

Conference on Grey Literature and Repositories: proceedings 2014: the Value of Grey Literature in Repositories [online]. Prague: National Library of Technology, 2014 [cit. 2014-12-19]. Available from: <<http://nrgl.techlib.cz/index.php/Proceedings>>. ISSN 2336-5021.

Enhanced publications in V4 countries

Edith Görögh, the University of Debrecen

Edyta Kędzierska, Natalia Kavalchuk, Jolanta Stepniak, the Warsaw University of Technology

Jozef Dzivak, the Slovak University of Technology, Chemical Library

Petra Pejšová, Hana Vyčítalová, the National Library of Technology

Abstract

The paper describes the project Enhancing scholarly communication: National initiatives to manage research data in V4 countries. The main goal of the project is a survey about state of research data management, repository contents, services and archiving policies in colleges, universities and research institutions on national levels. The results of the survey will be presented by representatives from Hungary, Czech Republic, Slovak Republic and Poland.

Keywords

Enhanced Publications, Open Access, Research, Research Data, Visegrad Funds, Digital Repositories, Scholarly Communication

This text is a collaborative work by authors participating in the “Enhancing scholarly communication: national initiatives to manage research data in the V4 countries” project.

This paper contains five parts:

An introduction to Open Access in scholarly communication, research data and projects

The report from the survey of Hungarian universities

The report from the Survey of Polish Scientific and Research-Development Units

The report from the survey of Slovak universities and scholarly institutions

The report from the survey of Czech research institutions and universities

This paper is licensed under the Creative Commons licence: CC-BY-SA-4.0 (<http://creativecommons.org/licenses/by-sa/4.0/>), via <http://www.nusl.cz/ntk/nusl-175813>.

An introduction to Open Access in scholarly communication, research data and projects

Edith Görögh,

editg@lib.unideb.hu

The University of Debrecen

Scholarly communication involves the transmission of knowledge in which information is transferred through personal interaction, by e-mail, via submission to a database, the creation of a video or through the formal writing and printing process. The expansion of the internet is slowly overwriting the traditional connections within the communication process.

Scholarly communication is changing and the structure of the process, the dynamics of the participants and the methods of interaction are altering. Knowledge is being shared instantaneously (Trends in Scholarly Communication, 2014). Blogs, personal and institutional websites, videoconferences and online meetings help researchers to acquire information about new research and new developments in their fields quickly. An enhanced dialogue within the research community not only contributes to fast communication among researchers leading to research collaboration, but also provides an opportunity to scientists to easily address wider audiences and communicate with leaders in their fields.

Visibility is an essential element of scholarly work. Visibility has long been the defining element of careers in academia and the popular motto of “publish or perish” urges researchers to make their work visible. The more a researcher publishes, the more he/she is recognized within the given research communities. However, academia has become larger and more demographically diverse. A new generation of researchers is employing different channels of dissemination than traditional printed formats. The networked scholar is born and is taking an active part in the dissemination of information about his/her work. The interactive presence of researchers in social media enhances the visibility of their work as well.

Visibility has also become a key issue in funders’ requirements during the past few decades. In most cases funders require that the research results which they finance should be demonstrated to all shareholders in an open and transparent manner. Such requirements are usually met not only by implementing the visibility and branding policies of the given funder, but also by providing Open Access to the research results.

Open Access in scholarly communication

Open Access is the practice of providing free on-line access to and free re-use of scientific information. Authoritative definitions of Open Access can be found in key political declarations on this subject. These definitions describe Open Access as including not only the right to read, download and print information, but also the right to copy, distribute, search, link, crawl and mine it.

The European Union pays special attention to publicly-funded scientific research. Its main objective is to optimize the impact of these scientific endeavors, both at a European level (FP7, Horizon 2020) and at a Member State level. The EU has chosen Open Access to disseminate the research results more broadly and faster. However, Open Access requirements are based on balanced support to both 'Green Open Access' (immediate or delayed Open Access that is

provided through self-archiving) and 'Gold Open Access' (immediate Open Access that is provided by a publisher). The main objective of the dissemination of scientific output via Open Access is to enhance economic performance in the EU and improve the capacity to compete through knowledge.

The recent EU Research Framework Programs, namely FP7 and Horizon 2020, have concentrated on the Open Access to research data which refers to the right to access and re-use digital research data including accessing, mining, exploiting, reproducing and disseminating data free of charge for the user.

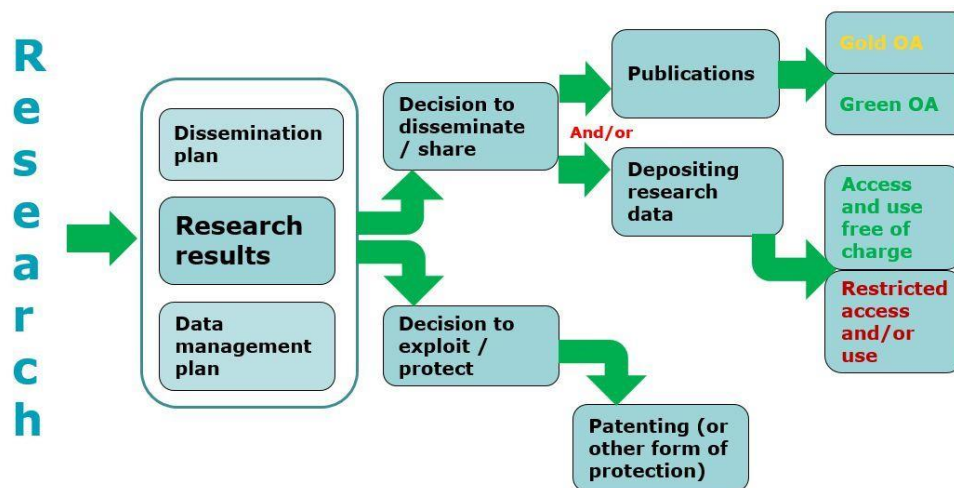


Figure 1: Open Access to scientific publication and research data in the wider context of dissemination and exploitation ¹

