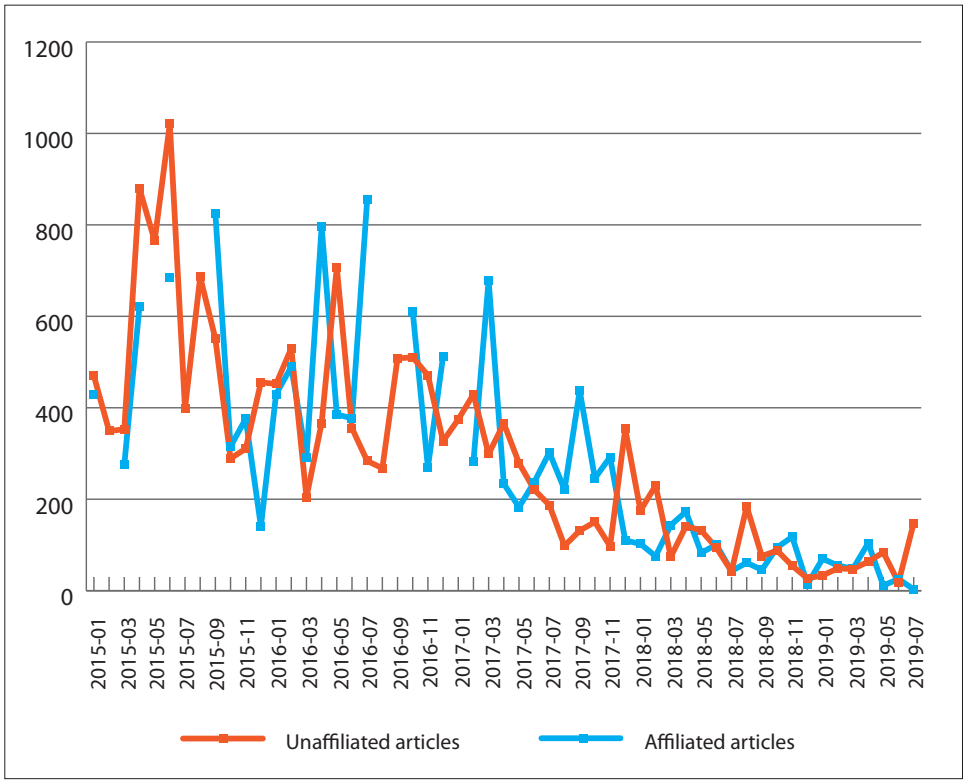


Fig. 6. Mean of views by publication date and by affiliation

## 5. Interpretation

### *5.1. Community engagement*

The current study found that there is a difference in participation in depositing research data on Figshare across the disciplines in Humanities. The number of articles in Linguistics, Law and History was twice as large as in Philosophy, Art, and Literature. We also noticed that the growth curve increased at different rates according to the disciplines. The disciplines with a higher number of affiliated articles (Philosophy, Art) had more stable evolution curve. However, while an almost identical number of articles had been uploaded on Figshare each year, the disciplines with a higher number of affiliated articles had consistently very low contribution. The contribution in disciplines with a large number of unaffiliated articles (Linguistics, Law) increased quickly from year to year and these were the disciplines with the highest number of articles published. For History and Literature, the distribution of affiliated and unaffiliated articles was identical, but overall, the number of contributions was higher for History, and their respective growth curves changed in very different ways.

The results showed that while the Humanities community engagement for Figshare varied across the disciplines, overall, it remained very weak. Indeed, the total number of datasets

published in the six disciplines was very low (1361 articles for History) for almost five years. The community engagement was not related to affiliation to institution. The presence of humanities data on Figshare was not always framed by institutional membership policies. Humanities researchers were using use Figshare to for opening their research data even if the institution they work for is was not affiliated to this data warehouse.

