



# Enhancing scholarly communication

National initiatives to manage research data  
in the V4 countries

**Authors:**

Edit Görögh (Hungary)

Edyta Kędzierska, Natalia Kavalchuk, and Jolanta Stepniak (Poland)

Jozef Dzivak (Slovakia)

Petra Pejšová and Hana Vyčitalová (Czech Republic)

A large, light gray arrow points from the bottom left towards the top right. In the bottom right corner, there is a graphic of several overlapping, slightly tilted rectangular shapes representing papers or documents.

Sponsored by  
IVF Standard Grant  
2013

## Content

<b>Introduction .....</b>	<b>3</b>
Open access in scholarly communication.....	4
Research data publication .....	7
Enhanced publication .....	8
The IVF project .....	9
<b>Survey results .....</b>	<b>11</b>
Hungary .....	11
Czech Republic.....	29
Poland.....	51
Slovakia.....	77
<b>References .....</b>	<b>97</b>

## Introduction

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

The rise and spread of networked communication had a significant impact on scholarly communication, as well. The new models introduced to communication systems, to publications, and to publishing practices shared a common factor: they have become Internet-mediated activities. These models affect all facets of the communication process; they offer new forms of presentation, new interaction practices between authors and readers, new business models, etc.

Scholarly communication is changing the structure of the process, the dynamics of the participants, and the methods of interactions. Knowledge is being shared instantaneously. (Brandon University). Blogs, personal and institutional websites, videoconferences and online meetings help researchers to acquire information quickly on new developments in their field. Enhanced dialogue within the research community not only contributes to fast communication among researchers leading to research collaborations, but also provides opportunity to scientists to easily communicate with leaders in their field.

Scholars are also utilizing different channels of communication to reach non-scholars to involve them in collaborative research projects. The internet is also valuable in this sense not only for reaching audiences outside of academia for research purposes, but also for informing them about the latest findings in science.

Institutions of higher education make information available to the general public in a variety of formats. Scholars introduce their findings to wider audiences and invite people to learn more about the research process, which in turn generates interest and support for the next research project.

Visibility is an essential element of scholarly work. It has been long defining the careers in academia, since 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 their researcher communities. However, academia has become larger and more demographically diverse. A new generation of researchers is employing various channels of dissemination besides the traditional printed formats. The networked scholar is born, taking an active part in dissemination information about his/her work and him/herself as well. The interactive presence of researchers in social media enhances the visibility of their work.

Visibility has also become a key issue in funders’ requirements in the past decades. In most cases funders require that the research results financed by them should be demonstrated to all shareholders in an open and transparent manner. Such requirements are met usually 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**

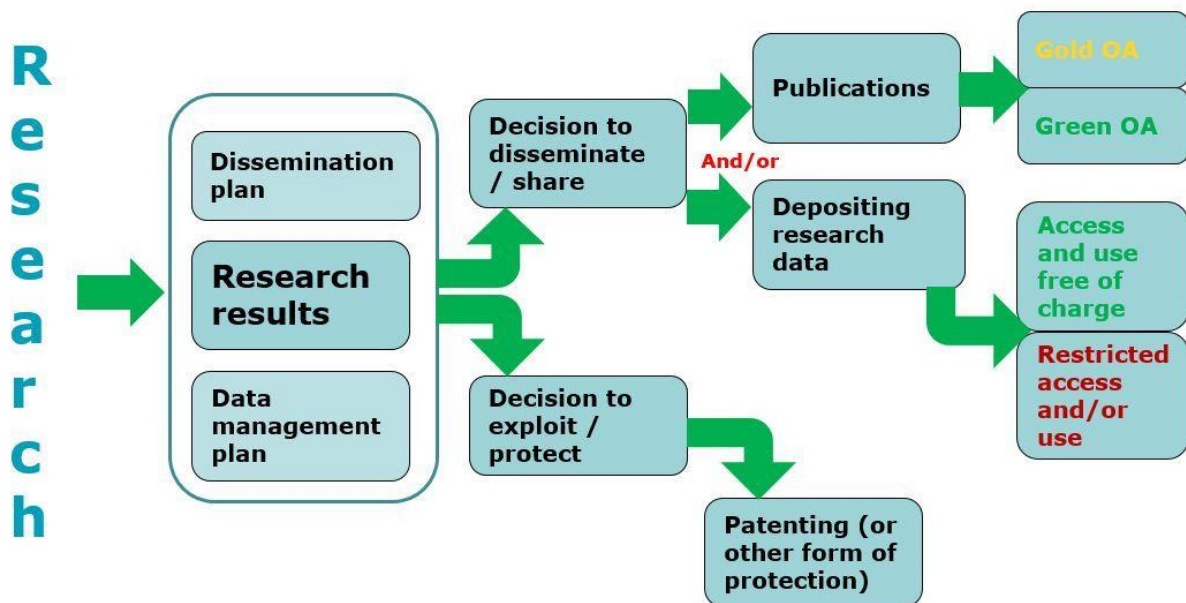
Open access is the practice of providing free on-line access to and free re-use of scientific information. Authoritative definitions of open access can be found in key political declarations on this subject (Budapest and Berlin Declarations). These definitions describe open access as including not only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

There are two main routes towards open access to publications:

- A. Self-archiving (also referred to as 'green' open access) means that the published article or the final peer-reviewed manuscript is archived (deposited) by the author, or a representative, in an online repository before, alongside or after its publication. Repository software allows authors to set a delay access to the article ('embargo period').
- B. Open access publishing (also referred to as 'gold' open access) means that an article is immediately provided in open access mode as published. In this model, the payment of publication costs is shifted away from readers paying via subscriptions. The business model most often encountered is based on one-off payments by authors. These costs (often referred to as Author Processing Charges, APCs) can usually be borne by the university or research institute to which the researcher is affiliated, or to the funding agency supporting the research. In other cases, the costs of open access publishing are covered by subsidies or other funding models.

The European Union pays special attention to publicly-funded scientific research. Its main objective is to optimise the impact of such research both at European level (FP7, Horizon 2020) and at 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 a 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 open access dissemination of scientific outputs is to enhance economic performance in the EU and improve the capacity to compete through knowledge.

The recent EU Research Framework Programmes, namely FP7 and Horizon 2020, concentrate on 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.



Open access has divided opinions about 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, 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.”

It is argued that open access publishing results in increased visibility and higher citation rates. The fast and toll-free access to results contributes to more readers of the article and possibly more reference to it by other researchers. In open access publishing, either by seeking open access journals, or by uploading the articles into an institutional repository, authors are more directly involved in the publishing process; therefore, they are more consciously involved in retaining exploitation rights.

Open access is often considered to be the solution to the serial crisis which was the result of the gradually growing subscription prices of scientific journals at the end of the twentieth century. The monopoly of scientific publishers encumbered the publishing process and made it difficult for the growing number of scientists to publish their results in quality periodicals. Furthermore, due to increasing subscription prices, libraries are forced to cancel journal subscription which restricts access to research results. Thus, open access removes the limitations from the free flow of results and improves the information supply. Through the green road of open access publishing, which includes uploading and archiving publications in an open access repository, the long-term availability of documents is also ensured.

Besides the many advantages of the open availability of research results, there are many opponents to this movement. One of the main reservations against open access concerns the quality control of this publishing process. Researchers are reluctant to choose open access journals because they think the peer review process is not as rigid as in the case of traditional scientific journals, and therefore the scientific validity and reputation of the journal is not as high as that of other, well-established journals in their field. It is true that numerous open access journals have tried to achieve quick visibility and citation statistics by accepting articles for publication without the necessary quality control. Authors need to check the background of the journals they intend to publish in, just as in the case of other printed or electronic not open access journals they would inquire about before sending the manuscript. The fact that authors pay for the publication, directly or through sponsorship, is regarded a primary weakness of the open access model. As a result, scientific outputs may have a less effective filter, and so a new post-publication filtering mechanism should be developed to match the quality control of traditional publishing process.

Another major concern on the researchers' side is the financeability of the author-pays model. Open access journal switched the financial aspect of access from the reader's side to the author's side. The authors are obliged to pay for the publishing process erasing this way the limitations from access for the readers. Open access journals offer open access publishing packages for institutions urging them this way to include author fees in their budget, and also suggest to authors to plan project funds with such fees in mind.

The increasing acceptance of open access publishing models compels traditional publishing houses to introduce open access into their product portfolio. A growing number of traditional publishers now offer free access to the articles they publish if the author pays the open toll. However, the new publishing models (open access, open toll, hybrid, etc.) require new legal considerations of the author's rights.

Traditional publishing houses are reluctant to rewrite their copyright agreements and transfer to a more author-centred solution. Instead they combine the Creative Commons licences, which is advised to be used by most of the open access publisher, with their present copyright agreement. (Heather Morrison)

Considering the many obstacles researchers face in publishing Gold OA, the Green OA publishing strategy offers a more accessible option. Institutional repositories provide the technological and human infrastructure for the long-term preservation and sharing of research results. Repositories are usually developed by the institutions in accordance with the best interests and demands of their researchers. Repositories also provide the opportunity to deposit materials related to the research publications.

## Research data publication

Besides open access publishing of scientific results, there are a growing number of initiatives around data publication. The pilot on research data in Horizon 2020 and other funding requirements signifies a 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, publishing and archiving data is important to preserve scientific information for future research.

Open data are the building blocks of open knowledge. Open knowledge is what open data becomes when it's useful, usable and used. (LERU)

The key features of openness are:

- **Availability and access:** the data must be available as a whole, at no more than a reasonable reproduction cost, preferably by downloading over the internet. The data must also be available in a convenient and modifiable form.
- **Reuse and redistribution:** the data must be provided under terms that permit reuse and redistribution including the intermixing with other datasets. The data must be machine-readable.
- **Universal participation:** everyone must be able to use, reuse and redistribute — there should be no discrimination against fields of endeavour or against persons or groups. For example, 'non-commercial' restrictions that would prevent 'commercial' use, or restrictions of use for certain purposes (e.g. only in education), are not allowed.

Data are considered an important resource; therefore, providing open access to them contributes to development in economic, social and scientific contexts. The main reasons for open data are:

- **Transparency.** The main principle of a democratic society is that the citizens are well-informed, which also includes free access to government data and information and sharing that information with other citizens.
- **Releasing social and commercial value.** By opening up data, government can help drive the creation of innovative business and services that deliver social and commercial value.

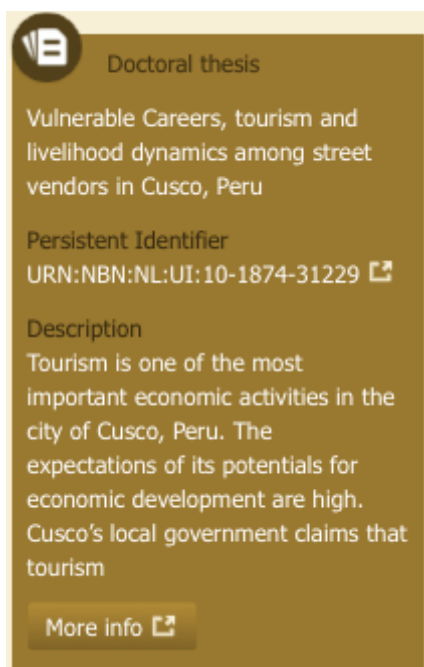
- **Participation and engagement.** By opening up data, citizens are enabled to be much more directly informed and involved in and contribute to the decision-making processes.

## Enhanced publication

As more focus is directed on research data, new methods of publishing emerge in scholarly communication. An enhanced publication (EP) is a totally new way of publishing in which a traditional publication (a book, an article or a report) is enriched with additional information. An enhanced publication relies on the linking possibilities of the web.

However, enhancing publications goes beyond just adding material that was necessary for writing the traditional publication. Readers of an enhanced publication will have the opportunity to comment on it. These comments will be added to the EP as well. The components of an EP will therefore vary in time, transforming it into a dynamic object. By doing so, research becomes more transparent to society.

In the Netherlands a preliminary data model has been designed that has been used by DANS to incorporate descriptions of EPs into its portal NARCIS. This unique system uses persistent identifiers for each published material. It allows the user to make connections between documents and obtain more information about the background and author of the publish result.



The image shows a digital library entry for a doctoral thesis. It features a circular icon with a stylized 'B' and the text 'Doctoral thesis'. The title is 'Vulnerable Careers, tourism and livelihood dynamics among street vendors in Cusco, Peru'. Below the title is the Persistent Identifier 'URN:NBN:NL:UI:10-1874-31229' with a small square icon. A 'Description' section follows, containing text about tourism in Cusco, Peru. At the bottom, there is a 'More info' button with a square icon.

Doctoral thesis

Vulnerable Careers, tourism and livelihood dynamics among street vendors in Cusco, Peru

Persistent Identifier  
URN:NBN:NL:UI:10-1874-31229

Description  
Tourism is one of the most important economic activities in the city of Cusco, Peru. The expectations of its potentials for economic development are high. Cusco's local government claims that tourism

More info

*Doctoral Thesis as a component of an EP, with its Persistent Identifier (URN:NBN)*



In the Netherlands, the SURF foundation and DANS have been working together to develop a visualisation tool for EPs that has been integrated into the NARCIS portal. (Hogenaar)

Enhanced publications foreshadow the future of publishing where access to the research results will be supplemented by information on the author, the research data collected and used, visual and audio materials produced during and after the research process, etc. Furthermore, based on the example of DANS innovations, all these materials will be able to be searched and visualized on one platform for the user.

### **The IVF project**

Developments in the scientific communications landscape and the advance of the green way of open access publishing urge researchers to deposit and archive not only their scholarly outputs but also the research data underlying their publications. IVF, focusing on the Visegrad countries (Hungary, Czech Republic, Slovakia, Poland), follows these changes in scholarly communication in this region and attempts to identify the milestones of the process through the projects it sponsors. Under the coordination of the University of Debrecen, four institutions from the Visegrad region started a cooperation to share experiences and map the national situations of research data management in this region. The participating members are the National Technical Library in Prague (CZ), the Chemical Library at the faculty of Chemical and Food Technology of Slovak University of Technology and Warsaw University of Technology.

There is a growing discourse about research data management: handling scientific data and linking them 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 preserve scientific information and improve access to it. However, there are national differences in technical and financial means to contribute to these programs.

The project, in the framework of which the present analysis was developed, 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 national initiatives to join library efforts to manage research data including collection, archiving and linking them to publications, and foreshadow preliminary national strategy plans to handle enhanced publication. Secondly, these national efforts are tied together in a joined analysis, which will review the national repository landscape in all V4 countries, and will discuss the current

issues facing repositories and data management in this region. Thirdly, the project intends to educate library/repository staff and researchers about the significance and management of research data in the form of workshops.

### *The survey*

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

NTK prepared a survey already in 2013 and tested it on research organizations in the Czech Republic. This survey subsequently served as a basis for a joint survey of all V4 institutions in 2014. Only four new questions concerning Open Access publishing were added.

The aim of the survey was to find out what research data are produced and archived by research institutions and public and state colleges and universities, as well as to provide a basic overview of Open Access publishing. In addition, the survey focused on the manner of data storage and archiving, the further use of the data and, most importantly, whether data are linked with research publications and whether we can consequently talk about enhanced publications in the Visegrad countries.

## Survey results

### Hungary

The project focuses on research data management, and discusses the transforming role of libraries in handling enhanced publications. The survey helps to map out the national repository landscape and identifies the main problems institutional repositories are facing in open data management. This project offers an opportunity for the participants to join the new European discourse of research data publishing and to increase the competitiveness of the Visegrad countries in the international open access movement.

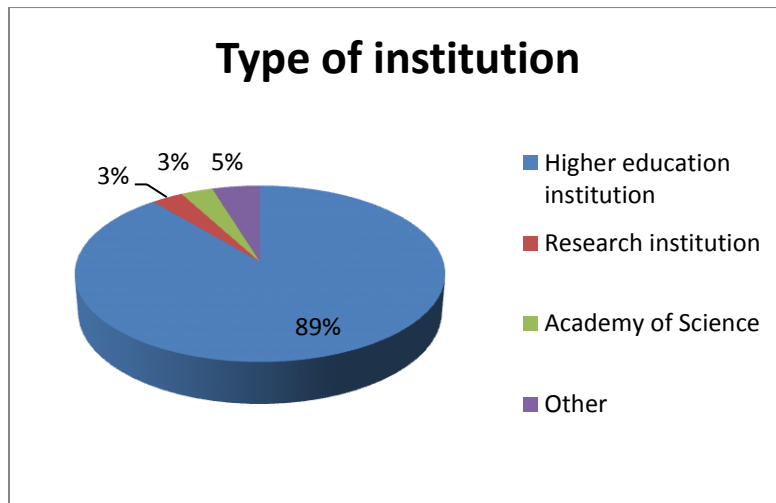
The survey executed in this project follows the questionnaire by the Czech partner, who started research in this subject a year earlier. Due to the fact that the other three participating institutions did not have prior experience in open research data, we felt it would be wise to include some questions on open access publishing in general. This way, an attitude on open access in general can be estimated besides the more detailed examination of data management practices.

The survey was distributed in two different channels. Both channels ensure communication among researchers and administrators of Hungarian higher education and research institutions. First, it was sent out to the HUNOR members. 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 outputs 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 a comprehensive cover of circulation. We requested from the libraries to send the survey to head of departments and researchers, as well. Questioning solely the leaders of departments or research fields would have not given us adequate information about research data managing practices, since it is the researcher and the research groups who actually deal with data on a regular basis.

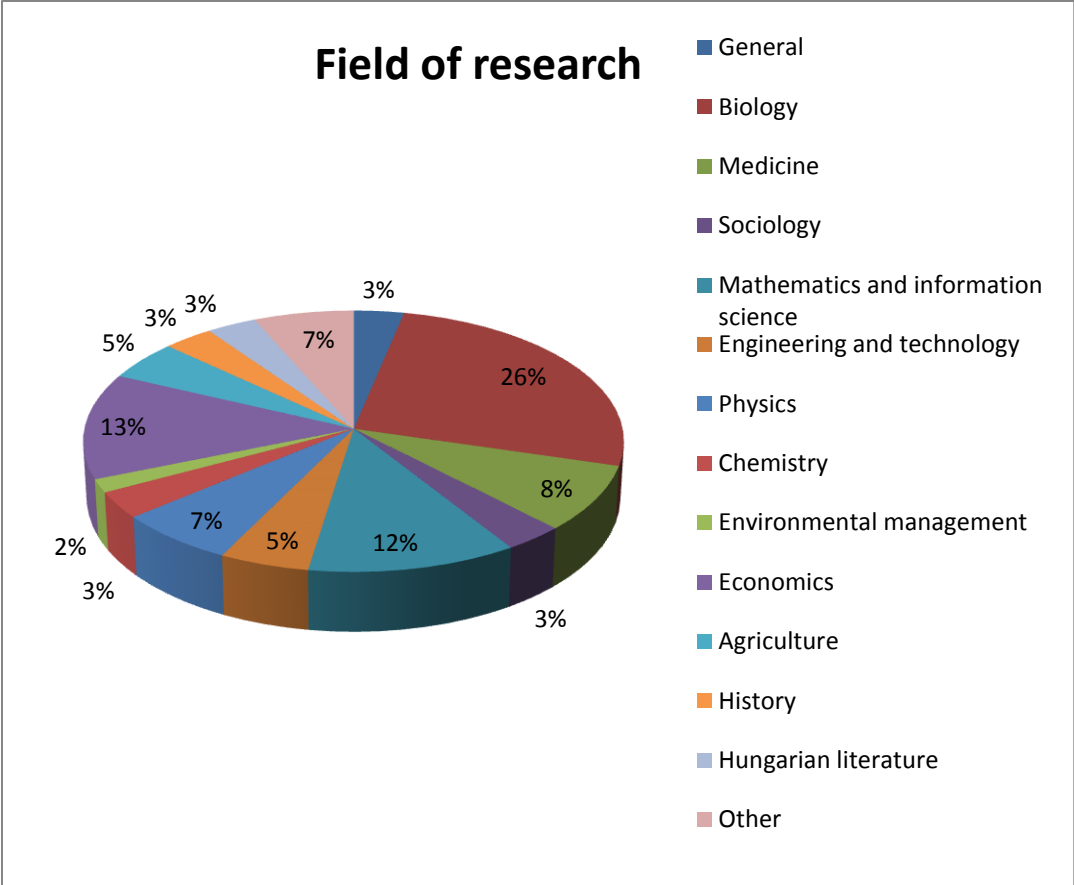
The other channel through which we tried to reach as many researchers as possible is the Database of Hungarian Scientific Documents. This national archiving initiative operates a network of administrators in all higher education institutions. The administrators are in direct contact with departments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

We received 70 answers, most of which arrived from higher education institutions, the majority of them from the University of Debrecen. Other universities participating in the

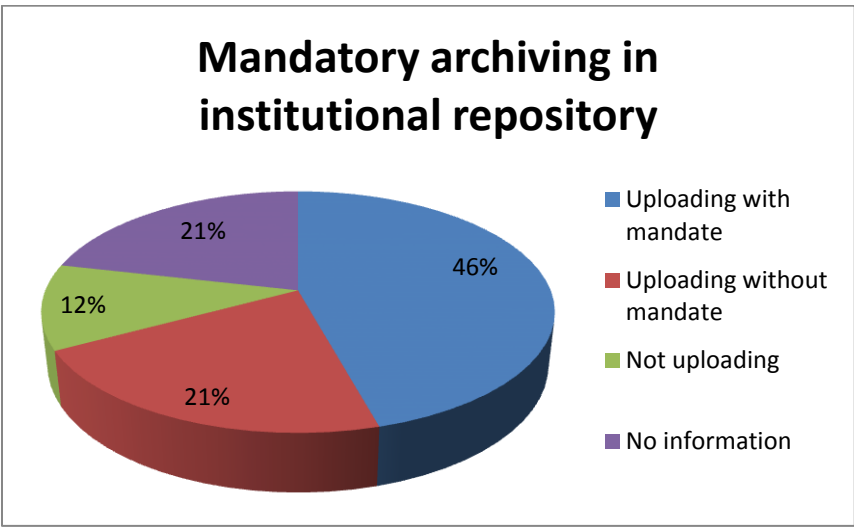
survey included Corvinus University and the University of Economics in Budapest, the University of Pécs, the University of Szeged, and the University of Miskolc. Through the responses from the Hungarian Academy of Science and the ALÖKI Applied Ecology Research Institution, the viewpoints of research institutions were also represented in the answers.



The field of biology was represented in the largest number of the responses, followed by the fields of economics, mathematics and information science. Researchers in these disciplines generally have to deal with research data on a regular basis; therefore, they have experience in collecting and storing data. We received responses from a large array of fields which gives us a nice overview of the research data management practices. In international research data initiatives the main objective in the long run is to regulate research data management and set up standards to make the sharing and reuse of research data more efficient. Therefore, the wider spectrum of research fields such analysis covers, the more information we have in the development of data management policies. On the flip side of the present analysis, which involves the examination of all research fields with one particular set of questions, it is a major shortcoming of such questionnaires that the differences of data management practices in various research fields are not considered. Therefore, the present questionnaire is mainly aimed at collecting general information on what type or size of data we should focus on in the future.