Open Access has divided opinion on the strengths and weaknesses of publishing and providing free access to scientific results and research data. According to Stevan Harnad, PhD, Professor of Cognitive Science at Southampton University, the UK, and a chief proponent of the Open Access Initiative, "anything that blocks access to research findings is ... going against the interests of research, researchers, their employers, their funders and the tax-payers that fund the funders." (Harnad, 2003)

Research data publication

As well as the Open Access publishing of scientific results, there is also a growing number of initiatives concerning data publication. The pilot on research data in Horizon2020 and other funding requirements signifies the strengthening advocacy of more open research in Europe. The Open Access movement and open data publication projects imply changing research and publishing practices.

Data publication is the process of making information, particularly data generated from research, available to all. Data archiving is the long term storage of such data and methods. In science, data publishing and archiving is important for the preservation of scientific information for future research.

¹ Guidelines on Open Access to Scientific Information and Research Data in Horizont 2020: Version 1.0 [online]. Brussel, 2013 [cit. 2014-10-16]. Available on http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/Open_Access_pilot/h2020-hi-Open_Access-pilot-guide_en.pdf. Guidelines. European Commission.

Open data is the building block of open knowledge. Open knowledge is what open data becomes when it's useful, usable and used (Ayrís, 2012). The key features of openness include (1) availability and access referring to the availability of data as a whole in a convenient and modifiable form, (2) data should be available to be re-used and redistributed and (3) everyone should be able to use it.

The IVF project

Developments in the scientific communications landscape and the advance of the green way of Open Access publishing have led researchers to upload and archive not only their scholarly output, but also the research data underlying their publications. The International Visegrad Fund (IVF), which focusses on the Visegrad countries (Hungary, the Czech Republic, Slovakia and Poland), follows these changes in scholarly communication in this region and attempts to identify the milestones in the process through the projects it sponsors. Under the coordination of the University of Debrecen, four institutions from the Visegrad region have started cooperation to share experiences and map the national situation of research data management in a project entitled "Enhancing scholarly communication: national initiatives to manage research data in the V4 countries". The participating members are the National Library of Technology in Prague (CZ), the Chemical Library at the Faculty of Chemical and Food Technology of the Slovak University of Technology and the Warsaw University of Technology; Biblioteka Glowna (PL). The project leader is the University of Debrecen (HU).

There is a growing discourse about research data management: handling scientific data and linking it to related publications. European programs, such as OpenAIREplus, focus on enhanced publications, and the improved visibility of research results in scholarly publishing. The aim of such Pan-European programs is to mark the directions of development in scholarly communication and to join international forces to implement the changes. The European research and innovation program defined in Horizon 2020 encourages national policy initiatives to improve access to and preserve scientific information. However, there are national differences in the technical and financial means to contribute to these programs.

The project addresses the main issues of managing and archiving research data, discusses the role of libraries in handling enhanced publications and examines the national repository landscape and the current issues facing institutional repositories and data management.

The project has three primary objectives. First, it serves as a national initiative to join library efforts to manage research data including their collection and archiving and linking them to publications. Secondly, it reviews the national repository landscape in all V4 countries, and discusses the current issues facing repositories and data management in this region. Thirdly, it identifies Open Access usage in IVF countries and national advantages and disadvantages.

The survey

The project's output represents a gradual data collection process and systematic evaluation of the results. First, national surveys were administered to the repositories. The results of the questionnaires were summarized, evaluated and published electronically on institutional websites. The results may serve as preparation for national system plans for research data management.

The NTK initially prepared a survey in 2013 and tested it on research organizations in the Czech Republic. This survey subsequently served as the basis for the joint survey of all V4 institutions in 2014. Due to the fact that the other three participating institutions did not have prior

experience in open research data, we felt it would be wise to include some questions on Open Access publishing in general. This way, an attitude on Open Access in general could be estimated alongside the more detailed examination of data management practices.

The aim of the survey was to find out what research data is produced and archived by research institutions and public and state colleges and universities and to provide a basic overview of Open Access publishing. In addition, the survey focused on the manner of data storage and archiving, further use of the data, and most importantly, whether data is linked to research publications and whether we can consequently talk about enhanced publications in the Visegrad countries. The national analyses of the survey are combined in a publication which is openly disseminated electronically via the project's website.

There is a tangible reluctance among the stakeholders (policymakers, researchers) to actively contribute to the current advances in scholarly communication in the Visegrad region. The project provides an overview of the national initiatives in order to give an insight into national developments and to join insular repository programs and research data management projects. The success of the project will be measured by the growing inclination among researchers to upload publications and related research data and by the strengthening interest in policymakers to provide a centralized stance on Open Access publishing. The planned activities will move us forward to become active members of the global repository and data management discourse and help to improve the institutions' visibility in the international academic landscape.

The report from the survey of Hungarian universities

Edith Görögh,

editg@lib.unideb.hu

The University of Debrecen

Introduction

In Hungary, the discourse on research data management is fairly new in scholarly communication. It is usually discussed within the context of Open Access publishing as an additional requirement of European funding agencies and programs, such as Horizon2020. As Open Access gains ground in this region, a growing number of institutions have included the main principles of Open Access in their publishing policies or even introduced Open Access mandates into their operational procedures. As well as funders' requirements, a growing number of publishers, especially Open Access ones, demand free accessibility to research data in connection with the research results they publish. Authors are urged to think about storing data in a location where free access can be ensured to them, and in formats which are eligible and compatible for readers.