Previous studies have shown that the management of research data is still not common practise among researchers in the humanities (Chowdhury et al., 2018; Neylon, 2017; Prost & Schöpfel, 2015; Stuart et al., 2018). In humanities, individual research can rarely generate shareable data, but digitization initiatives create shareable resources such as photographs of works of art, historical documents or cultural artefacts (Gorman, 2007).This explains why History is at the top of the list in terms of the number of articles published on Figshare. Therefore, it may be interesting to compare this results with other data warehouses like Zenodo or Dryad to study if this phenomenon is typical to Figshare or is it common in humanities practices.

Technically, Figshare allows its users to download multiple data formats, so it would seem that it overcame the challenge of managing the heterogeneity of data in human sciences (European Commission, 2015). However, we noticed that the descriptive categories used for the data types are more or less common in the six disciplines. Indeed, the most used item type were "journal contribution", then "dataset" followed by "figures", "presentations", and "media". Nevertheless, it is interesting to note that journal contribution which is at the top of the list for all the disciplines corresponds to any type of content officially published in a scientific journal, following a peer review process. This implies that Figshare is used to satisfy the requirements of funding agencies or publishers for opening research data. The majority of data published on Figshare was described by users as journal contribution, ignoring the other types even if they are more precise about the real data format.

### *5.2. Visibility analysis by disciplineand by affiliation or not to an institution*

More than 75% of articles was viewed less than 197 times and downloaded less than 45 times. This result means that the use of Figshare is far from being enough to insure the visibility and the reuse of data.

Even if some of the uploaded resources have generated much interest, the majority of research data in humanities was not viewed, and even less was downloaded. This calls into question the quality of the metadata. To be reusable, a data must first be findable. A recent study has shown that a large majority of researchers never or rarely use a metadata standards for describing research datasets (Chowdhury et al., 2018).

The program Figshare for Institutions ensures that their data management requirements are met transparently for the researcher. The system assists researchers in data management by providing conservation and metadata options for their research data. The researchers upload files and attach the appropriate metadata. The system can be configured to allow researchers to share files without intervention or route new downloads to an administrator for review before they are posted. Administrative control of file downloads can be done at the level of an institution or group, such as a department (Reed, 2016).

## 6. Conclusion and perspectives

This study allowed us to highlight that, even if Figshare is part of logic of openness, the design of each tool determines its use, and that the exploitation of data remains closely tied to the warehouse policy and the services it makes available. We encountered limitations with our attempts to determine the country of the researcher or institution responsible for depositing the dataset in terms of the design of Figshare research features and its API. The identification of the resource's country of origin was not possible.

Findings of this study support those of several abovementioned studies In the arts and humanities, it seems that sharing data globally and reusing it is not yet a common practice among the researchers (Chowdhury et al., 2018; Neylon, 2017; Prost & Schöpfel, 2015; Stuart et al., 2018). The humanities community engagement on Figshare remains low. Even when research data was published, 50% is rarely viewed and less downloaded.

The results of our study provided some specific details about the affiliated articles: 70% of humanities research data on Figshare was not affiliated with any institution, but for the 30% of affiliated data, a higher average of visibility was recorded. This study has shown that affiliated articles were more viewed than unaffiliated ones published in the same period.

Thus, it seems that the establishment of the warehouses, is certainly a step in the right direction, but is far from being a sufficient mean to ensure that the potential benefits will be fully exploited. We believe that the supply of these warehouses must become a systematic part of the process of producing scientific knowledge. One way to overcome the challenges we discussed could be the development of an institutional policies that points the recommended warehouse, guarantees copyright, organizes the data, and controls the quality of the metadata. In this study, we acknowledged that data published under the control of a data professional has more chance to be found than data published individually by a researcher.

We think that there is a gap between the one who transfers the content (infrastructure builders) and the one who creates the value (researchers). Defining the institutional policy at the university level more precisely could be a good way of bridging the gap and guarantee the exploitation of this data. The management of research data requires many skills, and the researcher cannot be expected to have them all. Thus, it is important for the researcher to have tools and adequate services, which features are flexible enough to suit her/his needs – an entity able to offer services, support users and meet new needs as they arise, to improve the entire system continuously. Achieving the FAIR goals is not just a matter of a specialist or researcher but rather a shared representation of an ecosystem (Lehmans, 2017). Thus, it is crucial to build on and unleash the potential of collaboration between researchers and specialists, guaranteeing quality, mastering the sharing modalities and embodying the global vision of the ecosystem.