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



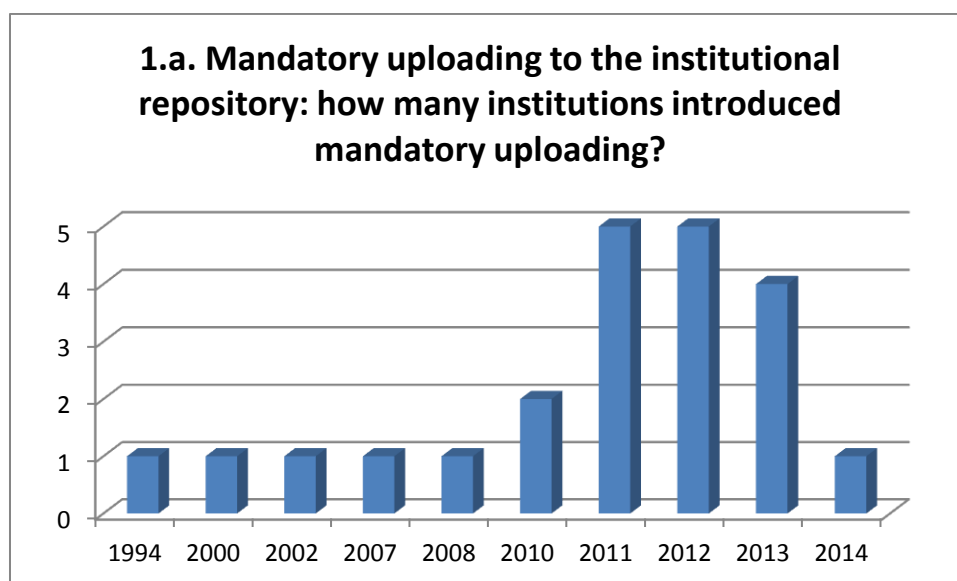
The answers to the first question imply that the majority of the respondents upload their academic work regularly to an institutional repository. The participating higher education

institutions operate repositories, since in most cases an institutional mandate requires the uploading of scientific results to an institutional archive.

At the University of Debrecen, following Government Decree 33/2007, which transferred the management of PhD dissertations to the jurisdiction of the universities, the Rector issued a mandate on the archiving procedure of PhD dissertations. This mandate was extended in the following year from the PhD dissertations to include all scientific publications (journal articles, conference proceedings).

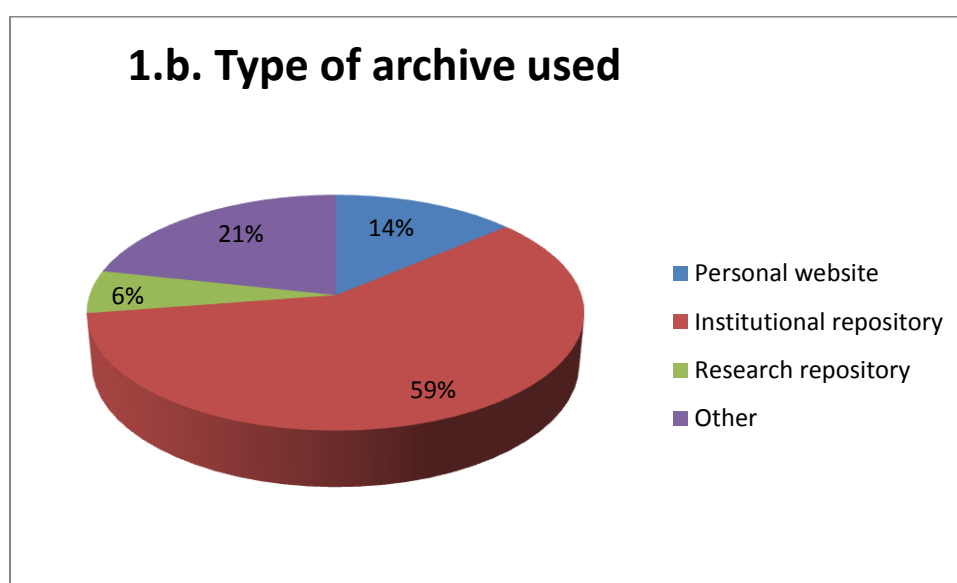
The visibility of and access to scientific outputs is not only regulated by the institutions themselves, but also requested by funding bodies outside of the institutions. In the past decades higher education institutions put more emphasis on the management of their scientific outputs. Due to continuously decreasing state funds, universities and research institutions have to find supplementary funds for operations. Projects funded by EU organizations require the accessibility of the research results to which they 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 universities use to attract more funding are based on the content of the repository.

Higher education institutions are inclined to regulate the archiving process of the scientific outputs of their researchers by institutional mandates ensuring the entry of the outputs into the repositories. The subquestion of Question 1 clearly indicates the growing inclination at universities to issue an institutional mandate for archiving.



### 1.b. What type of open archive (repository) is used most frequently?

The second subquestion of the Question 1 provided an opportunity for the respondents to provide details about the type of archive they use to store and share their scientific results. The majority of respondents named the institutional repository as a primary archiving solution. The “other” category was second most frequently chosen, which could be explicated further in the survey. Responses really varied, with most of them mentioning personal websites or websites of departments as a location of storage and access of articles and conference papers. Numerous respondents mentioned the central Archive of Scientific Works (MTMT), which is a bibliographic database of Hungarian scientific publications. These answers imply that the question was misunderstood by many respondents, since the question meant to refer to the accessibility of the publications themselves and not only the bibliographic data. In Hungary uploading the bibliographic data to MTMT is not mandatory; however, the publication lists prepared on the basis of this database are often requested to be included in appointment procedures or grant applications. The third most frequently chosen answer was personal websites as the storage location of scientific works. Interviews and discussions with researchers also reinforced the assumption that researchers use their personal computers to store their work, or at the best, create a personal website to provide access to it. Research archives were also designated as a primary storage place. Research repositories are used by researchers to upload scientific results of particular research fields (for example CogPrints for sociology, arXiv for physics). There is a long tradition within certain fields of study to upload preprints to the research repository of the filed in order to share results with colleagues. Numerous researchers at the University of Debrecen prefer using research repositories of their fields since these databases provide instant international visibility of their work within their research community.



## **2. What are/could be the main advantages of publishing or storing in Open Access for your institution and researchers?**

Open access has been long introduced to scientific discussions in Hungary. The University Library of Debrecen has been organizing workshops, 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 the 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 get more consciously involved in the publishing process. As open access publishing is becoming part of their research procedures, scientists begin to experience the effects of open access publishing more in practice.

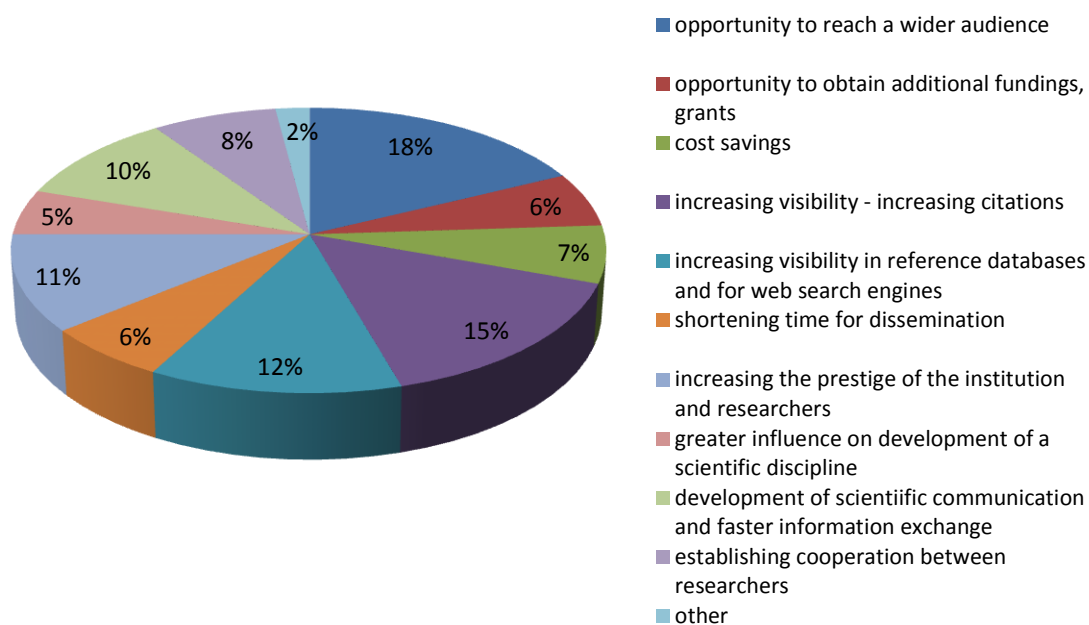
The majority of respondents thought that the most significant advantage of publishing or storing in open access is the opportunity to reach wider audiences. Since the goal of universities is to create and disseminate knowledge, researchers consider the principle of transferring information to wide audiences essential in 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 number of citations to their work. Therefore, the advantage of increased citation in open access publishing is valued highly among researchers. Visibility of an institution or scientific results 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 projects, etc. Another answer receiving high scores in this question was the development of scientific communication and faster information exchange as an advantage of open access. Closely related to this previous answer, several respondents valued open access in establishing new cooperation among researchers.

Based on the responses, cost savings was not considered an advantage of open access by many. Understandably, the transfer of expenses to the author's side in open access publishing (especially in the gold way of publishing in OA journals) does not imply a cost-saving solution for researchers. Respondents marked low on the priority list to receive the opportunity to obtain additional funding and grants through open access. Researchers do not think that providing access to their scientific outputs has a direct effect on successful grant application.

There were only a few negative comments on the advantage of open access publishing, the majority of the respondents agreed with the principles of free access and marked several advantages in the list.



## 2. Advantages of publishing or storing in open access for the institution and researchers



### 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?

The answers to this question clearly represent the reasons why open access is slow to gain ground within research communities in this region. Although it has been introduced to the scholarly dialogue, open access has not proved to be a vital publishing choice for researchers. The low prestige of OA publishing often derives from two main features of open access journals: relatively new in the market and costs money for the author to publish.

A common concern is that open access journals are of lower quality because they charge an article processing fee. Peter Suber in his book on Open Access confutes the costliness of OA publishing in general. Only about 30% of OA journals charge an article processing charge, but more than half of subscription journals work with some kind of processing fee, like page charges. (Peter Suber) The real reason for the low prestige lies in the fact that OA journals are newer and younger than subscription journals. All new journals need excellent publishing materials 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. However, we have to note that there are predatory open access journals in the market which abuse the OA business model. Their “mission is not to promote, preserve, and make available scholarship; instead, their mission

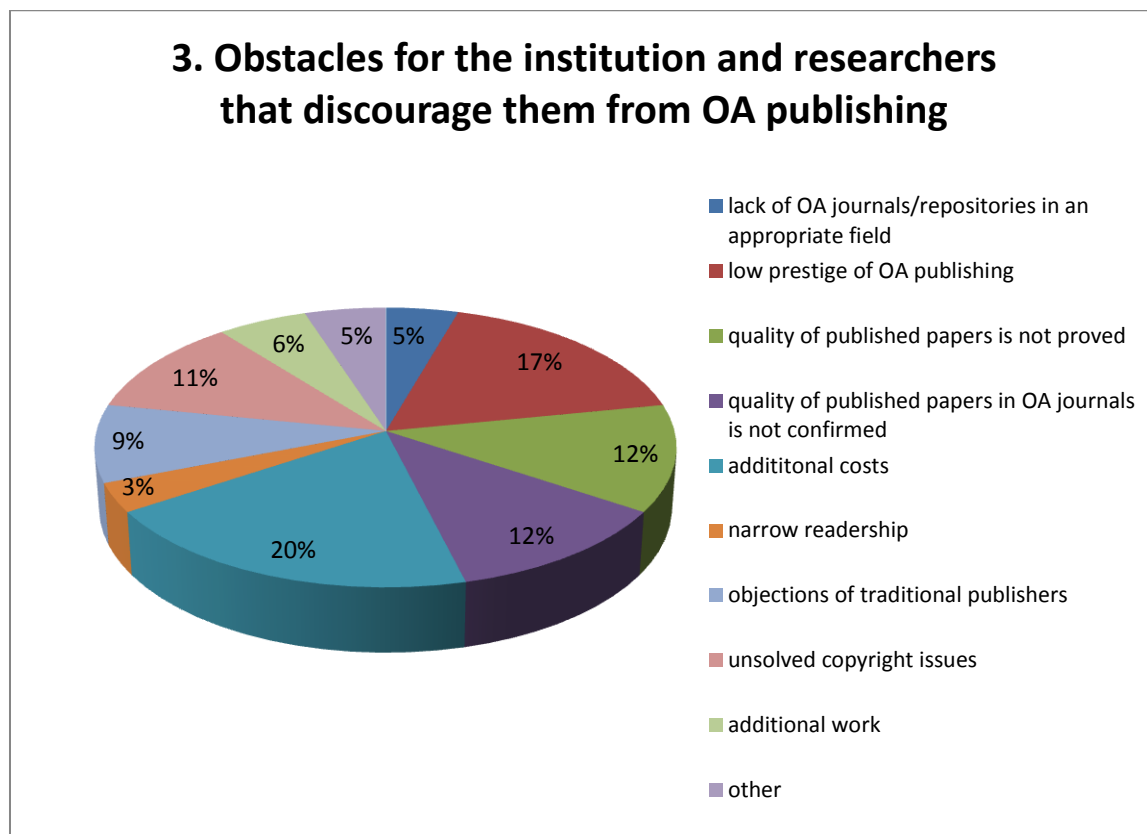
is to exploit the author-pays, Open-Access model for their own profit.” (Jeffrey Beall). Furthermore, such journals may put quantity before quality by accepting almost everything for publishing. They corrupt open access and promote unethical behaviours.

Considering the above factors, it is no surprise that researchers regard additional costs and the low prestige of OA publishing as the two significant obstacles that discourage them from publishing results open access. Many respondents feel that the quality of the published papers in open access journals is not proved and not confirmed. Such views on the quality of OA journals and on the works they publish will be overcome when researchers are educated about the open access publishing process itself. As soon as they get involved in the publishing process, they will see that open access journals do not differ from toll access journals in their commitment to peer review, but only in their business model, which has no bearing on the quality of the articles they publish. Quality-conscious open access journals use the same procedures and standards, including the same authors, editors, referees, as subscription-based scientific periodicals. Thus, researchers’ misconception about the quality of a given OA journal rather roots in the low profile of the publishing due to the recent introduction into the scholarly publishing realm. It has to be emphasized that low profile does not entail low quality (University of Rhode Island).

On the basis of the responses, another problematic question is related to the copyright issues of open access documents. The OA movement has triggered the debate on copyright in the scholarly communication system. All stakeholders in the system – publishers, academic institutes and libraries, authors – can adopt opposing positions in this debate. With the emergence of open access journal publishing new copyright models have been introduced. These copyright models oppose 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. Research shows that academic writers publishing in Open Access journals and gaining knowledge on copyright choices appear to be no longer satisfied with transferring copyrights to publishers.

Researchers consider the issue of the copyright 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. Copyright as a bundle of rights automatically assigned to the author(s) is a combination of moral rights (the most important of which is the right as an author to be properly and fully acknowledged) and exploitation rights. The debate focuses on the latter. In scholarly communication the exploitation rights ensure rather limited options: reuse for educational purposes and for commercial purposes. The right to reuse is essential in the OA models, since it gives free online access and permission to use the information for any responsible purpose. (Hoorn and Graf) 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 remade document.

Some respondents marked additional work as an obstacle in OA publishing. Additional work is performed in the green way of OA publishing with uploading the article into the institutional or research repositories. Researchers generally lose patience with file or data upload forms, either because they consider uploading manually articles one by one a waste of time, or because they do not know the protocols of how to upload all articles in one file. It is hard sometimes to convince them to follow and use the technological and cultural developments new trends offer.



#### 4. What types of research data are produced in your institution?

Research data can be defined as: “the recorded factual material commonly accepted in the scientific community as necessary to validate research findings” (OMB Circular 110). 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 be manageable for storing and reuse.

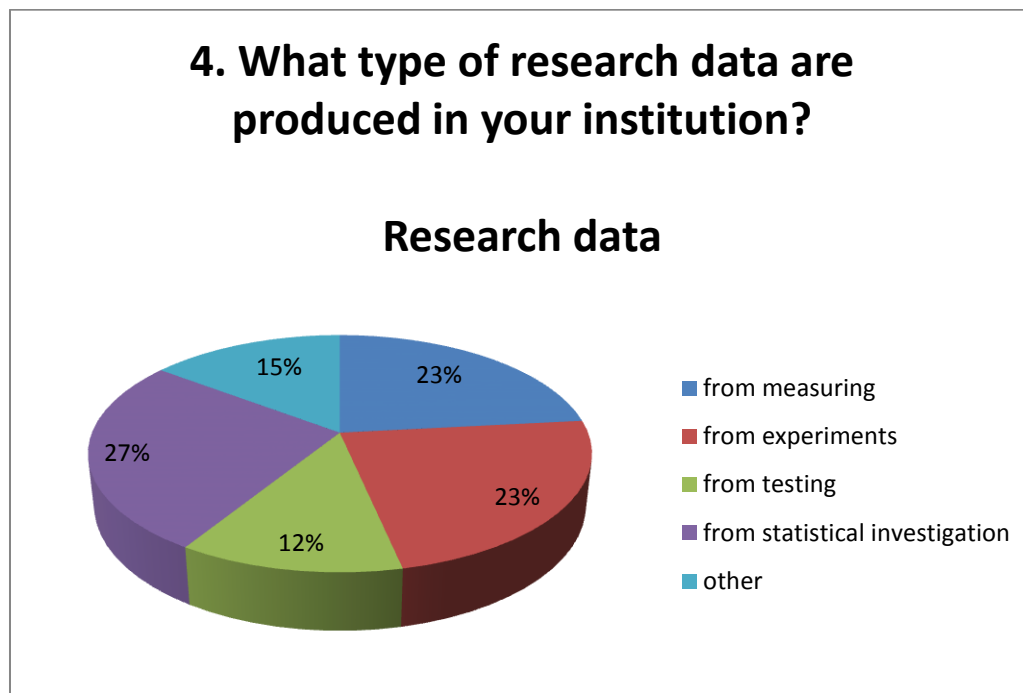
We can make a fundamental distinction between two types of data: **qualitative** and **quantitative**. Quantitative data are often described as information in numerical form, while qualitative data occurs mainly in text form. However, qualitative data could be much more than just words or text (e.g. photographs, videos, sound recordings and so on). Many researchers confute the sharp distinction between the two types since they are related to each other. All quantitative data is based upon qualitative judgments and all qualitative

information can be easily converted into quantitative, and there are many times when doing so would add considerable value to your research (Trochim).

Another categorization method divides research data into 5 categories, which all require different data management plans. Categories include observational data (captured in real-time, usually irreplaceable), experimental data (from lab equipment, often reproducible), simulation data (generated from test models), derived or compiled data (collection of smaller (peer-reviewed) datasets). Research data formats show much variety from text, spreadsheets, notebooks, reports, photographs, slides, and workflows to models and algorithms. (Boston University)

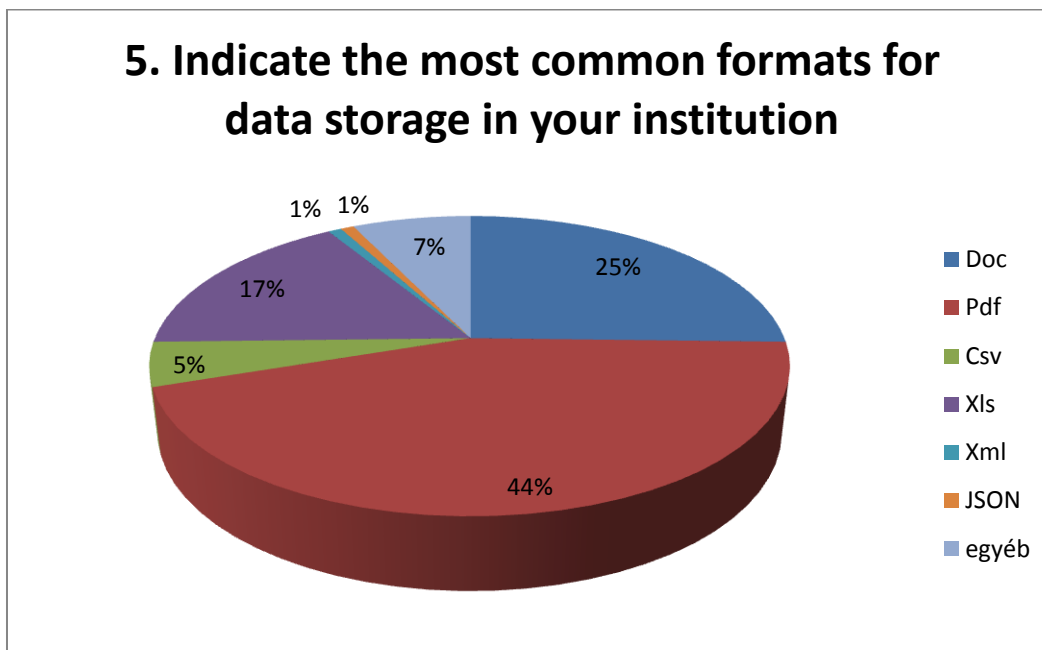
In case of an institutional repository where services have to be developed in order to meet the needs of researchers from different scientific fields, it is essential to get prepared to deal with such a variety of research data types and formats.

In the recent survey, the majority of respondents come from the fields of biology, sociology, mathematics, medicine, the research data here mainly are from testing, measurements and statistical investigation. These responses definitely do not give a well-rounded picture about the variety of research data an institution has to manage in research data management processes.



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

Responses indicate the most common storage form researchers prefer to use is PDF. The main reason for it is that PDF opens up the same way in any operational system without data loss or distortion. Doc files are also very popular. Inappropriate data storage leads to loss of data; however, most researchers do not think about using or reusing their own data. After publishing the research results, research data are not considered a priority to manage. Researchers should be educated about the methods of long term preservation of their raw data.



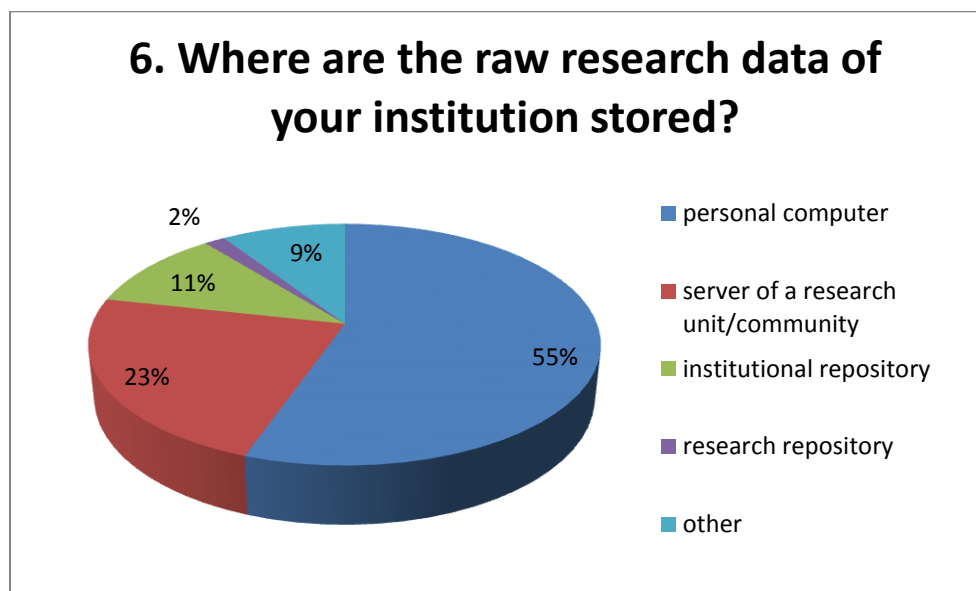
## 6. Where are the raw research data of your institution stored?