Survey description

The survey was distributed in two different channels. Both channels ensured communication among researchers and administrators of Hungarian higher education and research institutions. First, it was sent out to HUNOR members. The HUNOR (Hungarian Open Repositories) consortium was established in 2008 by the libraries of Hungarian higher education institutions and the Library of the Hungarian Academy of Sciences to advance national Open Access practices. The members of HUNOR are dedicated to promoting Hungarian research both nationally and internationally and to achieving the effective dissemination of scientific output through the implementation of a national infrastructure of Open Access repositories. Since libraries have central communication roles within institutions, the distribution of the survey through HUNOR libraries was meant to ensure the comprehensive cover of the circulation. We requested the libraries to send the survey to the heads of departments and to researchers, as well. Questioning solely the leaders of departments or research fields would not have given us adequate information about the research data management practices, since it is the researchers who actually deal with the data on a regular basis.

The other channel in which we tried to reach as many researchers as possible was through the Database of Hungarian Scientific Documents. This national archiving initiative operates a network of administrators at all higher education institutions. The administrators are in direct contact with the departments and research units within the institutions. We sent out a letter to the administrators at the University of Debrecen with a request to distribute the survey within the research communities they are responsible for.

We received 70 answers, most of which arrived from higher education institutions, the majority from the University of Debrecen. Other universities participating in the survey were the Corvinus University in Budapest, the University of Economics in Budapest, the University of Pécs, the University of Szeged and the Miskolc University. The viewpoints of research

institutions were also represented in the responses from the Hungarian Academy of Science and ALÖKI, the Applied Ecology Research Institution.

Questions about Open Access publishing

The first section of the survey focused on the Open Access policies of the institutions. The responses indicated that the majority of researchers use the repositories at their institutions to archive their research articles. The visibility of and access to scientific output is not only regulated by the institutions themselves, but also requested by the funding bodies outside the institutions. In the past decades, higher education institutions placed more emphasis on the management of their scientific output. Projects funded by EU organizations require the accessibility of the research results to which they have contributed finances. Furthermore, the prestige of a research institution, which usually reflects the ability of the institution to attract new students, new projects, more funds and to produce new knowledge, is often based on the number of publications and related citations the researchers accumulate and make accessible through the institutional repository. The statistical analyses which universities use to attract more funders are based on the content of their repositories.

Higher education institutions are inclined to regulate the archiving process of their researchers' scientific output by means of institutional mandates ensuring the entry of the output into the repositories. Responses indicate the growing inclination at universities to issue an institutional mandate for archiving.

Open Access has been long introduced into scientific discussions in Hungary. The University Library of Debrecen has been organizing workshops and presentations on Open Access in order to familiarize researchers and staff about the new international trends in scholarly communication. Although researchers in general are reluctant to embrace Open Access publishing in practice, they are knowledgeable about its benefits. The introduction of institutional mandates and the requirements of funding bodies urge researchers to comply with Open Access policies and to begin to be more consciously involved in the publishing process. As Open Access publishing becomes part of their research procedures, scientists begin to experience the effects of Open Access publishing more in practice.

The majority of respondents thought that the most significant advantage of Open Access publishing or storage was the opportunity to reach wider audiences. Since the goal of universities is to create and disseminate knowledge, researchers consider the principle of transferring information to wide audiences to be essential for their work. Another major issue in a researcher's career is the visibility of his/her work. The driving force of the "publish or perish" principle compels researchers to increase the citing of their work. Therefore, the advantage of increased citation in Open Access publishing is valued highly among researchers. The visibility of an institution is closely connected to the reputation of the institution or the researcher: the more people know and value the work, the higher prestige it receives. Institutions with high prestige attract more students to study there and more researchers to work there. The visibility and reputation of a researcher's work may result in successful grant applications, new research cooperation, etc.

Although Open Access has been introduced to the scholarly dialogue, it has not proved to be a vital publishing choice for researchers. The low prestige of Open Access publishing often derives from two main features of Open Access journals: it is relatively new in the market and it costs money for the author to publish. A common concern is that Open Access journals are of lower quality because they involve an article processing fee. The real reason for the low

prestige lies in the fact that Open Access journals are newer and younger than the subscription journals. All new journals need excellent publishing material to generate prestige. They need to develop a reputation for quality and the authors need to realize that by submitting excellent research to these new journals, they contribute to the prestige of Open Access.

Another problematic question based on the responses is related to the copyright issues of Open Access documents. The Open Access movement has triggered a debate on copyright in the scholarly communication system. Stakeholders in the system – publishers, academic institutes and libraries and authors can adopt opposing positions in this debate. New copyright models have been introduced with the emergence of Open Access journal publishing. These copyright models are opposed to the model used by traditional academic journals in which the copyright is transferred from the author to the journal publisher. The new models offer a wide range of choices for authors who should be informed about the advantages and disadvantages of these new models. Researchers have shown that academic writers publishing in Open Access journals and gaining knowledge on copyright choices appear to be no longer satisfied with assigning the copyright to publishers.

Researchers consider the issue of copyright to be problematic and unsolved, because they either do not possess a well-rounded view on the question or they do not agree with the rights offered by Open Access publishing. The right to reuse information is essential in the Open Access models, since it gives free online access and permission to use the information for any responsible purpose. The main problem lies in the free reuse of materials, since it is not guaranteed, although it is required, that the new user will cite the original author in the subsequent document. (Hoorn, 2006)

Questions about research data management

Research data covers a broad range of types of information and digital data can be structured and stored a variety of file formats. One of the main challenges of research data management is to start categorizing research data in order to make it manageable for storage and reuse. Furthermore, research data formats show a huge variety ranging from texts, spreadsheets, notebooks, reports, photographs, slides, workflows to models, algorithms, etc.

In the case of an institutional repository where services have to be developed in order to meet the needs of researchers from different scientific fields, it is essential to be prepared in order to deal with a variety of research data types and formats.