The researcher be must absolutely accompanied and so that she/he is not responsible for the management of the technical and legal problems related to the research data. Specialized entities must deal with these issues in a manner that is transparent to the researcher. We believe that information science professionals have a major role to play at this level. This discipline has investigated some of the most challenging issues related to the use of large-scale information resources, including the organization, access, management, and storage of research products in all their formats and encodings. They could support researchers in order to accomplish this task. Their participation can help increase the efficiency and

ensure the inclusion of default research data management in scientific production. This will ensure both the value of the data and the quality of the metadata.

To conclude, researchers do not get sufficient credit and other rewards for producing and sharing data. It is still the publication in a peer-reviewed journal that matters. Making data sets available should also be rewarded as an important scientific output. Journals adopting a data availability policy and data journals can be an important instrument to change this situation.

## References

Ali-Khan, S. E., Jean, A., MacDonald, E., Gold, E. R. (2018). Defining Success in Open Science. *MNI Open Research*[online], 2, 2, https://doi.org/10.12688/mniopenres.12780.2

Borgman, C. (2012). The Conundrum of Sharing Research Data. *Journal of the American Society for Information Science and Technology*, 6(63), 1059–1078.

Borgman, C. L., Darch, P. T., Sands, A. E., Wallis, J. C., Traweek, S. (2014). The Ups and Downs of Knowledge Infrastructures in Science: Implications for Data Management. In: G. Buchanan (ed.). *IEEE/ACM Joint Conference on Digital Libraries*[online]. New York: IEEE Press Piscataway, https://doi.org/10.1109/JCDL.2014.6970177

Bornmann, L. (2015). Usefulness of Altmetrics for Measuring the Broader Impact of Research: A Case Study Using Data from PLOS and F1000Prime. *Aslib Journal of Information Management*[online], 67(3), 305–319, https://doi.org/10.1108/AJIM-09-2014-0115

Carbou, G. (2017). Analyser les textes à l'ère des humanités numériques. *Les Cahiers du numerique*, 13(3), 91–114.

Chowdhury, G., Ünal, Y., Kurbanoglu, S., Boustany, J., Walton, G. (2018, October). *Research Data Management and Data Sharing Behaviour of University Researchers.* [online], [13.08.2019], http://nrl.northumbria.ac.uk/34963/

European Commission. (2015). *Validating the Results of the Public Consultation on Science 2.0: Science in Transition* [online], [13.08.2019], http://www.eesc.europa.eu/resources/docs/validation-of-the-results-of-the-public-consultation-on-science-20.pdf

Figshare (2018). *Figshare API User Documentation* (2018, 7 September) [online], https://doi.org/10.6084/m9.figshare.4880372.v2

Girard, C. (2017). Les mécanismes de centralisation des données de la recherche. Étendre l'accès libre à l'hébergement libre. *Revue française des sciences de l'information et de la communication* [online], (11), https://doi.org/10.4000/rfsic.3255

Gorman, G. E. (2007). The DeLorean or the Mini? *Online Information Review* [online], 6(31), https://doi.org/10.1108/14684520710841739

HLEG (2010, October). High Level Expert Group on Scientific Data. Riding the Wave – How Europe Can Gain from the Rising Tide of Scientific Data [online], [20.05.2019], http://www.eurosfaire.prd.fr/7pc/bibliotheque/consulter.php?id=2043

Jacobson, M., Larrousse, N., Massol, M. (2014, 7 July). *La question de l'archivage des données de la recherche en SHS (Sciences Humaines et Sociales)*[online]. Archives et données de la recherche (ICA/SUV 2014), [13.08.2019], https://halshs.archives-ouvertes.fr/halshs-01025106/document

Konkiel, S., Scherer, D. (2013). New Opportunities for Repositories in the Age of Altmetrics. *Bulletin of the American Society for Information Science and Technology* [online], 39(4), 22–26, https://doi.org/10.1002/bult.2013.1720390408

Koureas, D., Arvanitidis, C., Belbin, L. et al. (2016). Community Engagement: The "Last Mile" challenge for European Research e-Infrastructures. *Research Ideas and Outcomes*[online], 2, e9933, https://doi.org/10.3897/rio.2.e9933

Lehmans, A. (2017). Données ouvertes et redéfinition de la culture de l'information dans les organisations. *Communication Organisation*, 51(1), 15–26.