More than half of the respondents indicated that their research data are stored on personal computers. The practice of using a personal webstorage to archive research data carries dangers for the long-term preservation of data. Storing and backing up research data is a critical element of a research process. However, a simple saving is not necessarily sufficient to ensure the data's future usability. It is essential that after the project is completed, time and effort is taken to prepare an archived copy of the used research data. 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.

Servers of research units or departments are also reported by many respondents as a storage place of research data. At the University of Debrecen, there are departments in specific scientific fields, generating a lot of research data, which have set up local solutions to manage data. Either using a designated computer at the department or utilizing a cloud-

based storage facility, the data collected through research projects are managed and stored for reuse by PhD students and researchers of that department. Furthermore, research fields with a long tradition of research data management and practices of sharing data use specific databanks for long-term preservation.

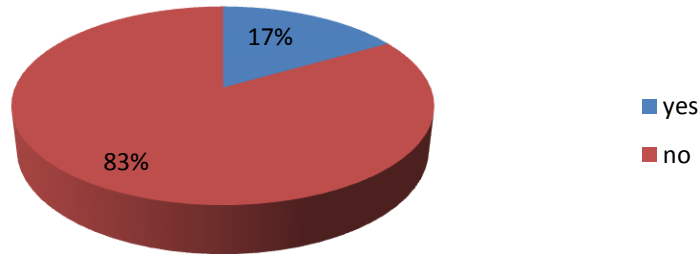
In case of the majority, choosing the personal computer as a primary storage facility indicates two things: (1) the archiving of research data is not organized on an institutional level, thus researchers have to manage data on their own, and (2) researchers are reluctant to share their data with others;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



### 7. Does your institution plan any changes in the way of storing and archiving the research data?

The responses indicate that researchers have no information about changes in the archiving policy of their institution. In the Hungarian version of the survey, the questions refer not only to the institutions, but also units of institutions and departments. This way, researchers may answer the questions in relation to their immediate research environment. In this regard, the responses imply that researchers are not informed about changes in data management even at the lower levels of the organization.

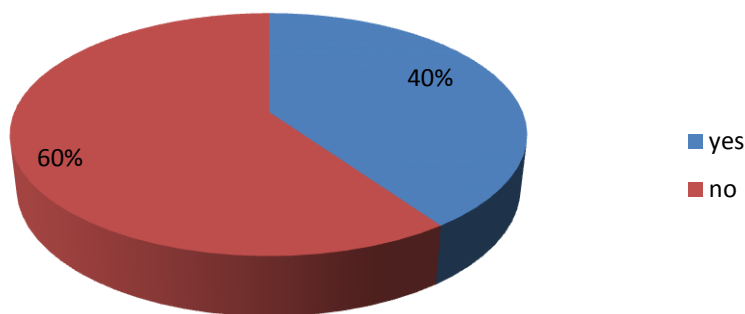
**7. Does your institution plan any changes for storing and archiving research data**



**8. Are the research data archived in your institution for more than 10 years?**

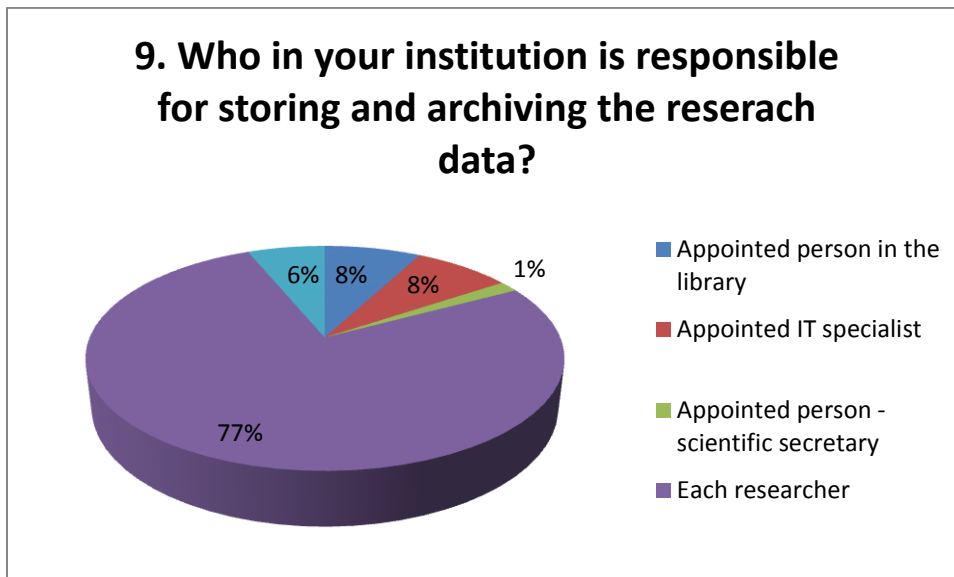
60% of the respondents do not have information about their institution's data management practices. The reason for this might be in the variety of practices of different research fields. As mentioned above, there are scientific fields which work with lots of research data, and therefore they have already developed data management practices, and there are fields which deal with a minimum amount of data, and therefore they do not have plans in this regards on an institutional level. The 40% positive answers imply, however, that where research data appear as a part of the research process, archiving practices have been developed for 10 years, at least.

**8. Are the research data archived in your institution for more than 10 years?**



### 9. Who in your institution is responsible for (takes care of) storing and archiving the research data?

As a continuation of question 8, the answers here imply that research fields and institutions where research data management is incorporated into the research flow do not necessarily have practices and policies developed for data archiving. 77% of respondents store their research data on personal computers. Only a small percentage of respondents work in an institution where an appointed librarian, IT specialist, or a scientific secretary takes care of data archiving for the researchers.

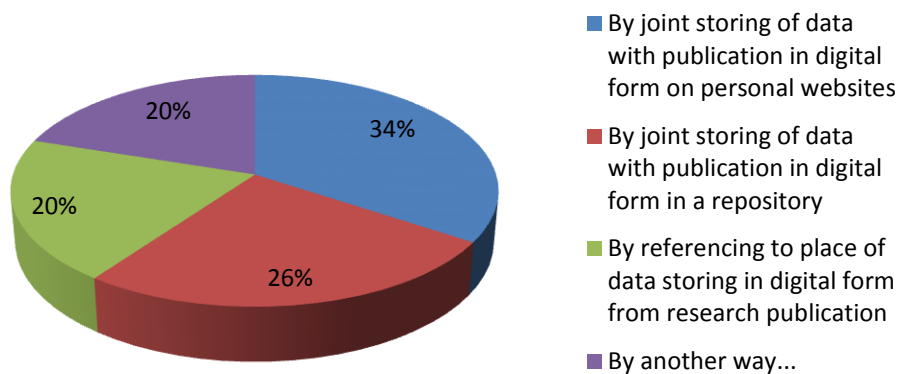


### 10. Are the research data produced in your institution linked with the research papers?

In this question again, the role of the personal computer in a research process overwhelms other options of publishing data in relation to research results, such as linking data to publications in a repository. Only half of the respondents answered this question, which means that real links between data and results through reference or actual online links are established only at a few institutions. Respondents also mentioned departmental servers or publishers' websites as possible locations where publications and data can be deposited and searched together. The answers show that some researchers and institution units have knowledge about research data publishing and management, and already have established storage locations and connections to research results. However, the majority of researchers do not deal with research data management, either because they do not have the infrastructure to rely on, or because their research does not produce data in quantities worth dealing with on an institutional level.



## 10. How are the research data linked with research papers?



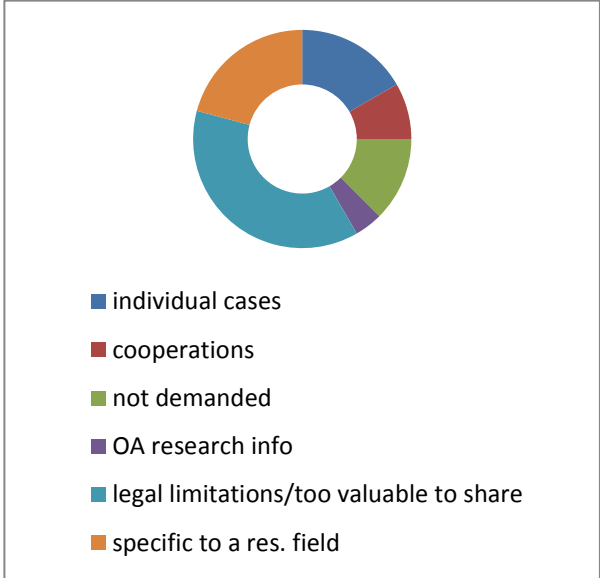
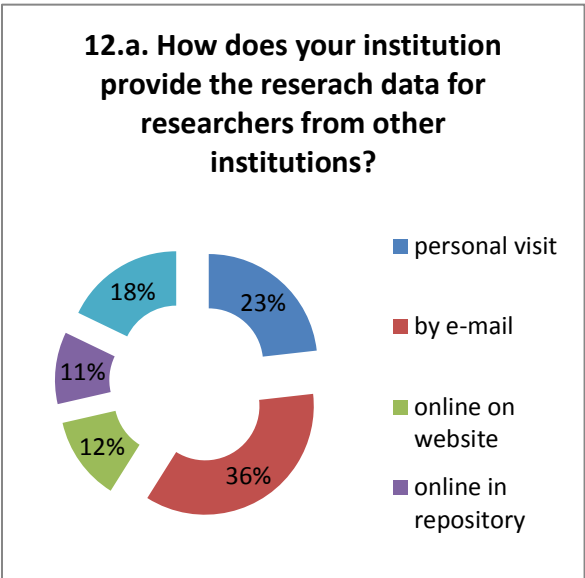
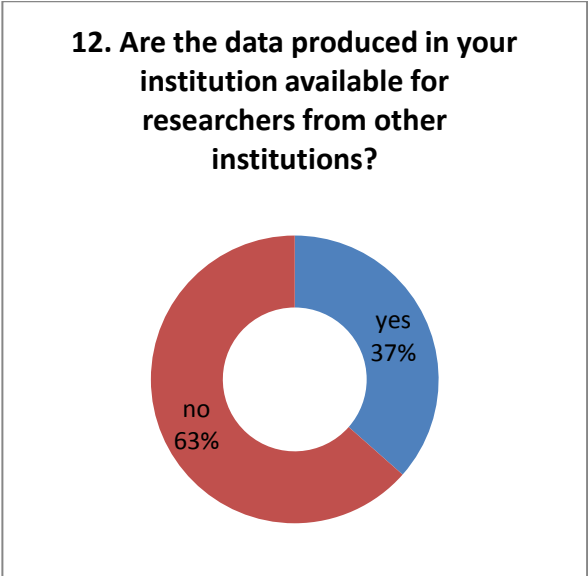
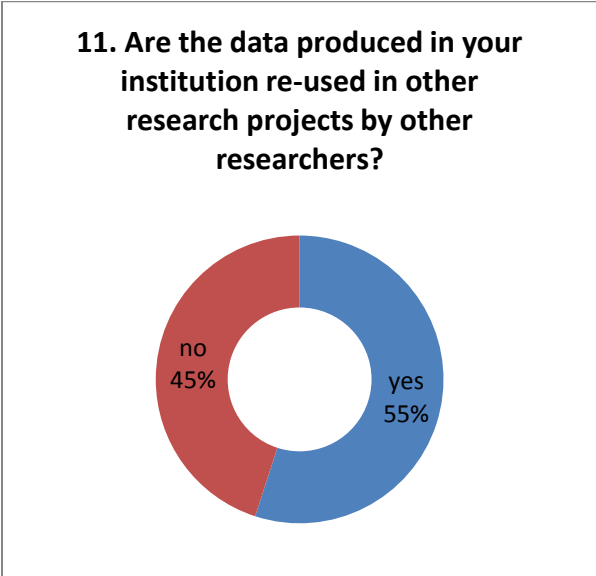
**11. Are the data produced in your institution re-used in other research projects by other researchers?**

**12. Are the data produced in your institution available for researchers from other institutions?**

Questions 11 and 12 inquire about the possible re-use of the research data within and outside of the institution. The responses show that collected research data, in more than half of the cases, are being re-used in other research projects. It is also implied by the answers that the re-use usually occurs within the same institution or department. Data are not often communicated outside of the immediate research community they are produced in. The reason behind this is mainly due to distrust with other researchers for misusing data or not crediting the producers appropriately. Researchers working in the same department or same research circles usually have faith in the fellow scientists to handle results and data in an ethical manner. The other reason for not sharing data with researchers from other institutions is basically that the data are way too valuable for them (answers to question 12.b confirm this presumption). Researchers use the same data sets for several research papers and articles or conference presentations. Sharing data with fellow scientists in the immediate working circles may result in co-authored articles or in further research leading to more co-authored results. This possibility is not necessarily ensured in case of cooperation outside of the institutions or immediate research groups.

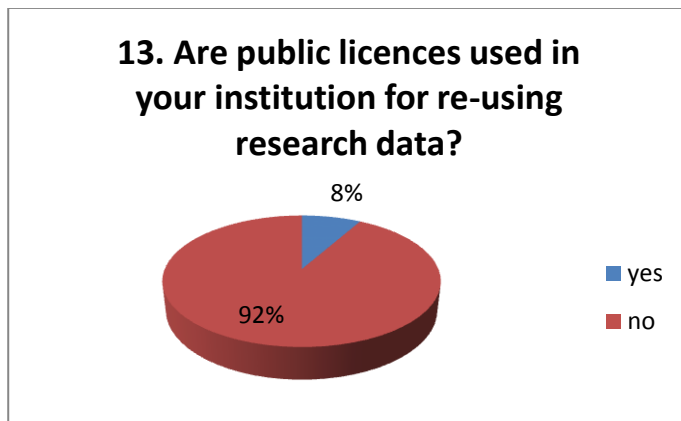
Sharing is also often undermined by the uncertainty of legal issues. Many researchers are not comfortable sharing information because of the unsecure transmission of data or due to the fear of improper data management by other researchers. Email and personal contact is still preferred to transmit information. In addition to traditional research environments and facilities for collaborations, virtual research environments have emerged which facilitate

data transfer, storage and providing access to data to various partners and institutions. However, while cloud-based file sharing services may be suitable for sharing certain types of data, they are not fit for managing sensitive and confidential data. That is why it is important to develop an infrastructure on an institutional level to ensure the long-term preservation of research data and the secure transmission of information among researchers. Besides the technical base of data storage, the legal background of data management should also be developed, so researchers may have a legal safety net to fall back on or use in their projects.



### 13. Do you use public licenses in your institution for providing and marking the research data?

Responses to Question 13 show that institutions do not usually use public licenses to protect research data. Researchers do not take a proactive stance in copyright issues; for example, they consent to full transfer of copyrights to the publishers in order for their articles to appear in a prestigious journal and they do not include copyright protection in their data management processes.



Data management plans (DMP) are created in order to standardize the processes involved in protecting data covering the collection, storage, transfer, and sharing of data. There are numerous advantages of developing and using a DMP:

- ensures continuity if project staff leave or new researchers join;
- prevents duplication e.g. re-collecting or re-working data;
- the data underlying publications are maintained, allowing for validation of results
- leads to more collaboration and advances research;
- ensures visible and greater impact;
- researchers gain credit for the data collection; (DCC)

As information on research data management is communicated to researchers, they become more knowledgeable about the advantages and necessity to use DMPs in their research processes. Due to increasing pressure from funders to comply in this question, researchers are faced with the urgency of the issue and institutions are advised to develop the technological and social/ legal infrastructures for data management.

### **Conclusion**

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 include the

main principles of open access in their publishing policies or even introduce open access mandates into their operational procedures. Besides funders' requirements, a growing number of publishers, especially open access ones, demand the free accessibility of research data in connection to the research result 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 available and processible for readers.

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

## **Czech Republic**

NTK conducted a 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. 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 individual faculties of colleges and universities were addressed. Where a smaller college or university 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 such representatives of these institutions that we expected to have a comprehensive knowledge of research activities of their institution. We chose persons 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 when it was not possible to determine who holds such posts in the institution or no such posts have been set up by the institution, representatives from the institution's management were selected.

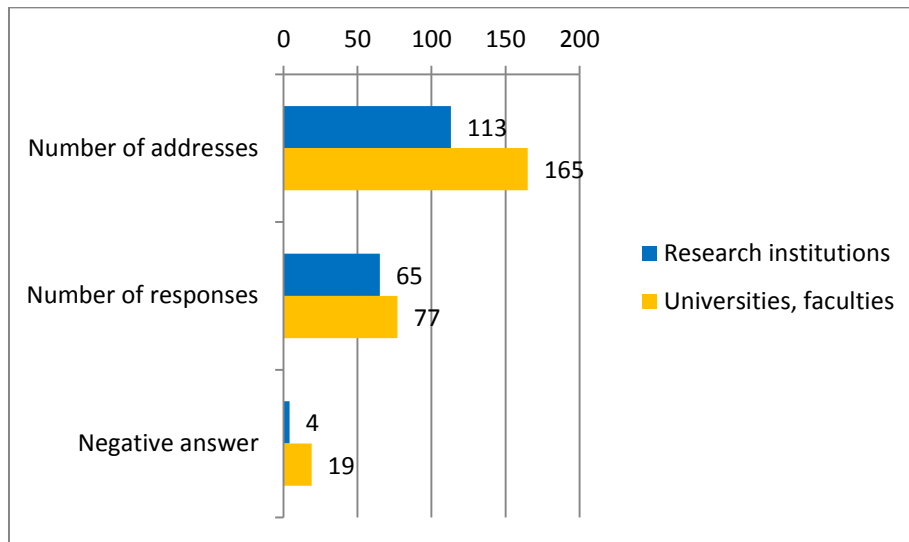
The survey was carried out in the form of an online questionnaire, consisting of 25 questions, including questions identifying the institution – the name of the institution, the respondent's e-mail, etc. A structured questionnaire was used, where the majority of questions were close-end questions with the possibility to choose one or more answers. There were several questions with a yes – no tick box option. Once a question was answered, another question was displayed, depending on the previous one. Multiple answer questions were marked accordingly. The online web application Vyplňto.cz was selected as a way of disseminating the questionnaire.

Selected representatives of the institutions received an informative email with a request to complete the questionnaire. The email contained information on the notion of enhanced publication, the aim of the survey as well as the use of the data obtained. The letter also included a request asking the addressee to pass the information and the link to the questionnaire on to some of their colleagues in case they cannot complete the survey themselves.

At first, we received only a small portion of completed questionnaires in both survey rounds; therefore we contacted the respondents with the same request again by phone.

In the first round, we addressed 113 research institutions and obtained 69 responses, including 65 completed questionnaires. Four institutions refused to or were not able to complete the questionnaire. The questionnaire response rate was 57.5%.

In the second round, we addressed in total 165 faculties of 26 Czech public and state colleges and universities. We obtained 96 responses, including 77 completed questionnaires, while 19 representatives of the faculties informed us that they were either too busy to complete the questionnaire or could not complete it. Questionnaire response rate was 46.6 %.



#### The results of the questionnaire - overview

There were various reasons behind the decision not to complete the questionnaire. The most frequent ones were secrecy or the protection of research information. The representatives of some of the institutions did not want to provide any information without giving any 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; although they are engaged in research, they form the summary of their research findings directly into published outputs. Other institutions stated that they did not have any research data since they only work with textual documents. One institution stated that it was not directly engaged in research, but rather in preparation of supporting materials for research for other institutions, i.e. it does not have research data of its own. This problem of a lack of raw research data primarily concerned research institutions focused on the humanities and, as far as colleges and universities are concerned, art colleges and faculties of arts or law.

The response rate was by 10.9 % lower for colleges and universities, as the answer that they do not have either time or research data was more frequent.

### *Survey of enhanced publications*

What follows are the four questions dealing with publication of research documents in Open Access mode that were addressed only to public and state colleges and universities in the second survey round.

*Question no. 1: Are the research papers of your institution stored and accessible in open archives and digital repositories?*

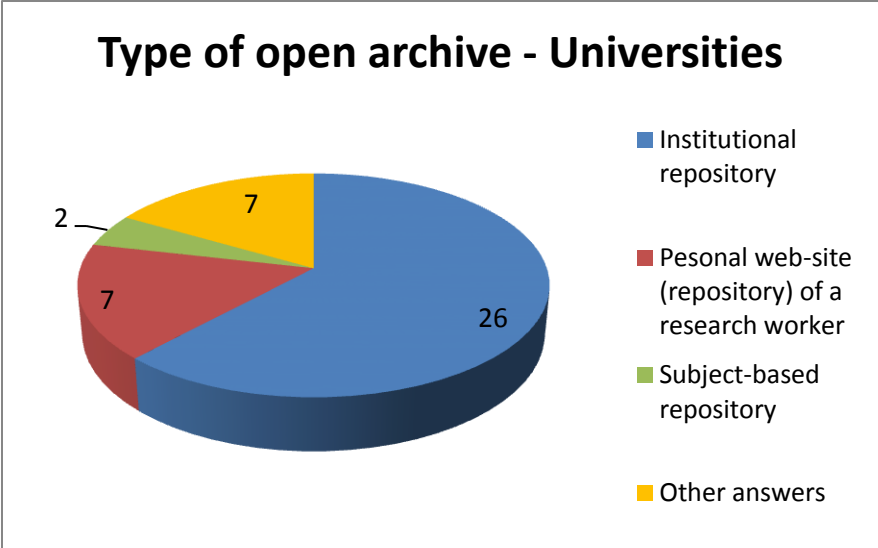
Yes, it is obligatory	6
Yes, though it is not obligatory	26
No	36
Hard to say	9

77 respondents, one answer possible

The results show that only close to half of the respondents (faculties of colleges and universities) that completed the questionnaire follow the practice of making research publications available in open repositories. The fact that some of the colleges and universities and their faculties do so without such requirement being imposed by an internal directive or decree is encouraging. The respondents that stated that making research publications available in open repositories was obligatory for them were asked to add information as to when the obligation was introduced. The answers differed, but the majority indicated that this obligation was introduced in the past 10 years.

*Question no. 2: What type of open archive (repository) is used most frequently? (Answered only by those who answered "Yes" in the previous question no. 1)*

Only respondents that responded that they store data to in any open archive answered this question. The respondents answering this question could choose more than one of the suggested answers or provide their own answer. The most common way of making research publications available in Open Access mode is the use of institutional repositories. It is followed by personal websites of scientists, researchers. Subject-based repositories are quite rare.