Inappropriate data storage leads to a loss of data, but most researchers do not think about using or reusing their own data. After publishing the research results, research data is not considered a management priority. Researchers should be educated about the methods for the long-term preservation of their raw data.

More than half the respondents indicated that their research data is stored on personal computers. The practice of using personal web storage to archive data present a danger for the long-term preservation of data. Storing and backing up research data is a critical element in the research process. However, simply saving the data is not necessarily sufficient to ensure its future usability. It is essential that time and effort is taken to prepare an archived copy of the used research data after the project has been completed. Archiving research data includes data protection which implies safeguards and periodic checks of file integrity. Organizing and documenting data is necessary to ensure that the data can be re-used in the future by other researchers.

In the case of the majority, the choice of a personal computer as the primary storage facility indicates two main things: (1) the archiving of research data is not organized at an institutional level, so researchers have to manage the data on their own, (2) researchers are reluctant to share their data with others, therefore they are not willing to upload data into institutional, departmental or research archives. Another reason for not using organized archiving facilities can be the lack of awareness among researchers of the advantages and methods of long term data preservation.

The survey also indicates that publications are only linked to research data on researchers' personal computers. Institutions have not started developing an infrastructure which ensures the long-term preservation of research data and the linking of data to publications.

Conclusion

Institutions cannot ignore the urgency of research data management at an institutional level. Authors will look for solutions outside their organization, because they are pressured by funders and publishers to comply and they lack an institutional infrastructure. Institutions should consider the development of their repositories to accommodate research data management principles to be a long-term investment. An institutional repository which ensures the long-term preservation of research data alongside the publications of researchers, has numerous advantages: (1) it contributes to the comprehensive collection of the institution's knowledge base, (2) the higher visibility and growing reputation of the institution and (3) more active participation in the international discourse on Open Access publishing and research data management (including participation at conferences, grant applications and consortia in this subject area). It is in the interests of the institution to channel and solve the researchers' demands internally and not to let the institution's scientific assets be stored and used externally.

The report from the Survey of Polish Scientific and Research-Development Units

Edyta Kędzierska, Natalia Kavalchuk, Jolanta Stepniak

E.Kedzierska@bg.pw.edu.pl, N.Kavalchuk@bg.pw.edu.pl, J.Stepniak@bg.pw.edu.pl,
Warsaw University of Technology

Introduction

The primary objectives of the research survey conducted with Polish scientific and research-development units include: describing the research data management at the surveyed institutions and illustrating ways for using open internet resources, providing source data and examining publications. One of the most important objectives was to determine the users' knowledge levels as well as the usage details of enhanced publications. We attempted to answer the following questions. What kind of data is produced at Polish institutions? Where and how long is it stored? Who is responsible for research data management? Are the research publications linked with the source data used within them? Is the research data re-used and provided to outside institutions? Are the data and publications provided in open internet resources? What opinion currently exists about the usefulness of the Open Access model and creating an open inter-institutional repository in a scientific environment? The answers to these questions have allowed us to determine the specificity of the production and distribution of scientific materials in the Polish scientific environment.

Survey description

Four hundred and eighteen Polish scientific and research-development units constituted the general population of the survey. The organizations that complied with these conditions included public scientific institutes, public institutions of higher education and the scientific institutes of the Polish Academy of Science. The research survey did not include commercial economic entities, private institutions of higher education, museums, archives or libraries. The analyzed units were individual institutions. We gathered information about the organization's research activities, but not the individual opinions of the researchers.

A research sample was selected by means of statistical methods for 268 units. Institutions were drawn randomly for the research sample. The survey was carried out using a questionnaire that was completed electronically and distributed via e-mail.

The survey was conducted over two months from 12.05.2014 until 13.07.2014. The selection process revealed that 22 scientific or research institutions had either been liquidated or merged into the administrative structures of other institutions. The general population was reduced accordingly to 396 units. During the survey, we verified that the number of completed questionnaires was less than 40%. We therefore decided to send the questionnaire to the reserve group of institutions which were not primarily selected for the research survey. As such, we ultimately gathered 207 questionnaires, which represents a 77% participation rate.

Taking into consideration the research sample size, we acknowledge that the survey was carried out using representative sampling. Based on the results, we can generalize about the situation regarding the entire defined general population.

About 63% of respondents filled out the questionnaire anonymously. A significant number of the units (34%) did not specify their institution type. Among the collected opinions, 46% of them come from research institutes, 26% from public institutions of higher education, 21% from scientific institutes of the Polish Academy of Science and 7% from institutions that specified themselves as another type of institution not defined within the population. The total institutions (217) were twice as likely to represent scientific disciplines than the humanities (111).

Questions about Open Access publishing

About 48% of respondents declared that researchers publish their materials in Open Access, while the other 52% asserted that the researchers either do not publish in Open Access or it is impossible to verify. Higher education institutions and research institutions of the Polish Academy of Science (more than 50% of the institutions) publish in Open Access. The other types of institutions (77% of them) declared that they do not publish in Open Access. Thus, the situation is quite difficult to define. 16% of institutions could not precisely say, if the researchers publish in Open Access or not.

The institutions whose researchers publish in Open Access most often use institutional repositories (65% of units) or external subject-based repositories (48%). 22% of the institutions use more specific and unusual solutions.

The majority of institutions (82%) do not use public licenses for providing research data. 71% of organizations do not want to change the situation. Data survey analysis revealed that Open Access publishing promotes the use of public licenses. The dependency is reversed.

It is remarkable that publishing papers or source data in open internet resources is not widespread with regard to Polish research institutions. Researchers prefer to publish their results in peer-reviewed commercial publications instead of providing source data in Open Access for a wider audience.