Neylon, C. (2017, 16 June). *As a researcher... I'm a bit bloody fed up with Data Management* [online]. Science in the Open [28.04.2019], https://cameronneylon.net/blog/as-a-researcher-im-a-bit-bloody-fed-up-with-data-management/

Open Access (n. d.). *Open Access – H2020 Online Manual* [online], [20.07.2019], http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/open-access_en.htm

Prost, H., Schöpfel, J. (2015). *Les données de la recherche en SHS. Une enquête à l'Université de Lille 3* [online], [13.08.2019], http://hal.univ-lille3.fr/hal-01198379/document

Reed, R. B. (2016). Figshare for Institutions. *Journal of the Medical Library Association* [online], 104(4), 376–378, https://doi.org/10.3163/1536-5050.104.4.031

Research Infrastructures (n.d.). *Research Infrastructures, Including e-Infrastructures*[online]. Horizon 2020 sections [28.04.2019], https://ec.europa.eu/programmes/horizon2020/en/h2020-section/research-infrastructures-including-e-infrastructures

Ruiz, R. M. (2018, 28 February). *GDPR Tool for Handling Data-Subject Rights & Requests* [online]. Figshare, https://doi.org/10.6084/m9.figshare.5853180.v2

Schöpfel, J. (2018). *Towards a Data Culture in Social Sciences and Humanities* [online], [13.08.2019], https://hal.archives-ouvertes.fr/hal-01846849

Snider, L. (2015, 21 June). *Geography 101 – Periodos Ges – Hydrographical Iconography in Quire 13 of Beinecke MS408, Folios 75 through 84* [online]. Figshare, https://doi.org/10.6084/m9.figshare.1309949.v4

Stuart, D., Baynes, G., Hrynaszkiewicz, I., Allin, K., Penny, D., Lucraft, M., Astell, M. (2018, 21 March). *Whitepaper: Practical Challenges for Researchers in Data Sharing* [online]. Figshare, https://doi.org/10.6084/m9.figshare.5975011.v1

Styles, S. (2017, 10 July). *The Commonest Speech Sounds: Prevalence Rates for Phonemes of the World* [online]. Figshare, https://doi.org/10.6084/m9.figshare.5047747.v3

Swan, A., Brown, S. (2008). *To Share or Not to Share: Publication and Quality Assurance of Research Data Outputs. A Report Commissioned by the Research Information Network* [online]. University of Southampton Institutional Repository [13.08.2019], https://eprints.soton.ac.uk/266742/

Taylor, M., Plume. (2014). Party Papers or Policy Discussions: An Examination of Highly Shared Papers Using Altmetric Data. *Research Trends* [online], 36. [21.07.2019], https://www.researchtrends.com/issue-36-march-2014/highly-shared-papers/

Waltman, L., Costas, R. (2014). F1000 Recommendations as a Potential New Data Source for Research Evaluation: A Comparison With Citations. *Journal of the Association for Information Science and Technology* [online], 65(3), 433–445, https://doi.org/10.1002/asi.23040

Weber, N. M. (2013). The Relevance of Research Data Sharing and Reuse Studies. *Bulletin of the American Society for Information Science and Technology* [online], 39(6), 23–26, https://doi.org/10.1002/bult.2013.1720390609