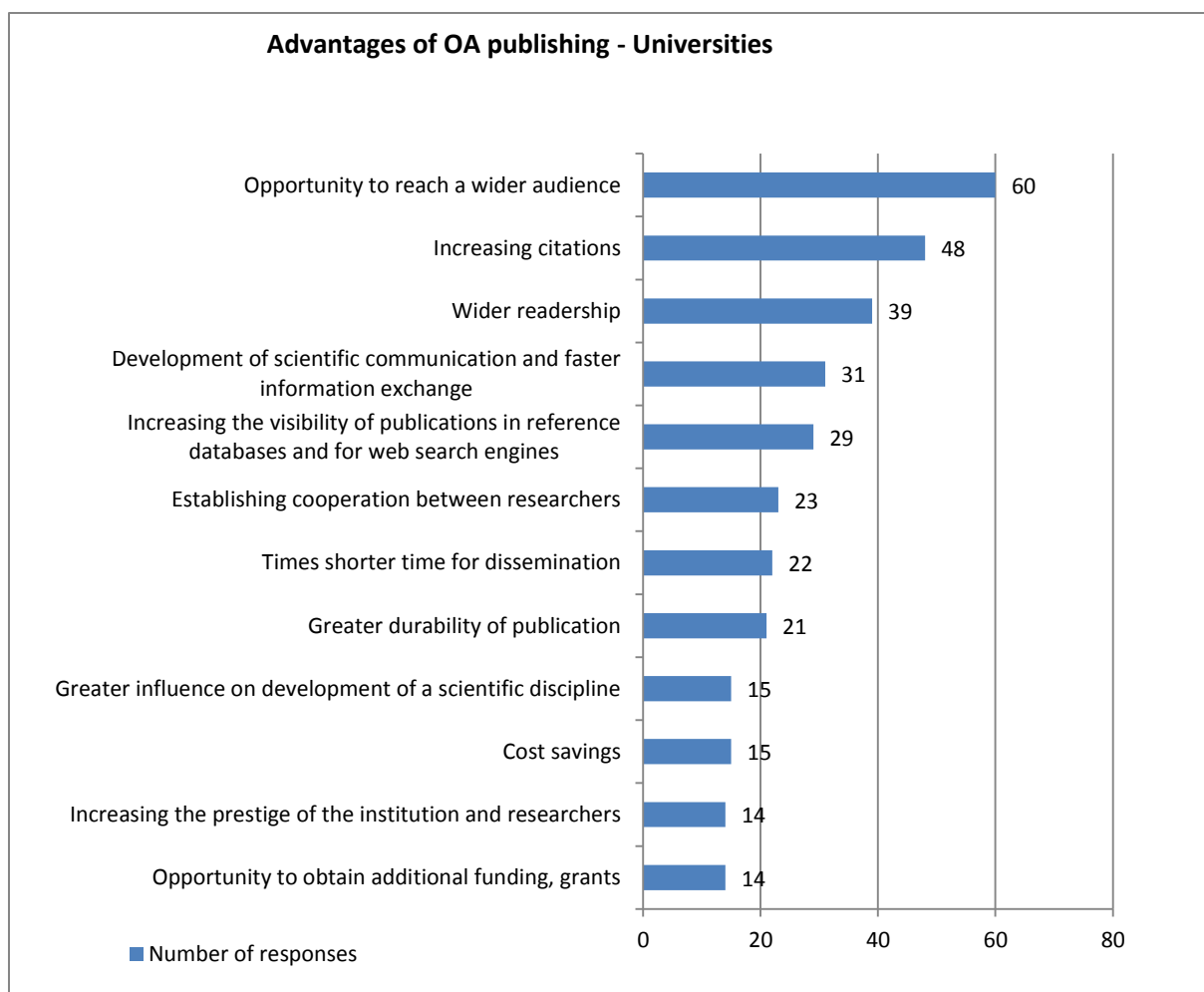


Used open archives at universities  
 32 respondents, multiple choice, possibility of own answer

*Question no. 3: What are/could be the main advantages of publishing or storing in Open Access for your institution and researchers?*

The respondents could choose from 12 given answers. The advantages that were highlighted most often include the possibility to reach a wider audience, achieve higher citation and readership rates and faster exchange of information among scientists.





Advantages of Open Access publishing  
77 respondents from universities, multiple choice

*Question no. 4: What are/could be the main obstacles for your institution and researchers that discourage them from publishing or storing research papers in OA model?*

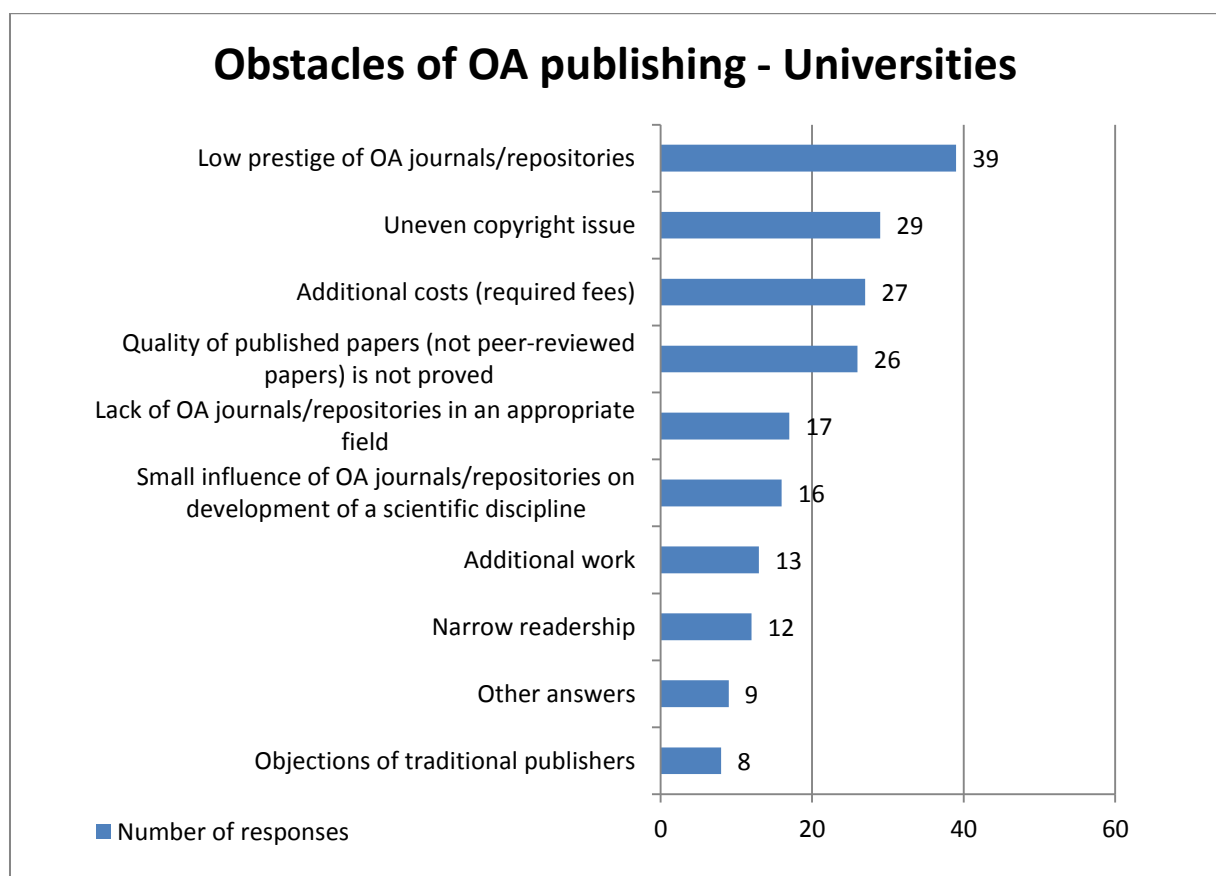
Respondents could choose from 9 given answers and also indicate their own reason. The most frequent obstacle was low prestige of Open Access journals and repositories, followed by copyright issues and fees connected with Open Access publishing. Therefore, it is evident that finance and prestige of journals and repositories still play a major role in scientific publication.

As to respondents' own answers, complications in reporting research activities for the Information Register of R&D Results (RIV)<sup>1</sup>, which serves as a basis for allocation of points to authors affiliated with the relevant institutions, clearly have a substantial impact. Every year,

<sup>1</sup> The Information Register of R&D Results, part of the Czech Information System of Research, Experimental Development and Innovations. It is a public administration information system securing collection, processing, provision and use of information on publicly funded research, development and innovations (<http://www.vyzkum.cz/FrontClanek.aspx?idsekce=610>).

the value of a point in CZK is defined and the more points an institution obtains, the greater amount is received for the institution's budget. Points are allocated for different research outputs. Due to this setup of the Czech system of evaluation of science and research, the authors of research publications focus on publications evaluated with points. The outcomes published in Open Access mode are usually not assigned "points" in RIV and thus they are not the primary preference; authors focus on publishing in major scholarly journals and resources as they receive more points for their results published in these resources.

Further, the respondents indicated copyright issues and concerns about copyright disputes with journal publishers, complications with ensuring a license for the National Library as far as legal deposit are concerned, but also wider issues of protection of intellectual and industrial property rights. Other reasons lied in insufficient organizational measures within the college or university or project obligations, when it is not possible to freely publish and disseminate all research results. There was a positive reaction to the general opinion that the quality of Open Access publications and repositories is usually not verified as there is no peer review procedure in most cases. A respondent described his personal experience: "I completely disagree with the statement in Question 4 that OA journals usually do not apply a peer review procedure. In all these journals in which my colleagues and I published or attempted to publish our papers there was a peer review procedure at least as rigorous and demanding as in traditional journals."

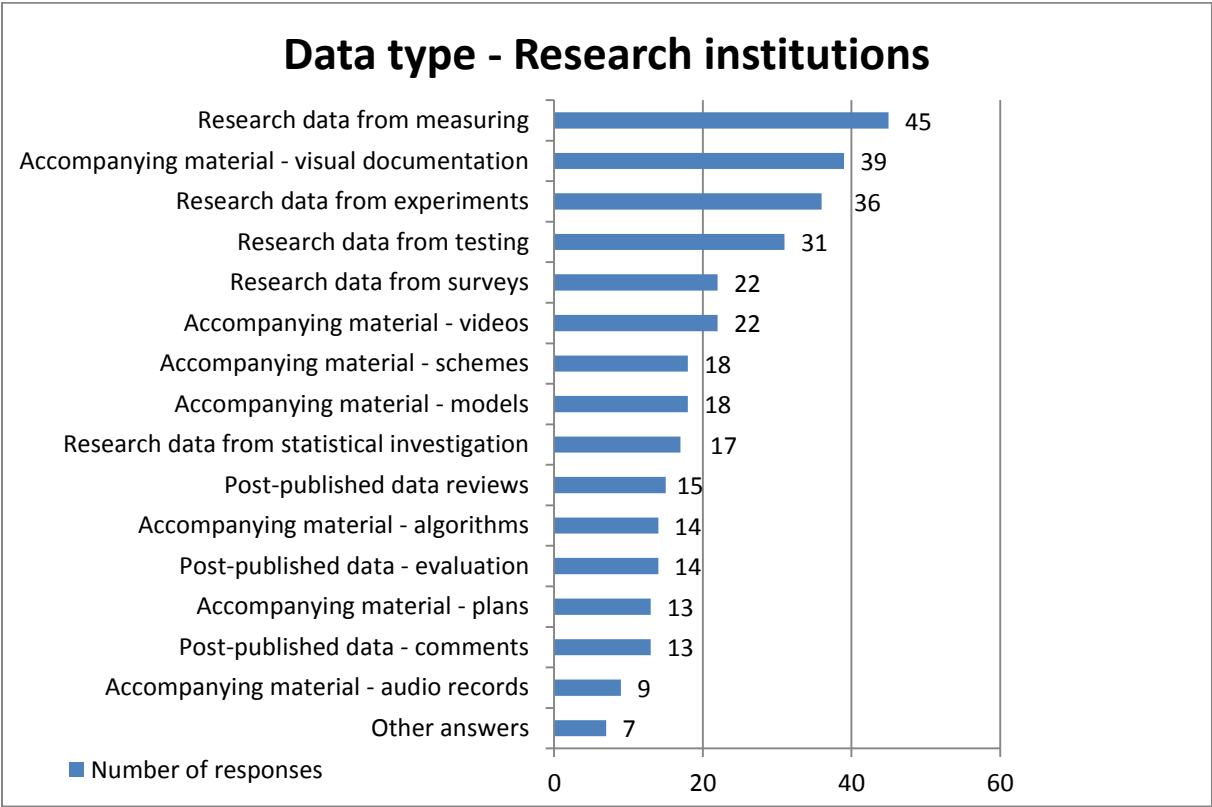


Obstacles of Open Access publishing  
 77 respondents from universities, multiple choice, possibility of own answer

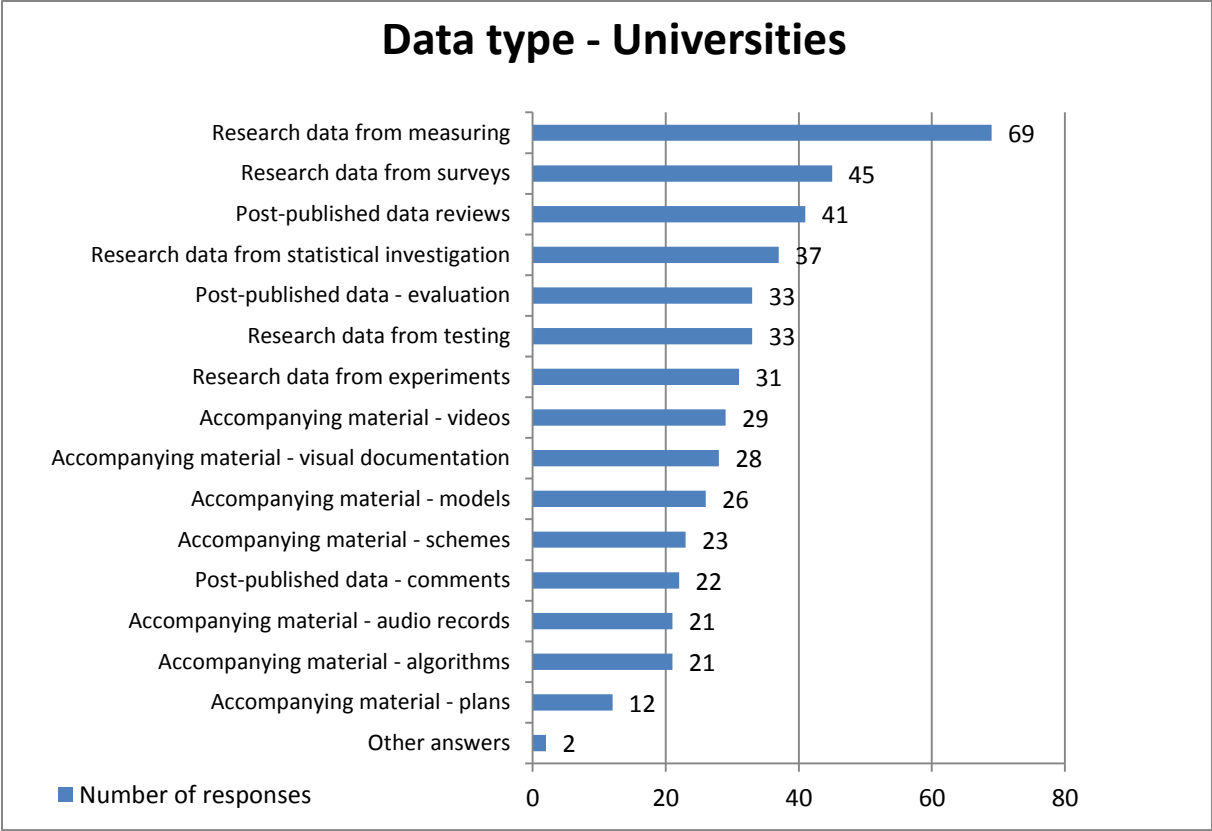
A crucial part of the questionnaire dealt with research data, their storage, archiving and access to them. There were altogether seventeen questions. These questions were addressed both to research institutions and faculties of public and state colleges and universities. In these questions, a comparison of the results for research institutions on the one hand and for colleges and universities on the other hand was made.

*Question no. 5: What types of research data are produced in your institution?*

An important aim of the questionnaire was to identify the types of enhanced publications that are produced in research. Respondents could choose more than one of the answers provided, i.e. several data types that they produce at their institution. In both groups, measurement data were the most frequent answer. Apart from them, research institutions indicated other standard research data – data from experiments, testing, surveys, etc. as the prevailing type of data. On the other hand, colleges and universities identified post-publication data (reviews, evaluations) as the predominant type of data, which were not common at research institutions.



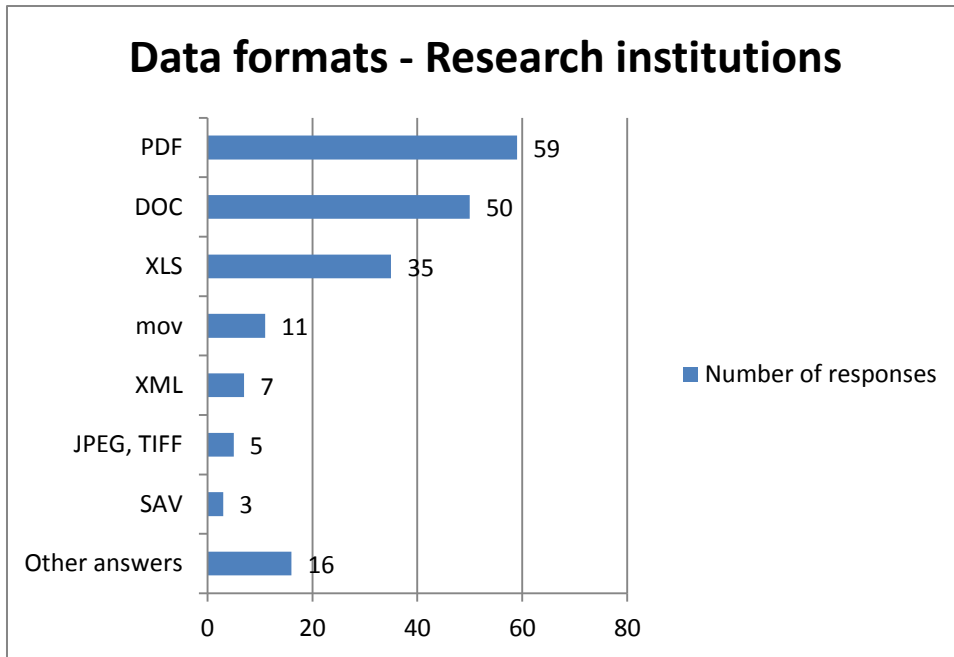
Types of data produced by research institutions  
 65 respondents, multiple choice, possibility of own answer



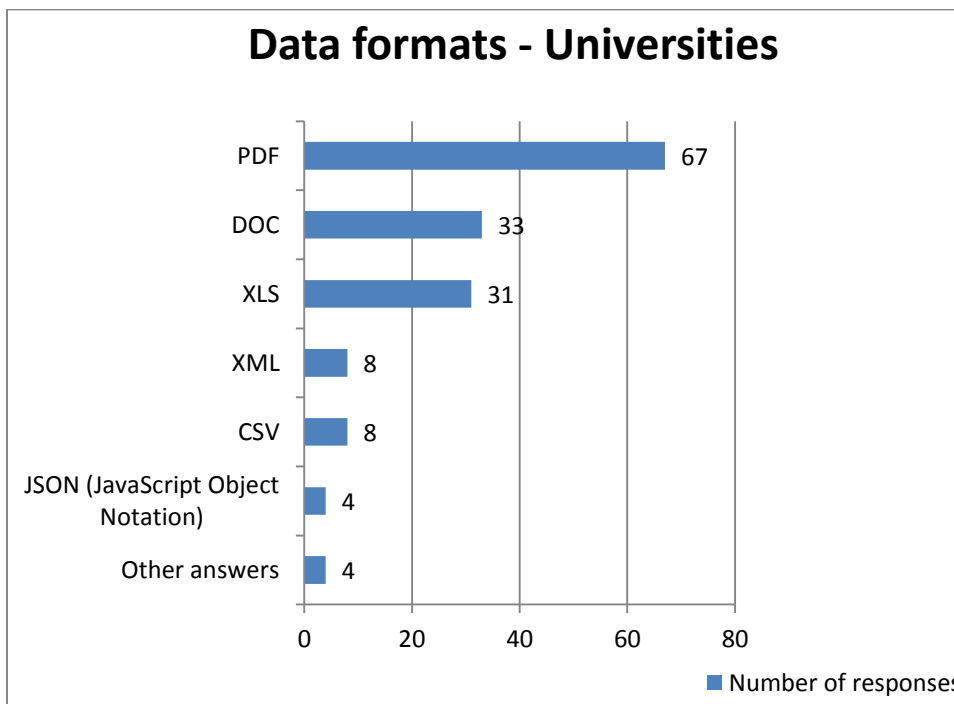
Types of data produced by universities  
 77 respondents, multiple choice, possibility of own answer

*Question no. 6: Indicate the most common formats for data storage in your institution.*

The most common format of data was, unsurprisingly, PDF, a universal format for file storage. Next were the DOC format and spreadsheet formats XML and CSV. Respondents could indicate also other data formats produced by them. These answers were very diverse and cannot be generalized. In addition to relatively widespread TXT, XML formats and audio and image formats, e.g. MOV, SAV, AVL, JWG, LPK, CIF, DAT and other formats were mentioned.