Opinions about Open Access publishing

About 80% of participants specified some of the advantages and obstacles of Open Access publishing. It seems that the awareness of the advantages of Open Access publishing is quite high. At least six advantages to an institution or researcher were indicated. However, about 40% of participants specified some of the obstacles of Open Access publishing. The most frequently mentioned advantages include: broadening the readership, increasing citation frequency, developing scientific communication and faster information exchange. The main obstacles include: unclear copyright issues, additional costs and the questionable quality of some published papers. Though awareness of the advantages of Open Access publishing is high, the majority of researchers do not publish materials in open internet archives and repositories.

71% of institutions wish to participate in creating a central inter-institutional repository. The experience of Open Access publishing, using public licenses for research data and providing and linking data with publications have led to a wish to participate in a project providing universal access to research data.

Questions about research data management

The research data produced within the institutions is related to both the type of unit and the cultivated scientific discipline. More than 60% of the institutions ascertained that they produce research data from measuring and experiments (exact science), while 10% mentioned other

types of unspecified produced data. On average the institutions stated that they produce four to five different types of data. In all types of institutions, the most common format for data storage is “.PDF” (87%), followed by “.doc” second place (67%). 11% of units use other atypical formats for data storage.

In most institutions, research data is stored in places which are only accessible to the institution's staff, for example, on individual workstations or a central server, which were the two most frequently given responses. Only 12% of units mentioned open resources, either their own or external ones. In addition, the employees at 71% of the institutions are responsible for research data storage and archiving by themselves. A minority of institutions (46%) used the central data management administered by designated departments or persons. 80% of the institutions declared that the data is stored for more than 10 years. The collected data shows that research data management in Polish scientific institutions is realized in an unsystematic and decentralized manner. Individual employee responsibility for data storage in the parent institution represents the clearly preferred option. Moreover, the organizations do not wish to change the situation. 80% of units do not plan to make changes concerning the method of research data storage.

About 77% of Polish institutions declare that they have linked the research publications to the source data used within them. About 23% of units do not undertake this activity. The institutions that link research data with the publications, do so by adding information about access to the data within the publication (68% of institutions). This way of linking data with publications is quite far from the model of enhanced publications, which is based on publishing digital forms of data and research papers in open internet resources. Less than 30% of institutions mentioned that they use this method of data linkage. More than 76% of institutions declared that they linked data with publications by publishing information about access to the data, but not digital objects in open resources. 80% of the institutions that do not currently link data with publications do not plan to make any such changes in the near future.

The survey found that linking research data with publications encourages researchers to publish their materials in Open Access resources. On the other hand the institutions that do not link data with publications also do not publish in open internet resources. The observed relationship is reversed: the fact that the researchers at institutions publish their materials in Open Access promotes the activity of linking publications with data.

It was checked if the institutional data storage method has an effect on the practice of linking data with to the publications. Data storage in open resources promotes the practice of data linking. A slightly smaller number of institutions that store and archive data in closed resources links data with publications. In turn, linking data with publications furthers the usage of open files.

80% of the institutions declared that the data is re-used in other research projects and by other researchers. 70% of the institutions provided data for outside research institutions. The preferable way of providing data is distribution via e-mail. More than 50% of institutions exchanged data via individual contacts. The main reasons of not providing data to outside researchers are: the confidentiality of the collected data, data/office secrecy and intellectual property protection. 88% of the institutions that do not provide data to other researchers do not plan to change their current state of affairs. The research data provided is repeatedly used by researchers from the parent institution as well as by researchers from outside institutions.

Conclusion

Polish researchers recognize the importance of linking research data with publications and re-using and providing data, but it is not their everyday practice. Most of the activities concerning the storage, provision and linking of data are realized inside individual institutions or at the workstations of individual researchers, not in institutional open repositories. It is not common practice to publish data and research papers in Open Access, but Open Access is indicated as an important factor for research promotion and the improvement of citations.

Survey results show a need for the better promotion of enhanced publications at Polish research institutions.

The report from the survey of Slovak universities and scholarly institutions

Jozef Dzivak

jozef.dzivak@stuba.sk

The Slovak University of Technology, the Chemical Library

Introduction

Research data (or “výskumné dáta” in Slovak) represents a valuable background for researchers, teachers, scholars, scientists, students and other groups constituting the intellectual expert community. Such data includes schemes, models, plans, surveys, documentations etc. In Slovakia, there is no comprehensive working system built on such data, which would interconnect the institutions and organizations dedicated to education, linking their data, complementing each other and integrating the data into a compact tool. It is expected that this data exists at its parent institutions and is accessible in some form to the visitors of the individual institutions or their websites. The survey should give an idea of how exactly it is formed, stored and made accessible.

Survey description

The survey evaluation aims to map the situation for research data in Slovakia in its scholarly environment and follows its creation, accessibility and storage. It shows the current position of data based on a survey and questions forwarded to leading Slovak universities and scholarly institutions.

In the case of Slovakia, the targeted organizations represented the most prestigious private and public education institutions with the biggest influence, which identified themselves as research institutions and institutions dedicated to higher education.

In total, there were 36 institutions contacted via an informative email explaining the background of the survey and the project. In those cases where no response was obtained, the survey collectors tried to reach the contact persons by phone.

In Slovakia, the survey itself was performed between 20th April and 30th June, reaching the peak of collected answers 20 days after the beginning of the survey and the second rather smaller peak 40 days after the beginning of the survey (after the institutions had been further contacted by telephone).

Unfortunately, only 44.4% of the contacted institutions responded. Other institutions refused to answer or were not able to answer. The reasons were: secrecy of information, lack of research data or undefined reasons.

Questions about Open Access publishing

This set of questions shows that no institution is obliged to store and make its research data accessible. Only a small portion of institutions keeps the data in a personal (website) repository of the individual researchers or in institutional digital repositories.

Paradoxically, many institutions see advantages in Open Access, the main advantages are opportunity to reach a wider audience, increasing the visibility of publications in reference databases and for web search engines and establishing cooperation between researchers.