# Eksploracja danych badawczych z zakresu humanistyki w serwisie Figshare

**Abstrakt**

**Cel/Teza:** Celem artykułu jest zbadanie obecności danych badawczych z zakresu humanistyki w serwisie Figshare oraz próba odpowiedzi na pytanie, czy samo opublikowanie danych badawczych przekłada się na ich wyszukiwalność.
**Koncepcja/Metody badań:** Projekt badawczy zakładał zidentyfikowanie zbioru cyfrowych artefaktów z obszaru humanistyki w serwisie Figshare za pośrednictwem jego API oraz analizę ilościową w oparciu o zmienne: liczba artykułów z podziałem na dyscyplinę, zmiany ilościowe w ujęciu chronologicznym, liczba wyświetleń i pobrań artykułów z uwzględnieniem informacji o afiliacji zbioru danych badawczych.
**Wyniki i wnioski:** Zaangażowanie społeczności humanistów w deponowanie danych badawczych w serwisie Figshare jest różne zależnie od dyscypliny. Nawet jeśli dane badawcze są tam publikowane, to dla 50% z nich odnotowano niską liczbę wyświetleń. Oznacza to, że sama otwartość danych badawczych nie jest wystarczająca dla ich widoczności w sieci i nie zapewnia ponownego użycia. Ostateczny wniosek z badań wskazuje na to, że informacja o afiliacji z daną instytucją badawczą dołączona do artykułów zwiększa szanse na ich widoczność w serwisie.
**Ograniczenia badań:** Jednym z ograniczeń badań jest fakt, że analizie zostały poddane tylko artykuły opublikowane w serwisie Figshare, co uniemożliwia podjęcie badań nad przyczynami małej obecności danych badawczych z obszaru humanistyki. Istnieje zatem potrzeba przeprowadzenia podobnych badań w odniesieniu do innych platform publikowania danych badawczych. Uzyskane rezultaty są zbieżne z wynikami podobnych badań nad praktykami zarządzania danymi badawczymi.
**Oryginalność/Wartość poznawcza:** Według wiedzy autorki, są to pierwsze badania nad obecnością danych badawczych z obszaru humanistyki w repozytorium danych badawczych Figshare.

**Słowa kluczowe**

Dane badawcze. Figshare. Nauki humanistyczne. Otwarte dane.

*SAMIA TAKHTOUKH is a doctoral student at GERiiCO Laboratory, University of Lille (France). She holds a Computer Sciences engineer diploma from ENSIAS School in Morocco and a Research Master's degree in Library and Information Science from University of Lille in 2016. Her area of investigation is humanities research infrastructures.*

*Contact to the author:*
*takhtoukh.samia7@gmail.com*
*University of Lille*
*Laboratoire GERiiCO*
*B.P. 60149*
*59 653 Villeneuve d'Ascq Cedex, France*

# Wskazówki dla autorów

Redakcja *Zagadnień Informacji Naukowej – Studiów Informacyjnych* 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 dobieranych spośród specjalistów problematyki w nim poruszanej. Każda recenzja zawiera jednoznaczne wskazanie czy tekst rekomendowany jest do publikacji w *Zagadnieniach Informacji 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**

## 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, identyfikator ORCID)**
- **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 limatations** *(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 Informacji Naukowej – Studiów Informacyjnych* 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 opatrzony 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 6th (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 6th. 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 CZASOPIŚMIE

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 Informacji 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 CZASOPIŚMIE 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

Bartalesi, V., Meghini, C. (2016). Using an Ontology for Representing the Knowledge on Literary Texts: The Dante Alighieri Case Study. *Semantic Web* [online], 8(3), 385–394, http://doi.org/10.3233/SW-150198

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

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

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**

## 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 and the ORCID numbers,**
- **the 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 Informacji 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

Bartalesi, V., Meghini, C. (2016). Using an Ontology for Representing the Knowledge on Literary Texts: The Dante Alighieri Case Study. *Semantic Web* [online], 8(3), 385–394, http://doi.org/10.3233/SW-150198

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

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. 22 608 28 24

Wydawnictwo Naukowe i Edukacyjne SBP – Warszawa 2019
Ark. wyd. 5,5.
Skład i łamanie: Justyna Grzymała-Łuszcz


WYDAWNICTWO
SBP