Data formats produced by research institutions  
65 respondents, multiple choice, possibility of own answer

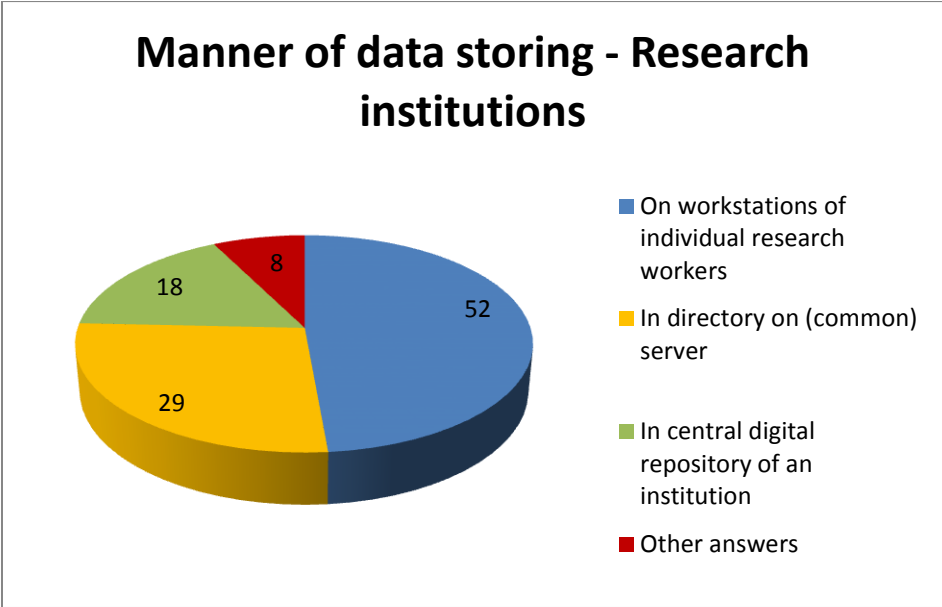


Data formats produced by universities  
77 respondents, multiple choice, possibility of own answer

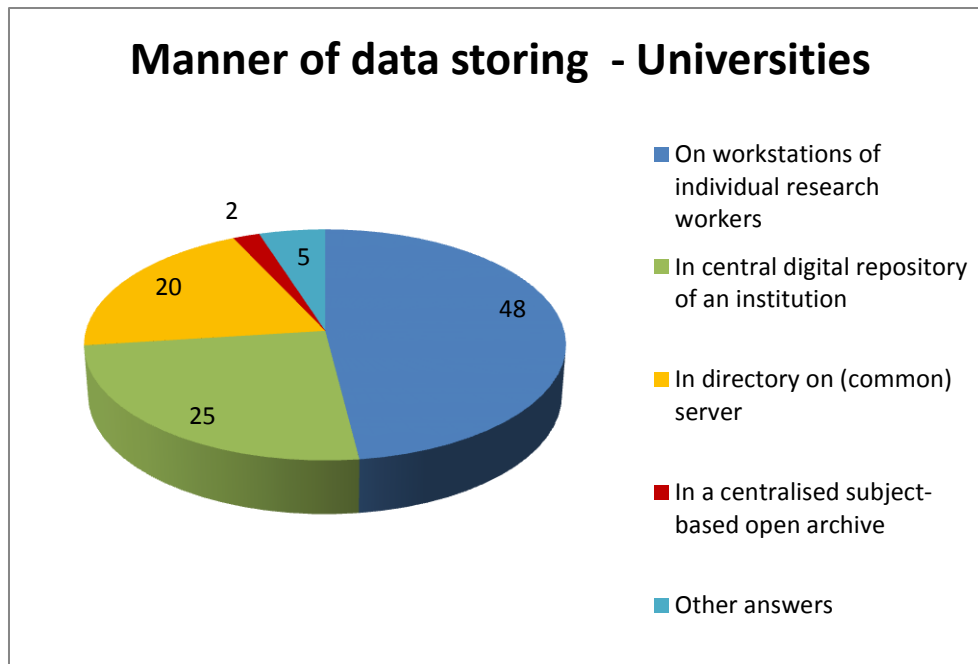
*Question no. 7: Where are the raw research data of your institution stored?*

This question was one of the most crucial ones as it bears witness to the attitude of the institutions to data archiving. Unfortunately, both surveys show that the majority of institutions do not have any centralized solution and data are most often stored only on

individual researchers' workstations. This approach does not guarantee security and long-term access to data. It may happen that important data disappear as computers, discs and technologies are replaced or when an employee leaves the institution. Yet, almost half of the respondents indicate this approach. Data are often stored in files on shared servers. This method ensures a higher degree of data protection and their centralized management, but does not result in improved work with data, their searchability and accessibility. 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. The option of storing data in a subject-based repository was selected only by a negligible number of respondents.



Manner of data storing in research institutions  
65 respondents, multiple choice, possibility of own answer



Manner of data storing at universities  
77 respondents, multiple choice, possibility of own answer

*Question no. 8: Does your institution plan any changes of the way of storing and archiving the research data?*

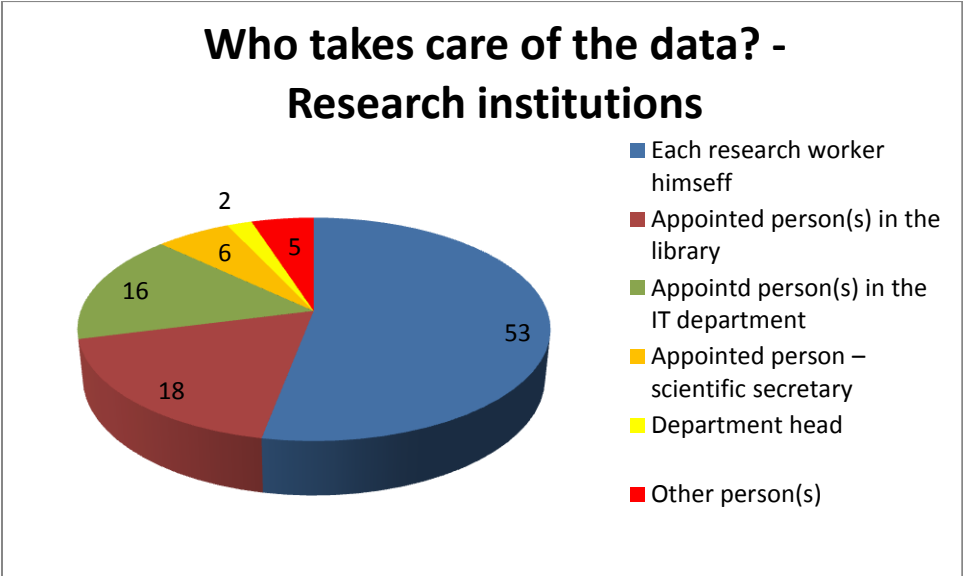
Most respondents both from research institutions and faculties of colleges and universities responded that they are not planning any changes in their present method of storing data. However, where the institutions are planning any changes, building a central digital repository was mentioned most frequently by the respondents. The situation in respect of the use of digital repositories is improving in particular at colleges and universities. Repositories are mostly used for archiving and providing access to thesis, publication activities of the employees and research publications and other research documents.

*Question no. 9: Are the research data archived in your institution for more than 10 years?*

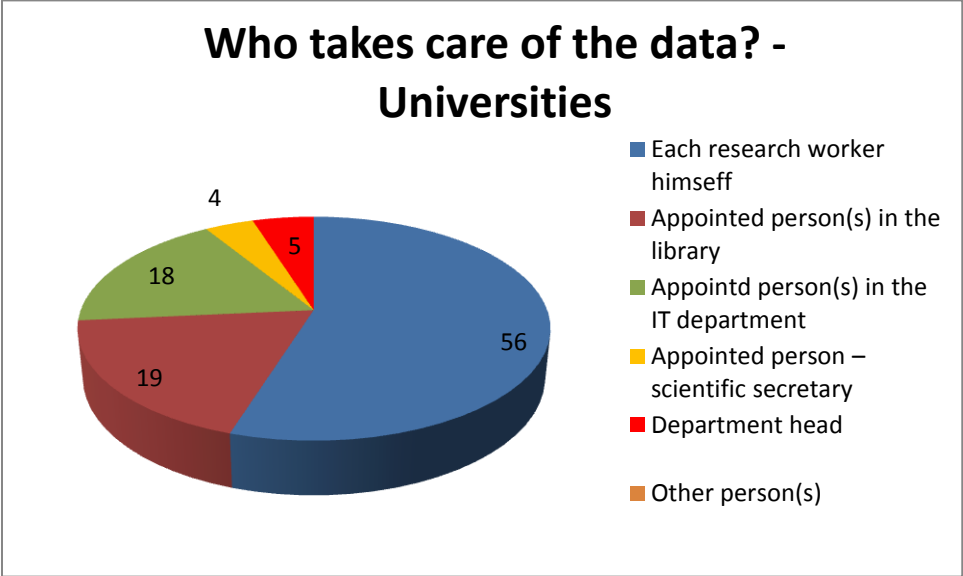
Permanent archiving of data might be already taken for granted at the present moment. Nevertheless, 30 representatives from research institutions and 33 respondents from colleges and universities stated that this was not the case, or they did not know the answer. About the same number of respondents confirmed long-term archiving. Where respondents stated that they had so far not archived data for ten years, but had already started long-term archiving, they were also asked to indicate how many years ago they started archiving. These answers were diverse, from 3 to 8 years. Some respondents also pointed out that it depends on data type.

*Question no. 10: Who in your institution is responsible (takes care) for storing and archiving the research data?*

The aim of this question was to determine what organizational arrangements were adopted to secure data storage. Once again, respondents were offered several options and could choose more than one answer or give their own answer. The answer that the researchers themselves take care of the data was clearly most frequent. This again testifies to the fact that the majority of Czech institutions engaged in research do not have any centralized data storage and archiving policy. Where there is a particular person or department authorized to take care of research data, it is usually the library, the IT department, scientific secretary or the head of the department or other organizational units. Some answers mentioned the archive of the institution and its staff, but it is not certain whether this really concerns digital data.



Persons responsible for data storing in research institutions  
65 respondents, multiple choice, possibility of own answer





Persons responsible for data storing at universities  
77 respondents, multiple choice, possibility of own answer

*Question no. 11: Are the research data produced in your institution linked with the research papers?*

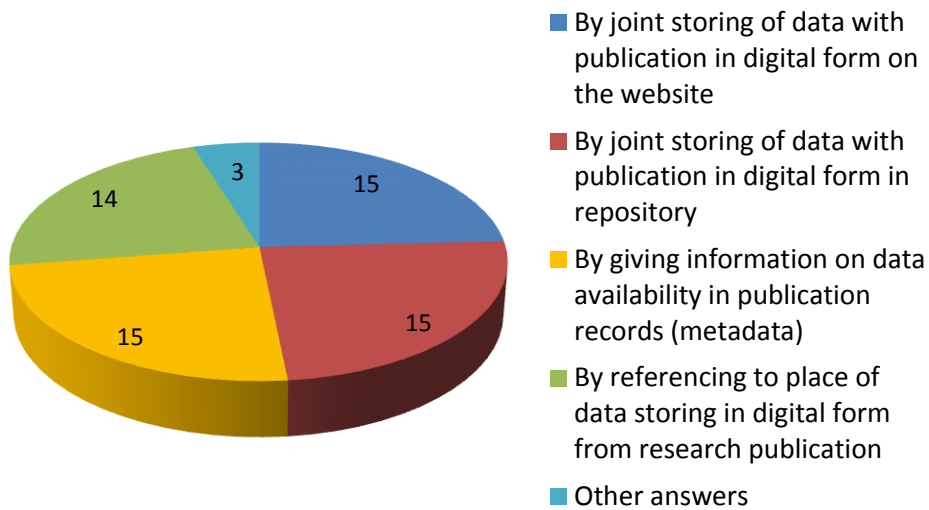
This was the key question of the entire survey. Using it as a basis, we can determine whether it is appropriate to talk about enhanced publications in the Czech Republic, even though the term has not been well established yet. The respondents received an explanation of the notion of enhanced publication, scientific publication and possible linking methods. We were pleased to find that an absolute majority of respondents responded in both questionnaire surveys that they linked research publications with research data in some way or other. In case of research institution this concerned 38 respondents out of 65 and 42 out of 77 respondents in case of colleges and universities.

*Question no. 12: If yes, how are the research data linked with research papers?*

*(Answered only by those who answered "Yes" in the previous question no. 11)*

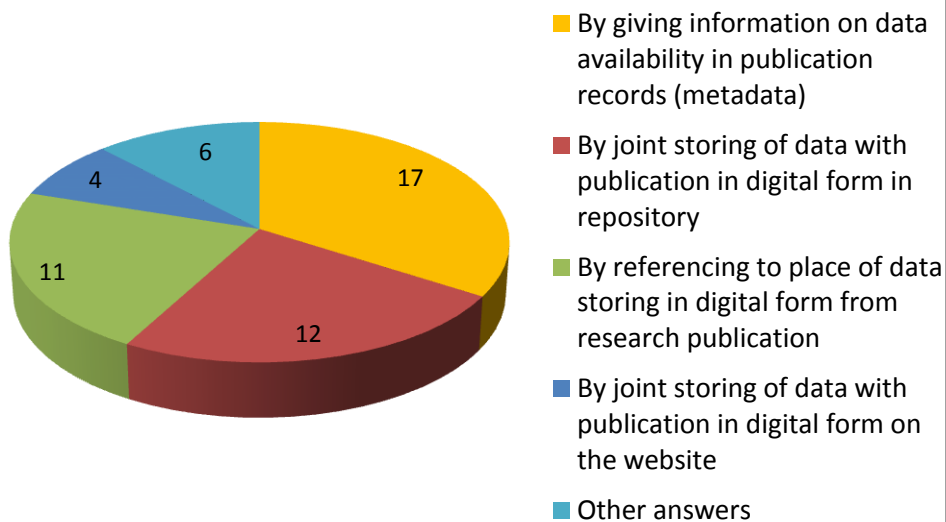
Several options of linking data to publications were suggested in the questionnaire. An ideal option is linking a research publication (research report, scholarly article, methodology, etc.) with raw data or other materials from research in a digital repository. This solution allows the user an easy way of working with the publication and data; everything is available at one place, while remote access is also possible. It is positive news that this option was the second most frequent in case of colleges and universities. Recording information about related research data in a bibliographic record of the research publication was the most frequent method of linking research publications and data at colleges and universities. This solution ensures that the user will learn about the existence of supporting documents from research, but it also depends on what kind of access to the very data is provided. If the user learns about the existence of the data, but has no opportunity to access them, it is not an ideal situation. The representatives of research institutions indicated the options of linking in a digital repository, joint storage of publications and data on the web, and information on availability of data in a research publication record in an equal number of responses. For representatives of colleges and universities a link from a publication to data in digital form was the third most frequent option; joint storage and linking of both in an electronic version on the web were selected less often.

## Manner of data linking - Research institutions



Manner of data linking in research institutions  
38 respondents, multiple choice, possibility of own answer

## Manner of data linking - Universities



Manner of data linking at universities  
42 respondents, multiple choice, possibility of own answer

*Question no. 13: Are there any plans for linking data with research publications in your institution?*

*(Answered only by those who answered "No" in question no. 11)*

Respondents who selected the answer that they do not link data with publications were asked to indicate whether they are at least planning this practice. A positive response was received only in a few cases, 5 respondents out of 27 from research institutions and 8 out of 35 from colleges and universities. Thus, these institutions are unlikely to seek improvement of this situation on their own initiative. The situation might be improved by greater public education, examples of good practice from abroad, promotion of the Open Access approach in the field of research, as well as the creation of a digital repository for a number of institutions that would enable depositing of both research publications and research data and their linking.

*Question no. 14: Are the data produced in your institution reused in other research projects and by other research workers?*

An overwhelming majority of respondents in both surveys (86%) confirmed that they reuse data from previous research projects as a basis or material for further research. This result was not surprising. Thus, it would be all the more appropriate if the institutions adopted a better way of storing, archiving and providing access to data. When data are stored in a well-organized, structured repository and thus they are easy to search and find, it is easier for the researchers to work with them and it saves their time.

*Question no. 15: Are the data produced during the research work in your institution available for researchers from other institutions?*

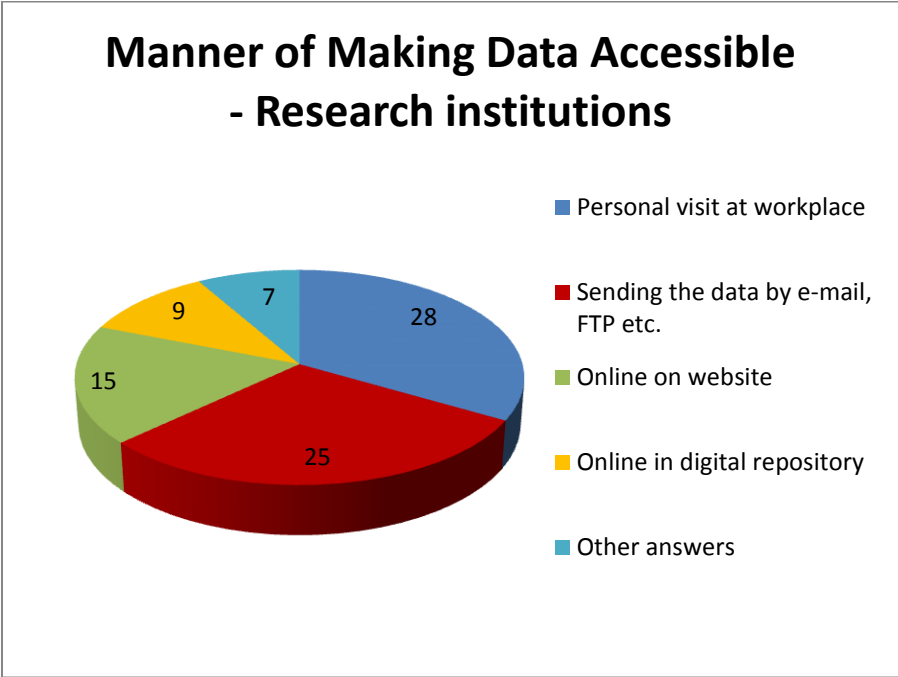
This question was used to find out whether research data are kept only for the purposes of the institution where they originated or whether research institutions are willing to provide them also to other researchers as part of science and research collaboration. 44 respondents out of 65 from research institutions and 41 out of 77 from colleges and universities stated that they are willing to provide data also to scientists outside their institution. They provide them in different ways, which is the focus of the following question.

*Question no. 16: How does your institution provide the research data for researchers from other institutions?*

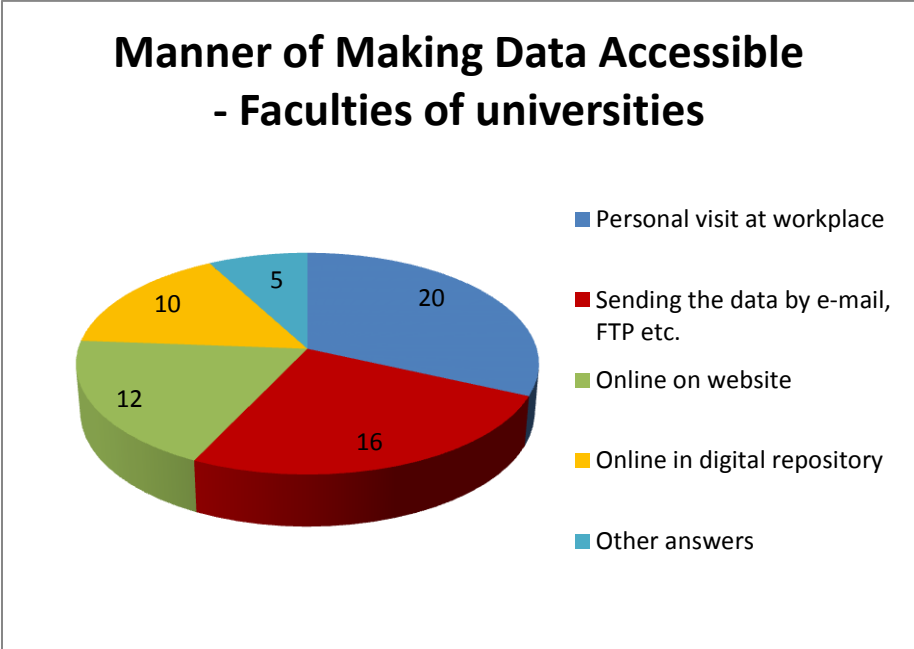
*(Answered only by those who answered "Yes" in the previous question no. 15)*

Respondents were offered various prepared answers; they could select more than one option as well as give their own answer. Both research institutions and colleges and universities indicated a personal visit of the individual interested in provision of data in most cases. Data are passed to the individual in person or provided for on-site study. The second most frequent option – sending of data by electronic mail or remote electronic data transfer (FTP) – seems to be more user-friendly. Online provision of data, either on the web or through a digital repository, was selected by a considerably smaller percentage of

respondents. Another option selected by respondents was that they publish data as a printed annex to a research publication, journal article or as part of presentation at scientific conferences.



Manner of data linking in research institutions  
44 respondents, multiple choice, possibility of own answer



Manner of data linking at universities  
41 respondents, multiple choice, possibility of own answer.

*Question no. 17: Why are the research data not available for the researchers from other institutions?*

*(Answered only by those who answered "No" in question no. 15)*

We were also interested in the reasons why institutions do not want to provide research data also to those interested outside the institution. Only respondents that responded that they did not provide data to any parties interested answered this question. There were several main reasons. Some institutions stated that they published only final official research outputs, i.e. research publications, articles in scholarly journals. These may include part of supporting materials connected with research, including data. In addition, research results are also presented at conferences and consequently in conference proceedings.

Private research institutions often cannot provide data due to commercial and copyright reasons. As far as tailored research, research made to order, is concerned, data become the property of the client, and it is therefore not possible to disclose them to other parties interested without the consent of the owner.

Copyright issues were mentioned also by other institutions. They treat research data as the property of the authors or the institution. Disclosure is sometimes also banned by agreements with partners and cooperating institutions; data are available only to cooperating partners.

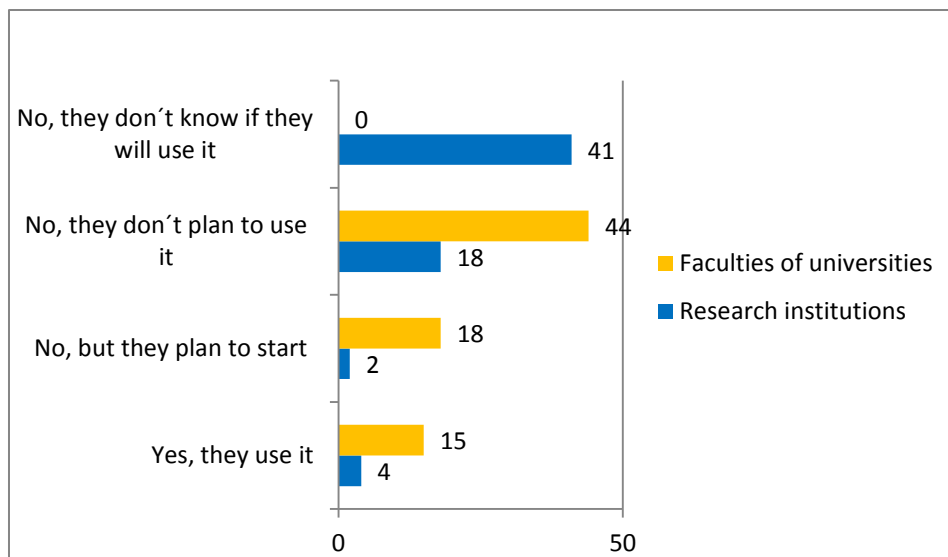
Research institutions also consider research data a trade secret or their know-how; they have concerns about their competitors, etc. Some data are expressly subject to a secrecy provision; research institutions fear that data may be misused. Provision of data is also complicated by patent protection.

*Question no. 18: Are there any plans for providing data stored by your institution for the outside researchers in the near future?*

*(Answered only by those who answered "No" in question no. 15)*

Most respondents that indicated that they did not provide access to data are not planning to provide data outside their institutions due to the above reasons.

*Question no. 19: Do you use public licenses in your institution for providing and marking the research data? (Creative Commons in the Czech Republic)*



### Use of Creative Commons – Research institutions, Universities

Creative Commons licenses comprise a set of public licenses empowering the authors in their decision as to under what conditions a work will be made publicly available. Using them, the author enters into a universally applicable agreement with all potential users of the work, on the basis of which the author provides certain rights to the work and reserves other rights. Creative Commons licenses do not deny the classical understanding of copyright. The popularity of Creative Commons licenses primarily stems from the fact that they are understandable in international context. The terms of license, i.e. the rights and obligations of a user of a work, are expressed graphically using simple pictograms.<sup>2</sup> More information about Creative Commons licenses and a simple guide for selecting the right license is available in Czech at the international website <http://creativecommons.org/licenses/>. Generally speaking, Creative Commons public licenses are currently not very widespread in the Czech Republic. Only 4 respondents from research institutions stated that these licenses were used by their institutions. The situation is better at colleges and universities where fifteen respondents indicated their knowledge and use of these licenses. We may assume that this situation is to a considerable degree due to a lack of familiarity with these public licenses and also due to concerns about misuse of copyright works.