Other very frequent answers included reaching a wider readership, increased citations, increasing the prestige of the institution and researchers, the development of scientific communication and faster information exchange.

However, institutions also saw negatives of Open Access. The most frequent negative answer involved the uneven copyright issue, which constitutes an external reason.

Questions about research data management

The second set of questions focused on the creation, type and format of the data. The set also included the past, present and future plans for data storage, who is responsible for the data storage, as well as the implementation of a system for data storage. The most used processes upon the basis of which data is produced were identified as measurements, experiments and testing.

And this is what the most used formats and types are dependent on. Schemes and models were the most frequently used forms of material accompanying enhanced publications. The most used formats for research data were also represented by visual materials such as tables or texts. As such, doc, pdf and xls were the top formats for research data.

The majority of this data is kept and stored by individual research workers, which may suggest that the majority of institutions do not have a centralized secured storage facility managed by the institution, even though the second most popular answer (over 21%) was that the data is stored on a common server. The third of institutions have an appointed librarian who stores the data. Individual researchers is responsible for data most often. Only a quarter of the institutions are planning to change the current situation.

The third set of questions focused on the links between research data and research papers and whether the data is reused by any other research workers/institutions and if it is marked with public licenses.

The survey showed that the research data is only linked with the research papers in a quarter of the responding institutions, mainly by giving information on the data availability in publication records (in metadata). However, in those cases where the data wasn't linked, only 40% of the institutions were planning to change it and to start linking the data and the papers.

The data produced by institutions is usually reused or available for reuse by other research workers in their projects. However, the data is also reused by researchers from other institutions to a lesser extent. It may be pointed out that the majority of such reuse and data acquisition by researchers from other institutions is based on personal communication – a personal visit to the workplace or sending the data by e-mail. Only a small amount of the distribution of data is executed by means of online services. In most cases, the institutions are not planning to change the current situation.

In most cases, the institutions do not use public licenses (80%) for providing and marking the research data and they also do not plan to use them (80%).

Even though there are not many plans to change the current situation and exploit more possibilities on how to use the research data, all the responding institutions are willing to

participate in building a common inter-institutional repository where the data would be stored and managed.

Conclusion:

The survey demonstrated that the questions of Open Access, research data, its availability and storage, licenses and other important issues are only beginning to be opened or in some cases are not yet open. Many institutions do not have online repositories or work on the availability of the research data. The institutions showed the will to cooperate on such repositories, however, they do not have a concept or any systematic plans to open up their data, which they often share with other researchers. They may also be lacking the necessary staff, know-how or leadership.

The report from the survey of Czech research institutions and universities

Petra Pejšová, Hana Vyčítalová

petra.pejsova@techlib.cz, hana.vycitalova@techlib.cz

The National Library of Technology

Introduction

While in Western Europe, in particular in the Netherlands, data repositories (i.e. repositories not only for publications, but also for the data as such) have been in operation for several years, accompanied by projects promoting enhanced publications, this practice is still new in the Czech Republic and the general situation of enhanced publications has not yet been mapped yet. “Enhanced publications allow for a fuller understanding of the process in which data and information are used and applied in the generation of knowledge.” (Farace, 2012)

Survey description

The NTK (National Library of Technology in Prague) conducted the survey of enhanced publications in two rounds. In the first round in 2013, we addressed public research institutions, private research institutions and other institutions engaged in research. We used the results of a survey focusing on linking research data to full-text publications in the community of grey literature as the basis of the survey. The conclusions that “Data exchange is becoming the norm in Open Access communities” (Farace, 2011) and “Research data should be preserved and accessible in order to enhanced scholarly communication” were the pivotal starting points.

In the second round in 2014, the survey was sent out to all public and state colleges and universities in the Czech Republic. Where possible, representatives of the individual faculties at the colleges and universities were addressed. Wherever a smaller college or university that is not divided into faculties was concerned, we tried to obtain a summary response for the entire institution. We also encountered cases where research was dealt with broadly on the level of the institution but the questionnaire could not be completed for individual faculties. The questionnaire was in such cases also completed for the entire college or university.

We addressed the representatives of these institutions whom we expected to have a comprehensive knowledge of the research activities at their institutions. We chose people holding the posts of science officers, science secretaries, deputy directors for research or science, vice-deans (or vice-rectors) for science and research and heads of science and research departments, as applicable. In cases where it was not possible to determine who holds such posts in the institution or where no such posts had been set up by the institution, representatives from the institution's management were selected.

In the first round, we addressed 113 research institutions and obtained 69 responses, of which 65 were completed questionnaires. Four institutions refused or were not able to complete the questionnaire. Questionnaire response rate was 57.5%. In the second round, we addressed in total 165 faculties at 26 Czech public and state colleges and universities. We obtained 96 responses, of which 77 were completed questionnaires, while 19 faculty representatives informed us that they were either too busy to complete the questionnaire or could not complete it. Questionnaire response rate was 46.6 %.

There were various reasons behind the decision not to complete the questionnaire. The representatives of some of the institutions did not want to provide any information without giving a reason or they did not have the time to complete the questionnaire. At some institutions, we encountered the problem that they essentially do not have any research data.

Questions about Open Access publishing

The first four questions of the survey focused on Open Access in scientific publishing. Nearly half of the respondents from the faculties of colleges and universities engaged in the practice of making research publications available in open repositories. Colleges and universities mostly use institutional repositories or employees' personal websites to publish research publications in Open Access mode. The advantage of Open Access publishing identified most often by colleges and universities was the opportunity to address a broader audience, the ability to achieve higher citation and readership rates and the faster exchange of information among scientists. As far as obstacles were concerned, they expressed concerns about the low prestige of Open Access journals and repositories, as well as copyright issues and the fees connected with Open Access publishing.

Questions about research data management

The main part of the survey focused on examining the situation with regard to providing of access to enhanced publications and research data in the Czech Republic. These questions were addressed both to research institutions and faculties at public and state colleges and universities.

The most frequent type of research data was measurement data, followed by research data from experiments, testing, surveys, etc. at research institutions, while, on the other hand, it was post-publication data, such as reviews and evaluations, at colleges and universities. The most common data format was, unsurprisingly, the PDF format, followed by the DOC, XML, spreadsheet formats.

One of the most crucial questions was how the institutions archive their data. The survey confirmed that there is no centralized solution for data storage at nearly half of the institutions. An overwhelming majority of respondents in both surveys (86%) confirmed that they reused data from previous research projects as the basis or material for further research. The surveys showed that the majority of institutions do not have any centralized solution and that data is most often stored only on individual researchers' workstations. This approach does not guarantee security and long-term access to data.

The ideal solution of data management and storage – in central digital repository of the institution – was indicated by 29 respondents from research institutions and 20 respondents from colleges and universities. Most respondents both from research institutions and faculties of colleges and universities responded that they were not planning any changes in their present method of data storage. The permanent archiving of data might be already taken for granted at present. Nevertheless, 30 representatives from research institutions and 33 respondents from colleges and universities either stated that this was not the case or did not know the answer. About the same number of respondents confirmed long-term archiving. Most often, the researchers themselves manage the data. Whenever there is a particular person or department authorized to take care of research data, it is usually the library, the IT department, the scientific secretary or the head of the department or other organizational units.

The survey's key question was, whether the research data are linked to the research papers. The fact that an absolute majority of research institutions and colleges and universities already

links research publications to research data in some way or other was a positive finding. In the case of research institutions this involved 38 respondents out of 65, while the result was 42 out of 77 respondents in the case of colleges and universities. Recording information about related research data in a bibliographic record is the most frequent method of linking research publications and data. However, such a procedure requires further steps to obtain the data. An ideal solution is linking a research publication to raw data or other research materials in a digital repository. This option ranked second, followed by the joint storage of publications on a website. Only a few of the respondents, who do not currently link data to publications, are planning to start.

60% of the respondents stated that they were also willing to provide data to researchers outside their institutions. This would take the form of a personal visit by the individual interested in the provision of data or the sending of the data by electronic mail or remote electronic data transfer (FTP). The online provision of data, either on a website or through a digital repository, was selected by a considerably smaller percentage of respondents. As another option, respondents stated that they published data as a printed annex to research publications. We were also interested in the reasons as to why institutions do not want to provide research data to interested parties outside the institutions. There were several main reasons. Some institutions stated that they only published final official research output, i.e. research publications, articles in scholarly journals. Private research institutions often cannot provide data for commercial and copyright reasons. Copyright issues were also mentioned by other institutions. Research institutions also consider research data to be a trade secret or their know-how; they have concerns about their competitors, etc. Some data is expressly subject to a secrecy provision; research institutions fear that data may be misused. The provision of data is also complicated by patent protection. Most respondents that indicated that they did not provide access to data are not planning to provide data outside their institutions for the above reasons.

We were interested in using Creative Commons licenses at research institutions and universities. The survey showed that Creative Commons public licenses that help to make the use of publications and data easier are still only used sporadically in the Czech Republic. There is a greater awareness and use of Creative Commons licenses at colleges and universities. Colleges and universities are therefore also more likely than research institutions to introduce Creative Commons licenses in the future.

The questionnaire also contained a question asking whether the institutions would be interested in establishing a central data repository. The reaction of the colleges and universities was fifty-fifty. Research institutions showed indecisiveness which probably stems from the fact that the topic of data repositories is new for these institutions in the Czech Republic.

Conclusion

On the whole, the survey showed that it is advisable to continue with the work on public education, opening up discussion on Open Access, creating enhanced publications and providing access to research data in the Czech context. The NTK has created an Enhanced Publications section on its website www.techlib.cz which provides information about enhanced publications and links to materials available globally and monitors the situation in the Czech Republic. In addition, NTK has included the topic of enhanced publication on the agenda of its educational events and conferences and it is cooperating with organizations dealing with this issue abroad.

Conference on Grey Literature and Repositories: proceedings 2014: the Value of Grey Literature in Repositories [online]. Prague: National Library of Technology, 2014 [cit. 2014-12-19]. Available from: <<http://nrgl.techlib.cz/index.php/Proceedings>>. ISSN 2336-5021.

References:

AYRIS, Paul. Open Research Data: Statement. *LERU* [online]. 2012 [cit. 2014-10-16]. Available from:

<<http://www.leru.org/files/general/Open%20Access%20to%20Research%20Data-FINALdocx.pdf>>.

FARACE, Dominic. Repurposing Grey Literature - Linking Research Data to Full-Text Publications: Some Preliminary Results. In: *Seminar on Providing Access to Grey Literature 2011: 4th year of the seminar focused on storage and making the grey literature accessible*, 25. 10. 2011. Prague: National Technical Library, 2011, s. 6. ISSN 1803-6015. Available from: <<http://www.nusl.cz/ntk/nusl-82067>>.

FARACE, Dominic, Frantzen JERRY, Stock CHRISTIANE, Laurents SESINK a Rabina DEBBIE. Linking full-text grey literature to underlying research and post publication data: An Enhanced Publications Project. In: *Thirteenth International Conference on Grey Literature: The Grey Circuit, From Social Networking to Wealth Creation*. Amsterdam: TextRelease, 2012, p. 10. ISSN 1386-2316. Available from: <<http://hdl.handle.net/10068/700153>>.

HARNAD, Stevan. Maximizing university research impact through self-archiving. *University of Southampton* [online]. 2003 [cit. 2014-10-16]. Available from: <<http://users.ecs.soton.ac.uk/harnad/Temp/che.htm>>.