*Question no. 20: Are there any plans for using public licenses in your institution in the near future?*

*(Answered only by those who answered "No" in the previous question no. 19)*

There was also a higher percentage of respondents from colleges and universities that responded that they might introduce the use of and marking with Creative Commons

<sup>2</sup> Creative Commons Czech Republic <http://www.creativecommons.cz/>

licenses. On the other hand, there were only 2 respondents from research institutions that provided this kind of answer.

*Question no. 21: If there would be a project to build a long-term central interinstitutional repository for research data, would your institution be interested in participating in it?*

The state of affairs as regards enhanced publications and, more generally, research data in the Czech Republic might be improved by creating a central repository for data or for both research data and publications. If it is not possible to centralize this effort on a national level, the situation might be improved by a repository that would be shared by several institutions. Therefore, a question seeking to determine whether research institutions would be interested in this solution was added to the questionnaire. The reaction was neither absolutely positive, nor negative. 20 respondents from research institutions stated that they would be interested, 45 respondents were not interested or did not quite know. Respondents from faculties of colleges and universities showed greater interest. 38 respondents said that they were interested, 39 were not.

## **Conclusion**

The survey of the general situation of enhanced publications at Czech research institutions and public and state colleges and universities brought in many respects a validation of the assumptions, but also some surprising findings.

The first four questions of the survey focused on Open Access in scientific publishing. Nearly half of the respondents from faculties of colleges and universities engage in the practice of making research publications available in open repositories. This practice is usually based on internal regulations that were put into force in the past 10 years. However, there are also a few exceptions where this practice is followed without internal regulations. Colleges and universities mostly use institutional repositories or employees' personal websites to publish research publications in Open Access mode. Subject-based<sup>3</sup> repositories are rarely used. This is also due to the fact that there are only a few of them globally.

The advantage of Open Access publishing identified most often by colleges and universities is the opportunity to address a broader audience, achieve higher citation and readership rates and faster exchange of information among scientists. As far as obstacles are concerned, respondents expressed concerns about the low prestige of Open Access journals and repositories, as well as copyright issues and fees connected with Open Access publishing. In the Czech Republic, complications in reporting research activities for the

---

<sup>3</sup> [http://en.wikipedia.org/wiki/Disciplinary\\_repository](http://en.wikipedia.org/wiki/Disciplinary_repository)

Information Register of R&D Results (RIV)<sup>4</sup>, which serves as a basis for the allocation of points to authors affiliated with the relevant institutions, clearly play a major role. Institutions do not make publication decisions on the basis of the openness of the journals, but are mainly interested in whether the journal is on the list of journals with points assigned.

The main part of the survey focused on examining the situation with respect to providing access to enhanced publications and research data in the Czech Republic. The most frequent type of research data were measurement data, followed by research data from experiments, testing, surveys, etc. at research institutions and, on the other hand, post-publication data, such as reviews and evaluations, at colleges and universities. The most common data format was, unsurprisingly, the PDF format. Next were the DOC and spreadsheet formats XML and CSV.

Another interesting finding was that data are in most cases taken care of by researchers themselves. The survey confirmed that there is no centralized solution for data storage at nearly half of the institutions. Most often, data are stored only on workstations of individual researchers. However, this does not secure long-term access to data and data security is seriously jeopardized. Where a centralized solution for data storage exists, it usually means storing the files on shared servers. This method ensures a higher degree of data protection and their centralized management, but does not result in improved work with data, their searchability and accessibility. The ideal solution of data management and storage in a digital repository of an institution was ranked third and it is the assumed solution in cases when changes are planned in the future. It is important to note that the use of digital repositories is increasing mainly at colleges and universities. Repositories are mostly used for archiving and providing access to theses, but providing access to publication activities of the employees and to other research documents is also starting to be considered. The long-term archiving of data might already be considered commonplace in the present era of digital data, yet only half of the addressed research institutions and colleges and universities engage in this practice.

The fact that an absolute majority of research institutions and colleges and universities already link research publications with research data in some way or other was a positive finding. 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 data. An ideal solution is linking a research publication with raw data or other materials from research in a digital repository. This option ranked second, followed by joint storing of publications on a website.

---

<sup>4</sup> The Information Register of R&D Results, part of the Czech Information System of Research, Experimental Development and Innovations. It is a public administration information system securing collection, processing, providing and use of information on publicly funded research, development and innovations (<http://www.vyzkum.cz/FrontClanek.aspx?idsekce=610>).



Researchers mostly reuse data from previous research projects as a basis for further research. 60% of respondents stated that they were willing to provide data also to researchers outside their institution. This would be in the form of a personal visit of the individual interested in provision of data or sending of the data via electronic mail or remote electronic data transfer (FTP). 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.

Where no access to the data is provided, it is important to know the reasons behind this decision. Private research institutions often cannot provide data due to commercial and copyright reasons. As far as tailored research is concerned, data become the property of the client, and it is therefore not possible to disclose them to other parties interested. Copyright issues were mentioned also by other institutions. They treat research data as the property of the authors or the institution. Disclosure is sometimes also banned by agreements with partners and cooperating institutions; data are available only to cooperating partners. Research institutions perceive research data as a trade secret, as their know-how, and they have concerns that their competitors may exploit them, etc. Some data are expressly subject to secrecy provisions in order to enable patent protection. Most respondents that indicated that they did not provide access to data are not planning to change this situation due to the above reasons.

The survey showed that Creative Commons public licenses that help to make use of publications and data easier are still used only 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. We may assume that this situation is caused by a lack of knowledge of these public licenses and by concerns about misuse of copyright works.

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

On the whole, the survey showed that in the Czech context it is advisable to continue with the work on public education and opening of discussion about Open Access and about creating of enhanced publications and providing access to research data. NTK created on its website [www.techlib.cz](http://www.techlib.cz) an Enhanced Publications section, which provides information about enhanced publications, links to materials available globally and monitors the situation in the Czech Republic. In addition, NTK includes the topic of enhanced publication on the agenda of

its educational events and conferences and cooperates with organizations dealing with this issue abroad.

## Poland

### *Open science – the diagnosis of the situation in Poland*

In Poland, there is no strategy for developing ideas of Open Access (open science) at the national level. The Ministry of Science and Higher Education published a declaration in which they stressed the need for free access to scientific publications only in December 2012. However, the document does not set out the principles (rules) for support (funding) this idea. It was not indicated which model for open access (green or gold) is preferred. Equally important, though rather only from a promotional point of view, was the statement of July 2013 published by the Polish Academy of Sciences and the Conference of Rectors of Polish Academic Schools. It was pointed out there that the green way (creating repositories) is the optimal model for Polish science and research institutions. An important step towards a wider availability of scientific works was introduced in July 2014, as an amendment to the Act on Higher Education, which requires that all universities have to upload the master theses of students graduated since January 2009 to the central (national) repository. This obligation also refers to PhD dissertations that are an important source of scientific information.

However, in various research institutions, libraries, or in a group of publishers of scientific journals may be observed a growing awareness of the need to become involved in the Open Access movement in order to increase the accessibility of scientific publications, or to obtain higher rates of citation. Evidenced by the number of digital libraries and repositories emerging from the early twenty-first century, they have been developing in parallel, mainly in a form of digital libraries. Currently, Poland has more than 100 digital libraries archiving almost 2 million publications, including mainly national heritage objects (publications not subject to copyright in the field of property rights).

An extensive list of digital libraries and repositories cooperating within the Federation of Digital Libraries is available at <http://fbc.pionier.net.pl/owoc/list-libs>. In this group there are about 15 repositories that collect current scientific publications (doctoral dissertations, journal articles, textbook and academic monographs). However, there are no research data in these repositories. No sources confirming the existence of open repositories of research data in Poland were found.

Recently, the report *Otwarta nauka w Polsce 2014 Diagnoza*<sup>5</sup> (*Open science in Poland in 2014 Diagnosis*, available at <http://pon.edu.pl/index.php/nasze-publicacje?pubid=13>) was released. According to the report 79% of Polish researchers support the idea of openness in science, but only 12% regularly make their publications available in open access. Surveys on direct accessibility of scientific data and their relationship to scientific publications have not been conducted before in Poland.

---

<sup>5</sup> *Otwarta nauka w Polsce 2014 Diagnoza*, ed. Jakub Szprota, Warszawa 2014, [on-line]. [Access 17.10.2014].  
Web address: <http://pon.edu.pl/index.php/nasze-publicacje?pubid=13>

## Objectives of the survey

The major objective of the research survey that was conducted in the V4 countries was to describe and explore the research data and materials management methods in a scientific environment<sup>6</sup>.

Table 1<sup>7</sup>.

1	Model OA	:	An exploration to determine if researchers' scientific papers are published in Open Access (OA); and a determination of the opinions about the usefulness of this model.
2	Research data management	:	An exploration of the principles of research data management, including production, storage, archiving, usage, and providing raw research data and materials in a scientific environment.
3	Enhanced publications	:	An exploration of the knowledge levels and use of enhanced publications, which are complex digital objects that link research publications to the source data used within them.

## Conceptualization

During the preparation of the research survey, the key concepts and connections between them were determined. The questions that could measure the incidence of the examined phenomenon in reality were specified (Table 2).

Table 2.

---

<sup>6</sup> The scientific environment is formed by the researchers and the institutions where research activities are carried out.

<sup>7</sup> Raw research data are statistical, sociological, geographic data, etc. In this context raw research data includes films, sounds recordings, photos, and both published and unpublished reports.

objectives	research problems	questions
Model OA	publishing	<ul style="list-style-type: none"> <li>➤ Publishing research papers in OA</li> <li>➤ The specific institutional requirements</li> </ul>
	archiving	<ul style="list-style-type: none"> <li>➤ The preferable types of digital archives/repositories</li> </ul>
	usefulness	<ul style="list-style-type: none"> <li>➤ Advantages of publishing in OA</li> <li>➤ Obstacles to publishing in OA</li> </ul>
Research data management	data characteristics	<ul style="list-style-type: none"> <li>➤ Type of obtained data/materials</li> <li>➤ The preferable formats for recording data digitally</li> </ul>
	storage	<ul style="list-style-type: none"> <li>➤ Methods of data storage</li> <li>➤ Data storage location</li> <li>➤ Time of archiving</li> <li>➤ Responsible person for research data storage</li> </ul>
Enhanced publications	use	<ul style="list-style-type: none"> <li>➤ Linking research data with publications</li> <li>➤ Methods of linking research data with publications</li> </ul>
	re-use	<ul style="list-style-type: none"> <li>➤ Data re-use by outside researchers and for other research projects inside the institution</li> </ul>
	providing	<ul style="list-style-type: none"> <li>➤ Data accessibility for outside researchers</li> <li>➤ Methods of providing research data for outside researchers</li> </ul>
	publishing	<ul style="list-style-type: none"> <li>➤ Use of public licenses to provide research data</li> <li>➤ Participation in the project of universal access to research data</li> </ul>

### *Research method*

A survey was carried out by means of a questionnaire that was prepared electronically and distributed by e-mail. The survey was created by the LimeSurvey web application. The questionnaire contained 27 questions that were ordered in four thematic blocks:

1. Open Access: this part gathered information about publishing research papers in OA;
2. Raw research data: this part contained information about raw research data management in institutions;
3. Linking research data with publications: this part is about creating enhanced publications;
4. Identification information: this part gathered personal information including the name of the institution, contact e-mail, type of institution, and specialization.

The questionnaire could be completed anonymously. Also, it was not necessary to put the name of the institution.

### *Sampling method*

Scientific and research-development institutions were the focus of the research. In these institutions individual researchers conduct their research activities. The survey gathered information about the form of scientific activities and principles of the functioning of the institution, but not the opinions of individual researchers.

## Population

According to the Statistical Yearbook of the Central Statistical Office of Poland<sup>8</sup>, 2733 Polish scientific and research-development units were registered in 2013. The research survey did not include commercial economic entities, private higher education institutions, museums, archives, and libraries. Thus, the population did include public institutions that carry out scientific and developmental activities. These types of institutions were classified into 3 groups:

Table 3.

Eligible general population		<i>Population size</i>
1	Scientific and research development units	200
2	Scientific institutes of the Polish Academy of Science (PAS)	82
3	Public higher education institutions	136
Total:		<b>418</b>

## Research sample

The research sample was selected by means of stratified sampling. In each stratum (Table 3) the specified number of elements was randomly drawn. Random selection allowed to obtain a representative sample and to conclude information about the population based on probability. That makes it possible to evaluate the accuracy.

**The sample size of the entire population (with a confidence level of 0.90 and an assumed error level of 0.05) is 268 units.** The calculation of the sample size with such parameters of confidence level and error level gives the confidence (90%) that the obtained results are not different from the results which could be obtained in general population of more than 5%.

Then the sample size of each stratum was calculated. Sample size of stratum is proportional to group's size in the general population.

Table 4.

Research sample		<i>Research sample size</i>
<b>1</b>	Scientific and research-development units	<b>128</b>
<b>2</b>	Scientific institutes of the Polish Academy of Science (PAS)	<b>53</b>
<b>3</b>	Public higher education institutions	<b>87</b>
<b>Total:</b>		<b>268</b>

<sup>8</sup> Rocznik statystyczny Rzeczypospolitej Polskiej 2013 = Statistical Yearbook of The Republic of Poland 2013 / [editor in chief Halina Dmochowska]; Główny Urząd Statystyczny (Central Statistical Office), s. 425.

## **Reserve questionnaires**

All other institutions not included in the sample were treated as a reserve group. In case of only a small number of completed questionnaires, it was planned to send questionnaires to the institutions from the reserve group. There were 150 reserve institutions.

### **Sampling mechanism**

We created a list of all institutions from the three specified groups. The institutions were alphabetized and then a number was assigned. Using a random number generator (<http://www.losowe.pl/liczba>), the appropriate number of institutions from each group was drawn. These institutions constituted the research sample.

### ***Data collection***

The time between the survey distribution and data collection was two months: from 12.05.2014 until 13.07.2014. The requests for completing the questionnaire electronically were sent to the selected institutions. The surveys were distributed via e-mail. We collected e-mails from each participating institution. The collection process revealed that 22 scientific or research institutions were either liquidated or merged into the administrative structures of other institutions. Such information was not discovered during the preparation of the research sample. Therefore, the actual eligible general population is 396 units instead of the previously assumed 418.

At least three e-mails to various offices of the majority of selected institutions were collected. E-mails with the questionnaire were successively sent to different e-mail addresses. In many institutions there was a problem to reach a competent person who could give information about collecting, storing and providing research data and materials. In case of large institutions having developed organizational structure (for example higher education institutions) it occurs that every organizational unit (department) has its own policy of research data and materials management. Accordingly, it was quite difficult to fill out the questionnaire that could describe the situation in entire institution. During the survey we verified the number of received questionnaires. The rate of return was less than 40% of the defined research sample. We then decided to send emails to the reserve institutions. Additionally, e-mails were also sent to institutions that were not initially selected for inclusion in the research sample.

### ***The degree of research sample realization***

A total of 207 questionnaires were gathered, 133 of them were filled out completely (all main questions were responded) and 74 completed partially. This means a 77% participation rate. It is impossible to determine the degree of sample realization for each stratum of the general population. The reason is that providing identification information was not

obligatory. Many of institutions failed to specify not only the institution name, but also the institution type and field of specialization.

## Data analysis<sup>9</sup>

### Participating institutions

Over 38% of the 207 institutions that completed the questionnaire specified the institution name or gave contact information. Most of the respondents (62%) decided to complete the questionnaire anonymously. Contact data was only necessary to calculate the number of filled out questionnaires.

Only 65.7% of respondents specified the type of their institutions<sup>10</sup>. Comparing research sample structure with population structure and taking into account the variable of institution type the degree of sample representativeness was determined.

Table 5.

Institution type:		
N = 136 (65.70%)	NR = 71 (34.30%)	Total = 207 (100%)
Response	Frequency	%
Higher education institution	36	26.47%
Research institute	62	45.59%
Scientific institute of PAS	29	21.32%
Secondary research unit	4	2.94%
Other	5	3.68%

The sample is representative if the percentage participation of each category in research sample is not more than 5% of percentage participation of the same category in population. If one of the sample categories appears rarely, it means that this category is underestimated. If it appears more frequently, it is overestimated. According to the table 6, it can be concluded that it was obtained a representative sample in the survey. The difference between both categories exceeds the given limit. In sample the category of higher education

<sup>9</sup> Table explanation:

**N** – number of respondents that answered the question. The number can be different taking into consideration the filter questions and incomplete questionnaires analyzed during survey research.

**NR** – no responses, the number of respondents that did not answered the question.

**Total** – total number of respondents entitled to answer the question.

**Number of indications** – number of responses to the multiple choice questions. In multiple choice questions the respondents could select more than one answer. Then total number of indications is more than number of respondents that answered the question.

**Frequency** - number of response indications

**%** - percentage participation of the chosen response among all respondents that answered the question.

<sup>10</sup> During the sampling process three groups of institutions were determined: research institutions, scientific institutes of the Polish Academy of Science and higher education institutions. Two response options were added to the list of institution types in the identification information of the questionnaire: *Secondary research unit* and *Other*. Only nine institutions selected these response options. Since the institutions that specified themselves as a "secondary research unit" or "other" institution have not been included in the sample, they are skipped in survey results presentation. The data obtained about these institutions are mentioned in general results of the survey (without division into three groups).



institutions is underestimated, while the number of research institutions is overestimated. Only the category of scientific institutes of Polish Academy of Science has a similar distribution in sample and population. There is no data on the number of secondary research units and units of other type.

*Table 6.*

	Research sample	Population	Difference
Higher education institutions	26.47%	34.34%	-7.87%
Research institutes	52.21%	44.95%	7.26%
Scientific institutes of PAS	21.32%	20.71%	0.62%

**Taking into consideration the 77% of research sample realization and the undertaken analyses related to the sample representativeness in terms of institution type, it was considered that the survey was carried out with the use of a representative sample. Based on the results, we can generalize about the situation regarding the entire population.**

In the identification information there was a question about the scientific disciplines that are represented by the surveyed institutions. Most of indications (almost 38% of respondents) referred to engineering and technology. A large group of units represents social sciences (more than 22%). Classifying disciplines into two general groups – sciences<sup>11</sup> (underlined) and humanities<sup>12</sup> – it can be noticed that the disciplines of the first group have twice as many indications (217, constituting 60% of all indications) than humanities disciplines (111, or 30% of indications).

---

<sup>11</sup>Sciences include Engineering and Technology; Medicine, Pharmacy and Related Sciences; Chemistry; Biological Sciences; Physics and Related Sciences; Earth Sciences; Agriculture and Related Sciences; Mathematics and Computer Science; Architecture and Spatial Planning; Astronomy.

<sup>12</sup> Humanities include Social Sciences; Economy; History and Philosophy; Linguistics and Literary Science; Law; Arts and Design; Psychology.

Table 7.

Scientific discipline (specialization):			
N = 137 (66.18%)	NR = 70 (33.82%)	Total = 207 (100%)	Number of indications= 361
Response	Frequency	%	
<u>Engineering and Technology</u>	52	37.96%	
Social Sciences	31	22.63%	
<u>Medicine, Pharmacy and Related Sciences</u>	29	21.17%	
<u>Chemistry</u>	28	20.44%	
<u>Biological Sciences</u>	26	18.98%	
Economy	19	13.87%	
<u>Physics and Related Sciences</u>	18	13.14%	
History and Philosophy	18	13.14%	
<u>Earth Sciences</u>	18	13.14%	
<u>Agriculture and Related Sciences</u>	18	13.14%	
<u>Mathematics and Computer Science</u>	17	12.41%	
General / Interdisciplinary	16	11.68%	
Linguistics and Literary Science	14	10.22%	
Law	12	8.76%	
Arts and Design	10	7.30%	
<u>Architecture and Spatial Planning</u>	8	5.84%	
<u>Astronomy</u>	3	2.19%	
Psychology	7	5.11%	
Other	17	12.41%	

*Responses of "Other" include<sup>13</sup>: physical culture (3 indications); defence, security, military(3); environmental protection; technical sciences; theology; management; new media, interior design, music education, instrumental studies; history of science, techniques, education; musical art; health sciences; building construction; workplace safety and ergonomics.*

## Open Access Model

### *Open Access publishing*

Almost half of the institutions (47.82%) declared that the research papers of academic staff are deposited in open digital archives and repositories; for over 3% of them it is an institutional obligation. Almost 16% of institutions admitted that it is quite difficult to state whether researchers use the model OA or not. 36.23% of institutions indicated that their employees' research papers are not released in open access format.

Among 8 institutions which have an institutional obligation to publish research results using OA model, two have admitted that it became obligatory quite recently, over the past two years.

<sup>13</sup> "Other (specify)" responses on open and semi-open questions are italicized.

It can be presumed that in most of Polish scientific and research development units the records concerning the open access publishing are not maintained. These data are not collected and included in the statistics. Therefore, it is very difficult to examine the size of the phenomenon. Almost half of institutions indicated that they use open access publishing, slightly more than half of them do not undertake this activity or cannot state this as a fact. It seems that the response on this question was selected intuitively, based on the presumption but not on the owned data.

### *Digital archives and repositories*

The institutions using OA model publish research papers in institutional repositories (almost 65% of respondents) or in outside subject based repositories (almost 50%). A quarter of the institutions use researchers' personal websites. Relatively often the institutions mentioned other places of publication: *institutions' websites (4 indications); open access journals' websites (3); outside digital libraries (for example Pomeranian Digital Library, CYBRA); institutional libraries (2); databases (2) like: e-publications of Polish Science, BazTech, EXPERTUS; specialised systems (2), for example System of Support for Leading Research-Scientific Works, Sharing and Research Works Publication (System Wspomagający Prowadzenie Prac Badawczo-Naukowych oraz Współdzielenie i Publikację Wyników Prac); national and subject based repositories; institutional or personal repositories.*

Analysing the data divided into three groups in accordance with the appointed groups of institutions we can notice some differences in the obtained results (Table 8). Research institutions and scientific institutes of the Polish Academy of Science indicated the use of institutional repositories most frequently. For higher education institutions the preferred places of storing, archiving and publishing research works are personal websites of researchers. In case of higher education institutions the distribution of responses is fairly uniform. Then, we suppose that the researchers of this type of institutions use all mentioned ways of archiving materials in Internet to a similar degree.

Table 8.

What type of archives (repositories) is most frequently used by researchers of your institution?			
N = 93 (93.93%)		NR = 6 (6.06%)	Total = 99 (100%)
		Number of indications = 150	
Response	Summary	Higher education institutions	Research institutions
Personal website of a researcher	<b>25 (26.88%)</b>	13 (36.11%)	4 (12.12%)
Institutional repository	<b>60 (64.52%)</b>	8 (22.22%)	25 (75.76%)
Subject based repository	<b>45 (48.39%)</b>	12 (33.33%)	15 (45.45%)
Other	<b>20 (21.51%)</b>	5 (13.89%)	6 (18.18%)
number of respondents	93	36	18

### *The usefulness of the OA model*

The respondents were asked to point out the advantages of Open Access publishing for their institutions. The responses were sorted in descending order according to the number of indications (Table 9). The most frequently indicated benefits were: an opportunity to reach a wider audience, increasing citations, development of scientific communication and faster information exchange. Responses of "Other" included *dissemination of knowledge about international affairs; promotion of unit/institution*.

Table 9.