HOORN, Esther a Maurits VAN DER GRAAF. Copyright Issues in Open Access Research Journals: The Authors' Perspective. *D-Lib Magazine* [online]. 2006, vol. 12, iss. 2 [cit. 2014-10-20]. DOI: 10.1045/february2006-vandergraaf. Available from: <<http://www.dlib.org/dlib/february06/vandergraaf/02vandergraaf.html>>.

Trends in Scholarly Communication. *Brandon University* [online]. 2014 [cit. 2014-10-16]. Available from: <<http://libguides.brandonu.ca/content.php?pid=223976&sid=1857707>>.

Appendix

Enhanced Publications Survey in V4 Countries

Questionnaire for research organizations and universities

Open access

Part 1 of the Questionnaire regards the gathering of information about publishing and archiving research papers in Open Access

1. Are the research papers of your institution stored and accessible in open archives and digital repositories?

- Yes, it is obligatory – go to question 1a
- Yes, though it is not obligatory – go to question 1b
- No - go to question 2
- Hard to say - go to question 2

1a. If it is obligatory – since when

1b. If yes, what type of archive (repository) is most frequently used:

- Personal web-site (repository) of a research worker
- Institutional repository
- Subject-based repository
- Other (please specify)

2. What are/could be the main advantages of publishing or storing in Open Access for your institution and researchers? (please indicate all appropriate)

- Opportunity to reach a wider audience
- Opportunity to obtain additional funding, grants
- Cost savings
- Wider readership for research papers published in OA journals
- Increased citations
- Increasing the visibility of publications in reference databases and for web search engines
- Shorter dissemination time
- Greater durability of publication
- Increasing the prestige of the institution and researchers
- Greater influence on development of a scientific discipline
- Development of scientific communication and faster information exchange
- Establishing cooperation between researchers
- Other (please specify)

3. What are/could be the main obstacles for your institution and researchers that discourage them from publishing or storing research papers in the OA model? (please indicate all appropriate)

- Lack of OA journals/repositories in the appropriate field
- Low prestige of OA journals/repositories

- Small influence of OA journals/repositories on the development of a scientific discipline
- Quality of the published papers (not peer-reviewed papers) is not proven
- Quality of the published papers in OA journals is not confirmed (lack of peer-view)
- Additional costs (required fees)
- Narrow readership for research papers published in OA journals
- Objections of traditional publishers
- Uneven copyright issue
- Additional work
- Other (please specify)...

Raw research data

Part 2 of the Questionnaire regards the gathering of information about raw research data management at the institution

4. What types of research data are produced in your institution? (please indicate all appropriate)

Research data produced:

- from measuring
- from experiments
- from testing
- from surveys
- from statistical investigation
- from another activity (please specify)

Accompanying materials

- videos
- audio records
- plans
- models
- schemes
- visual documentation
- algorithms
- other (please specify).....

Post-published data materials:

- reviews
- evaluation
- comments
- other (please specify)....

5. Indicate the most common formats for data storage in your institution:

- doc
- pdf
- csv

- xls
 - xml
 - JSON (JavaScript Object Notation)
 - other (please specify)...
6. Where are the raw research data of your institution stored?
- On the workstations of individual research workers
 - In a directory on a (common) server
 - In a central digital repository of the institution
 - In a centralized subject-based open archive
 - In other place (please specify)...
7. Does your institution plan any changes to the way of storing and archiving the research data?
- Yes (please specify what kind of changes your institution plans to make)
 - No
8. Is the research data archived in your institution for more than 10 years?
- Yes
 - No
 - Not yet, but we started.....years ago
9. Who in your institution is responsible (takes care) for storing and archiving the research data?
- An appointed person(s) in the library
 - An appointed person(s) in the IT department
 - An appointed person – scientific secretary
 - Each research worker himself
 - Other person(s) (please specify)...

Dissemination and re-use of the research data

Part 3 of the Questionnaire regards the information about linking the research publications to the research data and their provision and use.

10. Is the research data produced at your institution linked to the research papers?
- Yes – go to question 10a
 - No – go to question 10b
- 10a. If yes, how is the research data linked to the research papers?
- By joint storing of data with publication in digital form on personal website of research worker
 - By joint storing of the data with publication in a digital form in the repository
 - By giving information on the data availability in the publication records (metadata)
 - By referencing to place of data storing in digital form from the research publication
 - In another way (please specify)....

10b. Is it planned linking data with research publications in your institution?

- Yes
- No

11. Is the data produced at your institution re-used in other research projects and by other research workers?

- Yes
- No

12. Is the data produced during the research work at your institution available for researchers from other institutions?

- Yes – go to question 12a
- No – go to question 12b

12a. How does your institution provide research data to researchers from other institutions?

- Personal visit at workplace
- Sending the data by e-mail
- Online on website
- Online in digital repository
- In other manner (please specify)

12b. Why is the research data not available to researchers from other institutions?

12c. Is it planned the providing data stored by your institution for the outside researchers in the near future?

- Yes
- No

13. Is it used the public licenses in your institution for providing and marking the research data?

- Yes – go to question 14
- No – go to question 13a

13a. Is it planned the using public licenses in your institution in the near future?

- Yes
- No

14. If there would be a project to build a long-term, central inter-institutional repository for research data, would your institution be interested in participating in it?

- Yes
- No

Identification information

Name of your institution:

Contact e-mail:

Type of institution:

- Higher education institution
- Research institute
- Academy of Science
- Auxiliary scientific unit
- Other (please specify)

Field of specialization:

- General / interdisciplinary
- Biological sciences
- Medicine, pharmacy and related subjects
- Social sciences
- Mathematical and computer sciences
- Engineering and technology
- Physics and related sciences
- Chemistry
- Psychology
- Earth sciences
- Agriculture and related sciences
- Economy
- Historical and philosophical studies
- Language and literature studies
- Astronomy and space sciences
- Architecture, building and planning