What are/could be the main advantages of Open Access publishing for your institution and researchers?		
N = 164 (79.22%)	NR = 43 (20.78%)	Total = 207 (100%)
Number of indications = 948		
Response	Frequency	%
Opportunity to reach a wider audience	144	87.80%
Increasing citations	113	68.90%
Establishing cooperation between researchers	110	67.07%
Development of scientific communication and faster information exchange	97	59.15%
Wider readership of research papers	96	58.54%
Increasing the visibility of publications in reference databases and for web search engines	95	57.93%
Faster publication	75	45.73%
Increasing the prestige of the institution and researchers	55	33.54%
Greater influence on development of a scientific discipline	51	31.10%
Cost savings	40	24.39%
Opportunity to obtain additional funding, grants	39	23.78%
Greater durability of publication	29	17.68%
Other	4	2.44%

Indicating obstacles that discourage the use of the OA model the respondents displayed less solidarity with each other. In case of advantages 87.8% of respondents selected the same advantage. Indicating obstacles 57.49% of respondents had a common response. On average the respondents selected 5.7 advantages and only 2.7 obstacles. The main obstacles are **ambiguities in copyright law, additional costs and work, and the questionable quality of published papers. Other difficulties (responses of "Other") include *no sufficient knowledge about OA (2); necessity of applying for superior unit permission (2); specificity of institution activity; protection of research results; publishing research papers in other resources, for example in Journal Citation Reports that has an Impact Factor.***

Table 10.

What are/could be the main obstacles that discourage your institution and researchers from open access publishing?		
N = 167 (80.68%)	NR = 40 (19.32%)	Total = 207 (100%)
Number of indications = 457		
Response	Frequency	%
Ambiguities in copyright law	96	57.49%
Additional costs (required fee)	67	40.12%
Questionable quality of published papers (not peer-reviewed articles)	61	36.53%
Additional work	56	33.53%
No OA journals / repositories in an appropriate subject field	55	32.93%
Objections of traditional publishers	51	30.54%
Low prestige	35	20.96%
Small influence on development of a scientific discipline	17	10.18%
Narrow readership of research papers	9	5.39%
Other	10	5.99%

Over 80% of all respondents specified benefits and obstacles to Open Access publishing. On average they indicated more advantages than obstacles. It seems then that researchers of Polish scientific and research-development units are convinced that the use of OA model for publishing their research results is advantageous. However the above-mentioned obstacles hinder the open access publishing for over 50% of institutions.

## Raw research data management in the institutions

### *Data characteristics*

Respondents were asked what types of data and research materials are produced and stored in their institutions (Table 11). The most frequently chosen answers were related to research data produced from measuring and experiments (respectively 63% and 60% of institutions). Almost 20% less of them indicated research results obtained from testing and statistical data (third and fourth group of answers). 32% of institutions produce research data from surveys. 10% of respondents mentioned other types of gathered data than those included in the list of answer choices. These were following: *monographs (3), research papers, patents, prototypes, design documentation; technological procedures, maps, interviews, databases, texts, theoretical research results, mathematical modelling of phenomena and the structure of matter, new models of experimental data analysis, projects of research infrastructure, construction of measuring devices in experiments in physics and related sciences, analysis of political science; data from the study of the history of science, technology, education, analytical and historical works from theory of music and practices of musical education, software or visual pieces of art.*

Survey results show that higher education institutions more often produce statistical data, while research institutes – research data from measuring and Polish Academy of Science – experimental data. The distribution of variables in this question is related to the question about the scientific disciplines represented by survey institutions. Institutions usually produce the type of data that is related to the research in a particular field of knowledge.

Table 11.

What types of research data are produced in your institution?					
N = 166 (80.19%)		NR = 41 (19.81%)		Total = 207	Number of indications = 763
Response	Summary	Higher education institutions	Research institutions	Scientific institutes of PAS	
Research data from measuring	<b>105 (63.25%)</b>	17 (47.22%)	48 (77.42%)	19 (65.52%)	
Research data from experiments	<b>101 (60.84%)</b>	18 (50.00%)	47 (75.81%)	21 (72.41%)	
Research data from testing	<b>66 (39.76%)</b>	9 (25.00%)	37 (59.68%)	8 (27.59%)	
Research data from statistical investigation	<b>63 (37.95%)</b>	19 (52.78%)	20 (32.26%)	4 (13.79%)	
Research data from surveys	<b>54 (32.53%)</b>	21 (58.33%)	25 (40.32%)	6 (20.69%)	
Accompanying material - videos	<b>30 (18.07%)</b>	15 (41.67%)	8 (12.90%)	3 (10.34%)	
Accompanying material - audio records	<b>15 (9.04%)</b>	9 (25.00%)	1 (1.61%)	2 (6.90%)	
Accompanying material - plans	<b>21 (12.65%)</b>	5 (13.89%)	9 (14.52%)	4 (13.79%)	
Accompanying material - models	<b>44 (26.51%)</b>	10 (27.78%)	20 (32.26%)	9 (31.03%)	
Accompanying material - schemes	<b>37 (22.29%)</b>	10 (27.78%)	16 (25.81%)	6 (20.69%)	
Other accompanying material	<b>30 (18.07%)</b>	7 (19.44%)	9 (14.52%)	5 (17.24%)	
Accompanying material - algorithms	<b>32 (19.28%)</b>	6 (16.67%)	10 (16.13%)	9 (31.03%)	
Post-published data reviews	<b>67 (40.36%)</b>	16 (44.44%)	22 (35.48%)	17 (58.62%)	
Post-published data evaluation	<b>38 (22.89%)</b>	10 (27.78%)	16 (25.81%)	7 (24.14%)	
Post-published data comments	<b>43 (25.90%)</b>	12 (33.33%)	16 (25.81%)	7 (24.14%)	
Other post-published data	<b>17 (10.24%)</b>	4 (11.11%)	4 (6.45%)	5 (17.24%)	
Number of respondents	166	36	62	29	

The most popular and most frequently used data storage format is PDF, used by more than 86% of institutions (Table 12). The second format concerning to the number of indications is DOC. Every tenth of respondents identified also other, more specific formats such as jpg (4), tiff (3), tex (3) and others. The distribution of answers concerning data format used by various types of institutions is related to general results. Irrespective of institution type two most frequently used formats are the same.

Table 12.

Indicate the most common formats for data storage in your institution					
N = 166 (80.19%)		NR = 41 (19.81%)		Total = 207 (100%)	Number of indications = 383
Response	Frequency	Higher education institutions	Research institutions	Scientific institutes of PAS	
Pdf	<b>144 (86.75%)</b>	23 (63.89%)	51 (82.26%)	21 (72.41%)	
Doc	<b>114 (68.67%)</b>	33 (91.67%)	55 (88.71%)	26 (89.66%)	
Xls	<b>58 (34.94%)</b>	5 (13.89%)	9 (14.52%)	5 (17.24%)	
Csv	<b>22 (13.25%)</b>	9 (25.00%)	29 (46.77%)	13 (44.83%)	
Xml	<b>21 (12.65%)</b>	8(22.22%)	8 (12.90%)	2 (6.90%)	
JSON (JavaScript Object Notation)	<b>5 (3.01%)</b>	2 (5.56%)	3 (4.84%)	0.00%	
Other	<b>19 (11.45%)</b>	4 (11.11%)	7 (11.29%)	4 (13.79%)	
Number of respondents	166	36	62	29	

### Data preservation and archiving

The vast majority of the institutions' data is stored on the workstations of individual employees. Fewer than half of institutions store research data on central servers. Combining these two answers it may be said that the majority of raw research data is stored as internal data, not accessible to other researchers (76% of all indications). Open repositories were mentioned only by 13% of respondents<sup>14</sup>. That clearly shows the main tendency and preference of Polish research institutions for data storage methods — **data are mainly not available for external users, sometimes for other institutional researchers either, and they are stored on the researchers' own PCs**. Over 17% of institutions have also indicated other data storage locations, including the following: *libraries (7), archives (7), offices (3), media (4), institutional databases (2), repository of research team, closed external repository, backups, local servers, project manager PC or lab PC, servers of large international project teams (e.g. CERN), publication at homepage, measuring position*.

The Distribution of variables for each group of institutions is similar to general results of the survey.

Table 13.

Where are the research data of your institution stored?				
N = 164 (79.23%)		NR = 43 (20.77%)		Total = 207 (100%)
Number of indications = 262				
Response	Summary	Higher education institutions	Research institutions	Scientific institutes of PAS
On workstations of individual researchers	<b>129 (78.66%)</b>	28 (77.78%)	52 (83.87%)	26 (89.66%)
In a directory on (common) server	<b>71 (43.29%)</b>	9 (25.00%)	34 (54.84%)	12 (41.38%)
In a central digital repository of the institution	<b>19 (11.59%)</b>	2 (5.56%)	7 (11.29%)	4 (13.79%)
In a subject based repository	<b>15 (9.15%)</b>	5 (13.89%)	5 (8.06%)	2 (6.90%)
Other	<b>28 (17.07%)</b>	7 (19.44%)	12 (19.35%)	5 (17.24%)
Number of respondents	<b>164</b>	36	62	29

The majority of institutions (81.01%) did not plan to change the methods of raw data storage (archiving), while only 18.99% have such plans. Modifications planned include *creation of local open or closed repository (17), digital library implementation, creation of publication database (2), introduction of internal system for knowledge or data management (3), work on more precise rules, make research data open – if possible, improvement and modernization of journals' homepage, creation of central database for research data, purchase of streamer, usage of POL-on (Polish Central System of Researches), placement of data at CEON( Polish Centre for Open Science)*.

<sup>14</sup> Combining two questions we present data according to the number of indications, but not the number of responses. Each respondent could choose several responses to one question.



It seems that among all types of institutions, higher education institutions were those which mainly declared changes. Although still 59.38% of units of this type of institutions do not plan any changes, but 40.63% indicated a need for change. The difference between number of respondents which are in favour of changes and those which are against them is 20%. In other types of institutions these differences are much bigger: for research institutes it is 16.67% for changes and 83.33% against, for PAS units — 10.34% for changes and 89.66% against.

80.92% institution declare archiving of research data for over 10 years. The results of survey are similar in each type of units: higher education institutions (77.78%), research institutes (86.44%), as well as for PAS units (82.76%).

The institutions which do not store data for 10 years indicated other terms: *5 years period (7), 10 years (4); 5-10 years; 6 years, 2 years; no specific deadline (2), according to existing regulations, depends on: materials, researchers, rank of the data).*

In most of the surveyed institutions (71.34%) each researcher is responsible for his own raw data. This answer was the most frequently selected irrespective of the type of institution. For higher education institutions and PAS units, the superiority of this option was obvious. As the second possibility, respondents indicated that there are appointed persons in the libraries who take care on research data. However this option was on average only half as many times selected than the most frequently chosen answer. Among other persons or units responsible for raw data storage 17.68% of respondents mentioned the following: *head of / worker of archive (7), project manager (research manager) (4); head of unit: department / laboratory (3); Department of Research Planning and Programming (3); head of department / office employee (2); researchers are responsible for data storage (2); vice director for research (2); head of Department of Research and Publishing; the server administrator in international project; officer of documentation; staff of Department for Research Management, Department of Public Information; appointed employee; appointed employee of secretary designated for research; team for coordination of research work; or distributed management.*

Table 14.

Who is responsible for storing and archiving research data in your institution?				
N = 164 (79.23%)		NR = 43 (20.77%)	Total = 207 (100%)	Number of indications = 257
Response	Summary	Higher education institutions	Research institutions	Scientific institutes of PAS
Appointed person(s) in the library	<b>64 (39.02%)</b>	14 (38.89%)	25 (40.32%)	9 (31.03%)
Appointed person (s) in the IT department	<b>37 (22.56%)</b>	4 (11.11%)	18 (29.03%)	6 (20.69%)
Appointed person - scientific secretary	<b>10 (6.10%)</b>	2 (5.56%)	4 (6.45%)	0.00%
Each researcher himself	<b>117 (71.34%)</b>	30 (83.33%)	43 (69.35%)	26 (89.66%)
Other	<b>29 (17.68%)</b>	5 (13.89%)	14 (22.58%)	4 (13.79%)
Number of respondents	<b>164</b>	36	62	29

To sum up rules of research data management in Polish scientific institutions it has to be said that preferred actions are those directed inside of the institution, and mostly they are not centralized but distributed to the staff. It seems to be an informal rule that researchers are responsible for research data stored on their own PCs. This tendency was observed at most institutions, no matter what their type was. Some 40% respondents declared more centralized management (data are on a common server, monitored by specified person). Rarely institutions store research data in open repositories of their own or external institutions (only 13% of all indications).

#### Enhanced publications

##### *Linking data with publications*

About 77% of Polish institutions link research data with publications. Over 23.5% of them do not follow this practice (Table 15). Taking into consideration the institution type it can be observed that most of the institutions that link data with research papers are included into the group of scientific institutes of PAS (proportion 5:1), only few of them represent the group of higher education institutions.

Table 15.

Are the research data linked with research papers?				
N = 149 (71.98%)		NR = 58 (28.02%)	Total = 207 (100%)	
Response	Summary	Higher education institutions	Research institutions	Scientific institutes of PAS
Yes	<b>114 (76.51%)</b>	24 (77.42%)	48 (78.69%)	21(84.00%)
No	<b>35 (23.49%)</b>	7 (22.58%)	13 (21.31%)	4 (16.00%)
Number of respondents	<b>149</b>	31	61	25

It was examined if there is a relationship between open access publishing and the practice of linking data with research papers. Table 16 shows whether open access publishing influences the practice of linking data with publications. An inverse relationship is shown in Table 17: linking data with research papers promotes open access publishing. Regardless of whether institutions use the OA model for publishing their research papers or not, most of them indicated that they link research data with publications. Nevertheless much more institutions that make their works freely available on the Internet practise linking data with research papers, compared to the institutions that do not use the OA model. It can be concluded that there is no clear correlation, though open access publishing promotes linking data with publications. The correlation occurs in other described situation. Institutions that practise linking data with research papers in most cases publish their research works using the OA model, while the institutions that do not practise that do not publish in open access format. Thus, survey evidence indicates that linking data with publications is closely related to using an open access model.

Table 16.

Linking data with publications	Open access publishing <sup>1</sup>	
	Yes	No / Hard to say
Yes	83.75%	68.12%
No	16.25%	31.88%
Number of respondents	80 (100%)	69 (100%)

Table 17.

Linking data with publications	Open access publishing		
	Yes	No / Hard to say	Number of respondents
Yes	58.77%	41.23%	114 (100%)
No	37.14%	62.86%	35 (100%)

Answers on the question about research data storage location reflect the data management strategy. Responses were merged into two groups: *storing data in open institutional archives* and *storing data in closed archives*<sup>15</sup>. The relationship between methods of data storage and the practice of linking data with research papers was analysed: Table 18 shows the ways of data storage having an influence on linking data with publications; while Table 19 indicates that the practice of linking data with research papers has an influence on the ways of data storage. Irrespective of the specified method of data storage, most of the respondents stated that in their institutions data are linked with publications. However the institutions which store data in closed institutional archives indicated more often that they do not

<sup>15</sup> Closed archives include workstations of individual researchers and institutional central server. Open archives include open institutional repositories and open subject-based repositories.

practise linking data with research papers. Regardless of the assertion about linking data with publications, institutions mostly indicated that they store data in closed institutional archives.

Table 18.

Linking data with publications	Research data storage location	
	<i>Closed institutional archives</i>	<i>Open archives</i>
<i>Yes</i>	79.35%	87.10%
<i>No</i>	20.65%	12.90%
<b>number of indications</b>	184 (100%)	31 (100%)

Table 19.

Linking data with publications	Research data storage location		
	<i>Closed institutional archives</i>	<i>Open archives</i>	<b>Number of indications</b>
<i>Yes</i>	84.39%	15.61%	173 (100%)
<i>No</i>	90.48%	9.52%	42 (100%)

Respondents that declared linking data with research papers were asked to specify the linking methods. The most prevalent response selected by all three respondent groups referred to giving information about data access in publication records (metadata). Compared with other response options, this option seems to be quite distant from the model of enhanced publications (EP), which is based on publishing digital data and research papers in different places: for example on websites and repositories. In Polish institutes the depositing digital research data occurs much less frequently (about 40 percentage points less) than posting information about data access. Despite the assertion of the majority of Polish institutions (75%) that they link raw data with publications based on them, the preferred way to enable access to the research data is posting information about data storage locations. A minority of respondents (over 28%) link digital raw data with research papers.

Higher education institutions compared to other institutions indicated different ways of linking data with publications. The majority of scientific institutes of PAS (80% of them) use one prevalent method and less frequently other methods.

Table 20.

How are publications linked with research data used within them?					
N = 112 (98.24%)		NR = 2 (1.76%)		Total= 114 (100%)	
				Number of indications = 154	
Response	Summary	Higher education institutions	Research institutions	Scientific institutes of PAS	
By joint storing of data with publication in digital form on personal website of research	<b>20 (17.86%)</b>	6 (26.09%)	5 (10.42%)	3 (14.29%)	
By joint storing of data with publication in digital form in repository	<b>32 (28.57%)</b>	6 (26.09%)	15 (31.25%)	4 (19.05%)	
By giving information on data availability in publication records (metadata)	<b>76 (67.86%)</b>	16 (69.57%)	36 (75.00%)	17 (80.95%)	
By linking research papers with data storage location in digital form	<b>18 (16.07%)</b>	8 (34.78%)	7 (14.58%)	1 (4.76%)	
Other	<b>8 (7.14%)</b>	1 (4.35%)	4 (8.33%)	2 (9.52%)	
Number of respondents	112	23	48	21	
Average number of indications	1.4	1.6	1.3	1.2	

Responses of "Other": printed publications with data or information about data (2); institute's website; repository with limited access; traditional method, analogue; works accessed on institute's server; work reports available at the library; by citation.

Only 20.59% of institutions that do not link data with research papers expressed a willingness to change the current situation. The majority of these institutions (79.41%), irrespective of institution type, do not plan to undertake activities aimed at linking raw data with publications.

#### Using and providing research data

In over 80% of institutions, research data are re-used in other research projects and by other researchers of the institution. This was acknowledged by 90% of research institutes' employees, 78% of higher education institutions, 74% of scientific institutes of the Polish Academy of Science.

Table 21.

Are the data produced in your institution re-used in other research projects and by other researchers?					
N = 144 (69.56%)		NR = 63 (30.44%)		Total = 207 (100%)	
Response	Summary	Higher education institutions	Research institutions	Scientific institutes of PAS	
Yes	<b>116 (80.56%)</b>	22 (78.57%)	54 (90.00%)	20 (74.07%)	
No	<b>28 (19.44%)</b>	6 (21.43%)	6 (10.00%)	7 (25.93%)	
Number of respondents	144	28	60	27	

Almost 70% of Polish institutions provide data to external researchers. Most of the affirmative answers were selected by Institutes of PAS.

Table 22.

Are the data produced in your institution available for researchers from external institutions?				
N = 141 (68.12%)		NR = 66 (31.88%)		Total = 207 (100%)
Response	Summary	Higher education institutions	Research institutions	Scientific institutes of PAS
Yes	98 (69.50%)	17 (60.71%)	38 (66.67%)	22 (81.48%)
No	43 (30.50%)	11 (39.29%)	19 (33.33%)	5 (18.52%)
Number of respondents	141	28	57	27

Institutions that provide data to researchers from external institutions mainly indicated that the preferable way of sharing data is distribution via e-mail. More than 50% of institutions prefer to receive data personally, while every fifth respondent downloads data from websites. Over 16% of respondents pointed out other data providing methods, which include *personal contact (5); printed publications (3) or electronic publications (1) which contain data; posting data in subject based repositories and international registers (2); other methods according to the type of data (2); common data access based on the cooperation agreement; depositing data to the common databases both on the personal servers and outside servers with remote access conditioned by passwords; publishing in journals; permission of department head.*

Table 23.

How does your institution provide research data for researchers from external institutions?			
N = 97 (98.98%)		NR = 1 (1.02%)	Total = 98 (100%)
		Number of indications = 179	
Response	Frequency	%	
Personal receipt in the institution	49	50.52%	
Sending data by e-mail	63	64.95%	
Downloading data from websites	33	34.02%	
Downloading data from repositories	18	18.56%	
Other	16	16.49%	

Mentioned among the reasons of not providing data for external researchers were the confidentiality of collected data and office secrecy (6 responses). In some cases data is provided only with the permission of a competent person, such as the head of department / principal / author, or following signed agreements (4). It was also stated that data is the property of a researcher / principal / institution. Sometimes the data specificity makes it difficult to disseminate them (3). Part of institutions indicated protection of intellectual property and ambiguities in copyright law (3). Other reasons include (responses of "Other"): *no appropriate access security software, providing data only in the institution library*

according to the availability clause, providing data only after project realization or data that are processed in publications, proceedings, monographs, and no declared needs.

The vast majority of institutions (87.5%) which do not provide research data for external institutions are not planning to change this situation. Only 12.5% of them declared willingness to disseminate data to external institutions.

Over 82% of scientific institutions do not use public licenses for providing data. This kind of response was dominant in each group of units.

Table 24.

Does your institution use public licenses (for example <i>Creative Commons</i> ) for depositing data?				
N = 141 (68.12%)		NR = 66 (31.88%)		Total = 207 (100%)
Response	Summary	Higher education institutions	Research institutions	Institutes of PAS
Yes	25 (17.73%)	5 (16.67%)	8 (14.04%)	3 (11.54%)
No	116 (82.27%)	25 (83.33%)	49 (85.96%)	23 (88.46%)
Number of respondents	141	30	57	26

Two key questions that can help to identify whether open access publishing is popular among Polish researchers or not are those about open access publishing and depositing research data under public licenses. The results have shown that 52% of institutions do not publish in OA format and 82% of them do not provide research data. It was checked if there is a dependency between these variables. Both the institutions that publish and do not publish research papers in OA format in most cases do not use *Creative Commons* licenses (Table 25). However there are more institutions that publish open access and use public licenses for depositing raw research data than the institutions that do not publish open access and use public licenses. Thus, open access publishing encourages the use of public licenses for posting data. There is a reverse dependency between these variables (Table 26). Most of the institutions that deposit data under public licenses publish in OA format, while the institutions that do not use public licenses publish or do not publish in OA format in equal measure. Therefore, it can be suggested that the use of public licenses for depositing data promotes open access publishing.

Table 25.

Open access publishing		
Use of public licenses	Yes	No / Hard to say
Yes	18.95%	11.11%
No	63.16%	88.89%
Number of respondents	95	63

Table 26.

Open access publishing			
Use of public licenses	Yes	No / Hard to say	Number of respondents
Yes	72.00%	28.00%	25
No	51.72%	48.28%	116

Institutions that do not use public licenses for data providing were asked if they have any plans to change the current situation. 70.91% of institutions replied that they do not; 29.09% of them *answered in the affirmative*. Readiness to change was declared by 40.91% of higher education institutions, 31.82% of scientific institutes of PAS and 21.28% of research institutions.

The last question in the questionnaire was intended to check opinions and readiness of Polish scientific institutions to participate in the project of public access to research data. Over 71% of units expressed their interest in creating a central inter-institutional repository for research data. However, a considerable group of respondents (28%) does not want to take part in the project. The biggest disinterest in the project was shown by institutes of PAS (40%). One quarter of higher education institutions and research institutions do not wish to participate in creating an inter-institutional repository either.

Table 27.

If there was a project to build a long-term central inter-institutional repository for research data, would your institution be interested in participation in it?				
N = 138 (66.67%)		NR = 69 (33.33%)		Total = 207 (100%)
Response	Summary	Higher education institutions	Research institutions	Scientific institutes of PAS
Yes	99 (71.74%)	24 (75.00%)	41 (74.55%)	14 (60.87%)
No	39 (28.26%)	8 (25.00%)	14 (25.45%)	9 (39.13%)
Number of respondents	138	32	55	23

It was checked which factors could have an influence on the intention to participate in the project of public access to research data. Does the fact that the institutions publish research works using OA model reflect opinions about the project? Most of institutions that make their researchers' works freely available to all researchers wish to become involved in the project, while the majority of institutions that do not publish in OA format have no wish to take part in the project (Table 28). The difference in the responses is not big indeed and we cannot declare this to be an apparent relationship, but the fact that open access publishing promotes idea of public access to research data is not an overinterpretation.



Table 28.

Open Access publishing		
Participation in the project of universal access to research data	Yes	No / Hard to say
Yes	78.21%	63.33%
No	21.79%	36.67%
Number of respondents	78	60

Most of the institutions that deposit research data under public licenses expressed willingness to be engaged on the project as well. Some of respondents that do not use public licenses also wish to get involved in the process of creating inter-institutional research data repository, but the number of these institutions is much smaller. (Table 29). It can be concluded that the use of public licenses influences involvement in the project.

Table 29.

Use of public licenses		
Participation in the project of universal access to research data	Yes	No
Yes	88.00%	68.52%
No	12.00%	31.48%
Number of respondents	25	108

The same can be said about the results related to linking research data with publications. Most of the respondents who follow this practice want to participate in the project, while only few institutions which do not link data with publications give importance to building research data repository (Table 30).

Table 30.

Linking data with publications		
Participation in the project of universal access to research data	Yes	No
Yes	76.92%	54.84%
No	23.08%	45.16%
Number of respondents	104	31

## Conclusions

### Research data management

1. Produced within the institutions, research data 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. 10% of units referred to other types of produced data which were not specified as response options. In all types of

institutions the most common format for data storage is “.PDF” (87%), followed by “.doc” (67%). 11% of units use other, untypical formats for data storage.

2. In most institutions research data is stored in places available only for the institute's staff, for example, on individual workstations or a central server, which were two of the most frequently cited responses. Only 12% of units mentioned open resources, either their own or external. In addition, mainly the staff of institutions (71% of institutions) are responsible for research data storage and archiving. A minority of institutions (46%) use central data management administered by appointed 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 occurs in an unsystematic and decentralized way. The preferred method is for an individual employee to be responsible for the data storage in the parent institution. Moreover, the institutions do not wish to change the situation: 80% of units do not plan to make changes concerning the method of research data storage.

#### **Linking research data with publications**

3. About 77% of Polish institutions have declared the use of linking 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 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 distant from the model of enhanced publications, which is based on posting data and publishing research papers in digital form. Less than 30% of institutions indicated that they use this way of linking data. More than 76% of institutions admit to linking data with research papers through publishing information about access to data, but not digital objects in open archives/repositories. 80% of institutions that do not currently link data with publications do not plan to make any such changes in the near future.
4. The survey found out that linking research data with publications promotes that researchers publish their works in OA. On the other hand the institutions that do not link data with research papers also do not make their works freely available on the Internet. The observed relationship is reversed: the fact that the researchers publish in open access promotes the activity of linking publications with data.
5. It was checked if the institutional data storage method has an effect on the practice of linking data with publications. Storing data in open archives/repositories encourages the practice of data linking. A slightly smaller number of institutions that store and archive data in closed archives links data with publications. In turn, linking data with research papers promotes the use of open access archives.

6. 80% of institutions have declared data re-use in other research projects and by other researchers. 70% of institutions provide data for external research institutions. The preferable way of providing data is distribution via e-mail. More than 50% of institutions share data via individual contacts. The main reasons for not providing data for external researchers are confidentiality of collected data, data/office secrecy, and intellectual property protection. 88% of institutions that do not provide data for other researchers do not plan to change the current arrangements. Provided research data is repeatedly used by researchers from the parent institution as well as researchers from external institutions.

### **Open access publishing**

7. Over 48% of respondents declared that researchers publish their papers using an open access model, while the other 52% claimed that the researchers either do not make their works open access or it is impossible to verify. Open access publishing is popular in higher education institutions and scientific institutes of the Polish Academy of Science (more than 50% of units). The other types of institutions (77% of them) declared that they do not use the open access publication model. Thus, the situation is quite difficult to define. 16% of institutions could not explain precisely whether the researchers make their research works freely available on the Internet or not.
8. The institutions whose researchers publish their research results open access predominantly use institutional repository (65% of units) or external subject based repositories (48%). 22% of institutions find other more specific and unusual solutions.
9. The majority of institutions (82%) do not use public licenses for research data deposition. 71% of units 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.
10. Regarding Polish research institutions, it is remarkable that making research papers and data freely available on the Internet is not widespread. Instead, researchers rather prefer to publish their research results as peer-reviewed publications instead of providing research data for a wider audience.

### **Opinions about open access publishing**

11. 80% of respondents specified some of the advantages and obstacles to open access publishing. It seems that awareness of the advantages of open access publishing is quite high. At least six advantages to an institution or researchers were indicated. The most often indicated advantages include broadening the readership, increasing citation frequency, developing scientific communication, and faster information exchange. The

main obstacles include ambiguities in copyright law, additional costs, and the questionable quality of some published papers.

12. Though awareness of the advantages of open access publishing is high, the majority of researchers do not deposit research papers in open access archives and repositories.

13. 71% of institutions wish to participate in creating a central inter-institutional repository. The experience of open access publishing, use of public licenses for research data dissemination, and providing and linking data with publications all encourage taking part in the project of public access to research data.

## **Slovakia**

### *The subject of the project and “research data” as the key phrase*

Research data (or “vedecké dáta” in Slovak) represent the valuable background for researchers, teachers, scholars, scientists, students and other groups forming the intellectual expert community. Such data include schemes, models, plans, surveys, documentations, etc. In Slovakia, there is no complete working system built on such data which would interconnect institutions and organizations dedicated to education, linking their data, complementing each other and integrating the data into a compact tool. It is supposed that these data exist in their institutions of origin, accessible in some form to the visitors of the individual institution or their web site. This document should provide a description of how exactly these are formed, stored, and made accessible.

### *The purpose of this document within the project*

This document, based on the project titled “Enhancing scholarly communication: National initiatives to manage research data in the V4 countries” and led by the University of Debrecen, aims to map the situation of research data in Slovakia in its scholarly environment and follows their formation, accessibility, and storage. On the basis of the survey and questions forwarded to leading Slovak universities and scholarly institutions, it shows their current position. The project itself will compare the surveys and thus the development/current situation of the subject in Slovakia and other V4 countries. Besides this, the document shows the interest of the relevant parties in the creation of the format that would help make the research data more accessible. Mapping the bases and the work system within the research and education institutions, it shows which road needs to or should be taken and where to start.

### *Formation of the document*

The document is based on the survey performed simultaneously in Hungary, the Czech Republic, Poland and Slovakia including carefully formulated and chosen questions agreed upon by the project partners and targeted at the relevant institutions.

In 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 cases where no response was received, the survey collectors tried to reach the contact persons by telephone.

In Slovakia, the survey itself was performed between 20<sup>th</sup> April and 30<sup>th</sup> June, reaching the peak of collected answers in 20 days after the beginning of the survey and a second, smaller peak 40 days after the beginning of the survey, after contacting the institutions by the phone.

Unfortunately, only 44.4% of contacted institutions responded. Other institutions refused or were unable to answer. Their reasons included the confidentiality of the information requested, the lack of research data or undefined reasons.

### *The survey and the questions*

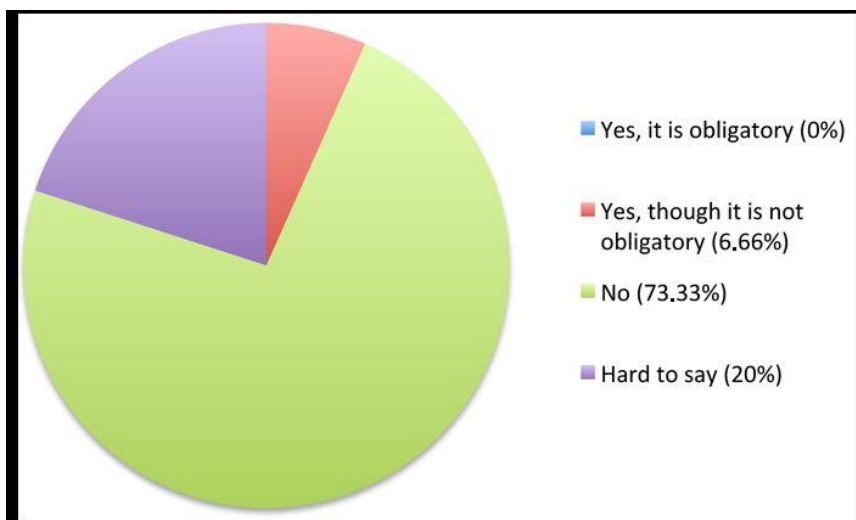
The survey was divided into 4 pages:

- page 1: Open Access
- page 2: Raw Research Data
- page 3: Dissemination and Reuse of the Research Data
- page 4: Identification data

#### **Page 1: 5 questions on publishing and archiving research papers in Open Access**

*Question no. 1: Are the research papers of your institution stored and accessible in open archives and digital repositories?*

The pie chart based on the collected responses:

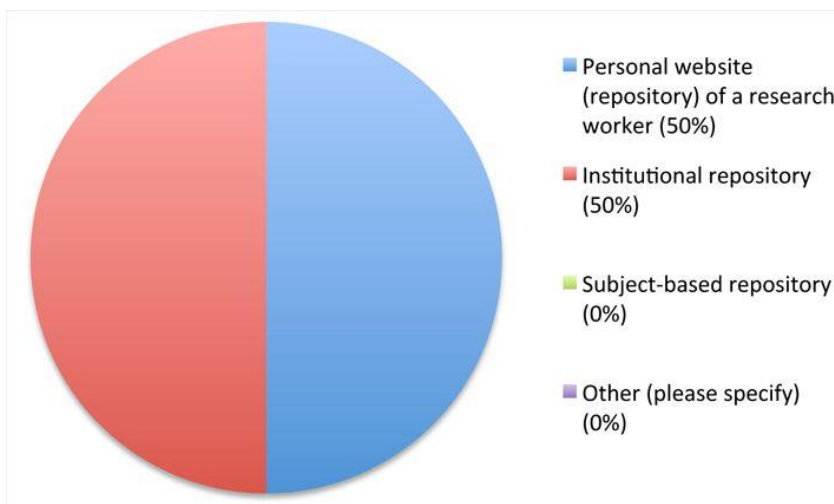


It is obvious that the institutions are not obliged to make research data accessible as there was no “Yes” answer. The institutions mostly do not make the data openly accessible.

*Question 1a: The question “If it is obligatory, since when?” was irrelevant as there were no institutions, which were/are obliged to create open archives and repositories.*

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

The question follows the “Yes” answer to “Question 1”. The already small number of positive answers were proportionally divided into the categories of personal website and institutional repository:



*Question 2: What are / could be the main advantages of publishing or storing in Open Access for your institution and researchers?*

The most frequently given answers to this question were:

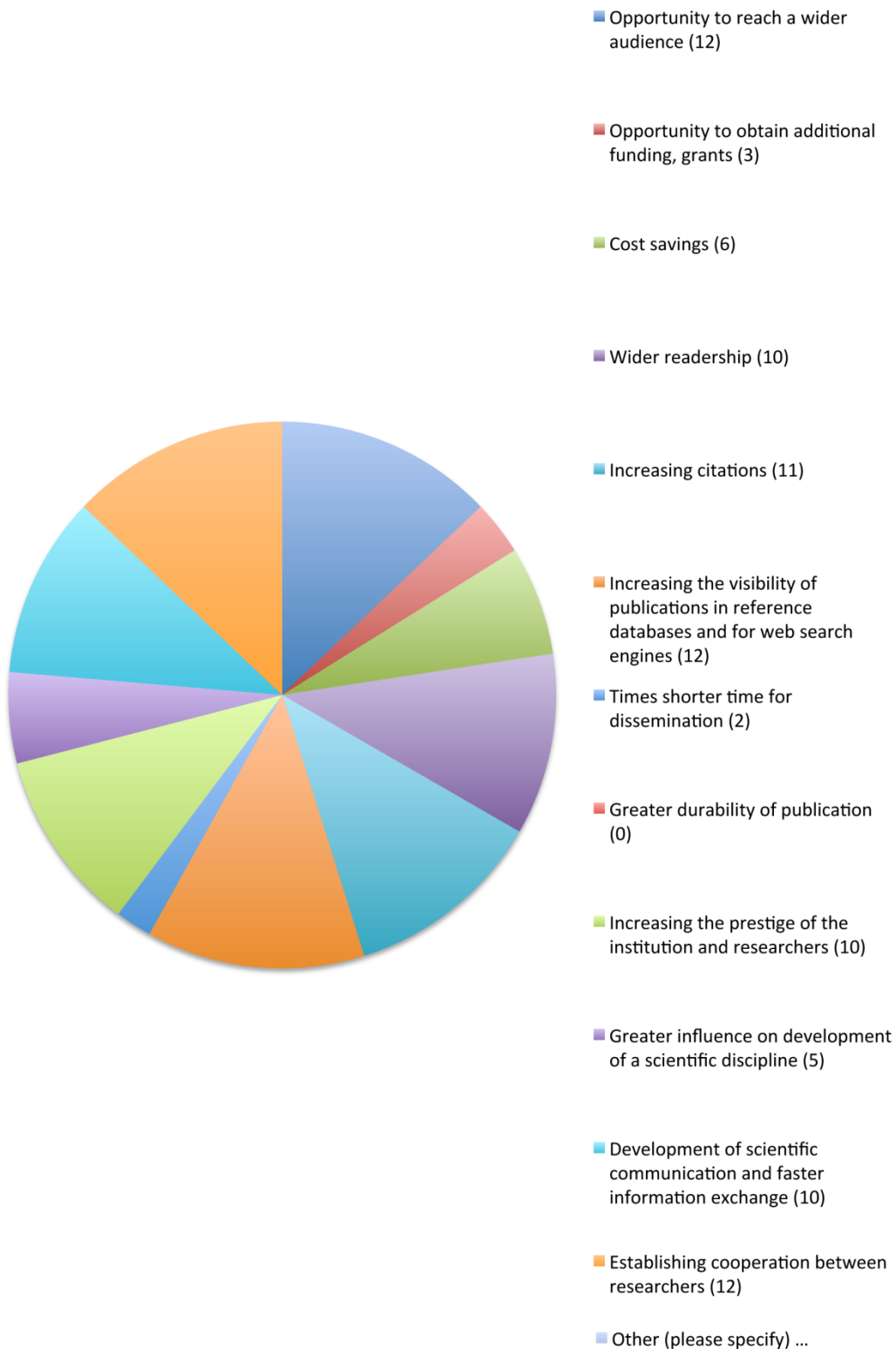
- opportunity to reach a wider audience,
- increasing the visibility of publications in reference databases and for web search engines,
- establishing cooperation between researchers.

Other very frequent answers were:

- wider readership,
- increasing citations,
- increasing the prestige of the institution and researchers,
- development of scientific communication and faster information exchange.

All responses and their frequency can be seen in the pie chart below, which shows the actual number of chosen answers:



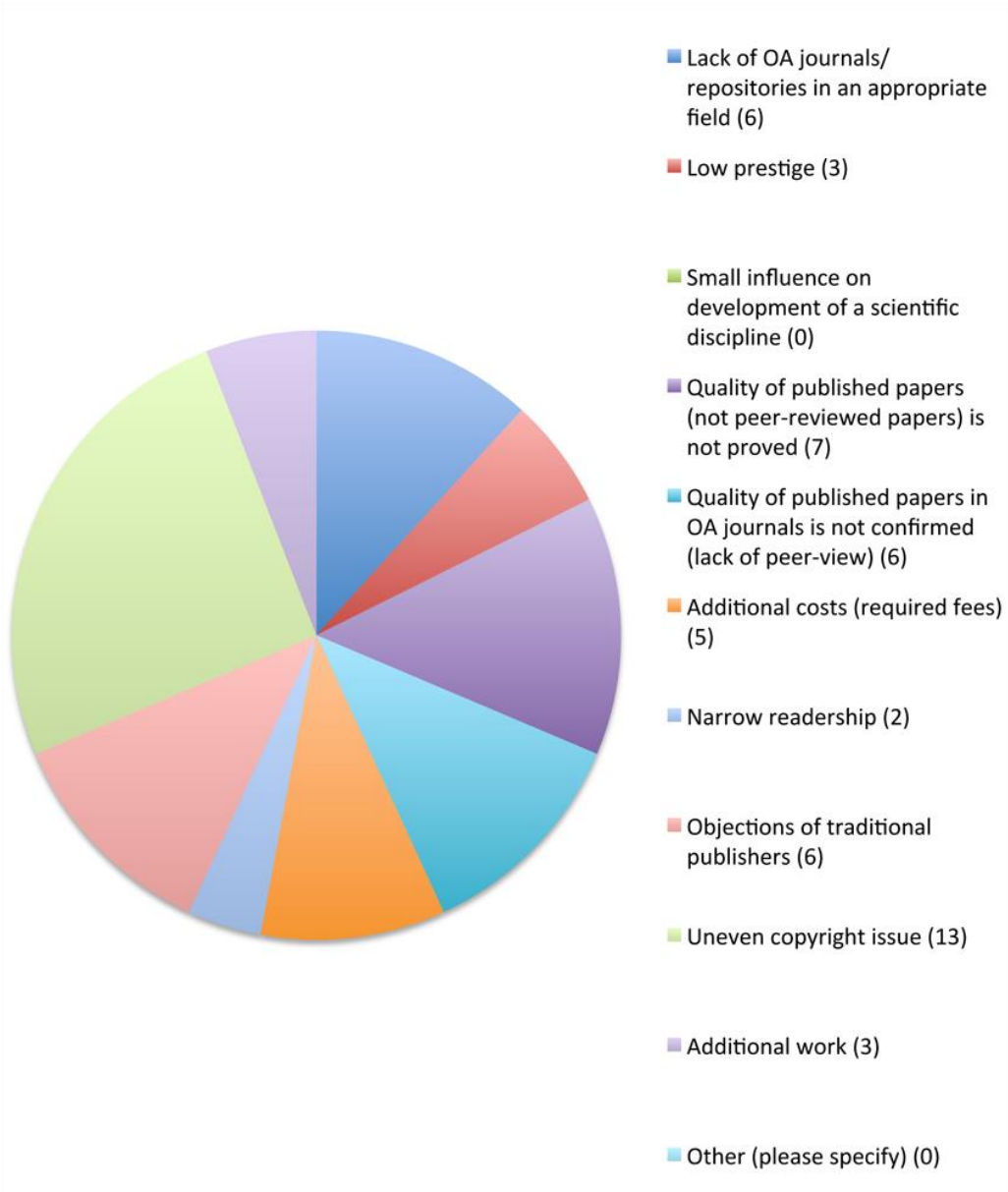


*Question 3: What are / could be the main obstacles for your institution and researchers that discourage them from publishing or storing research papers in OA model?*

All respondents agreed on the single issue, which represents the biggest obstacle in OA model: “uneven copyright issue”.

The institutions thus see more advantages of the OA model than disadvantages or obstacles (where there were more advantages chosen in the previous question, and so in total it received more hits). However, the copyright issues were selected the most frequently, including in response to the previous question, and so that may be seen as the major topic to discuss.

Other relevant obstacles can be seen in the pie chart below, which represents the number of checked answers:



## Page 2: 8 questions on Raw Research Data Management in an Institution

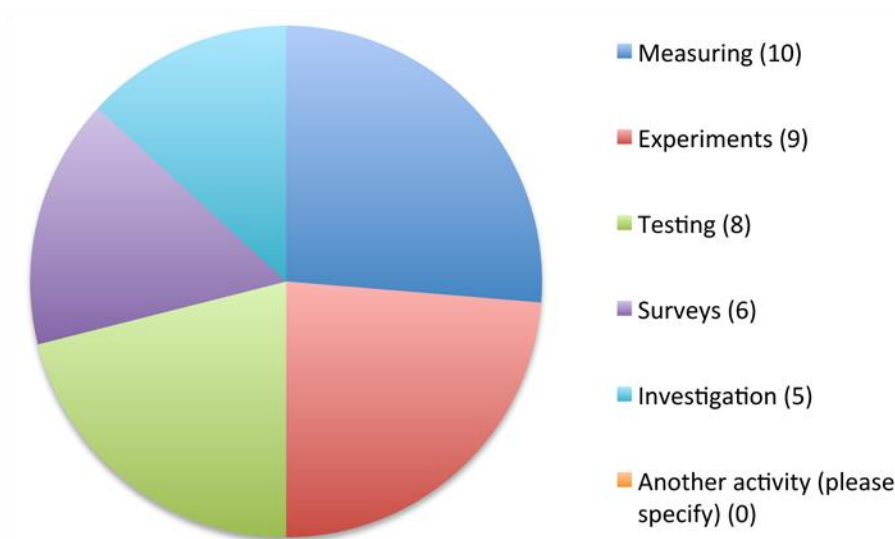
### *Question 4: What types of research data are produced in your institution?*

The question was subdivided into three questions on the basis of:

- a) Manner of the Research Data creation
- b) Type of accompanying material
- c) Types of post-published data materials

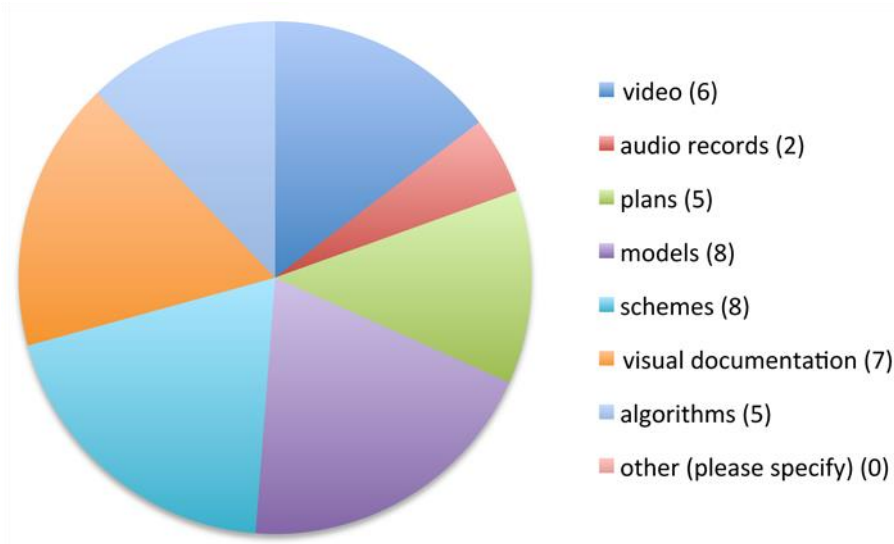
Question 4a presented the manner in which the data were created meaning what processes preceded the data creation: measuring, experiments, testing, surveys, statistical investigation or another activity.

The answers can be seen in the pie chart below, from which it is obvious that none of the activities needed for data creation predominated:

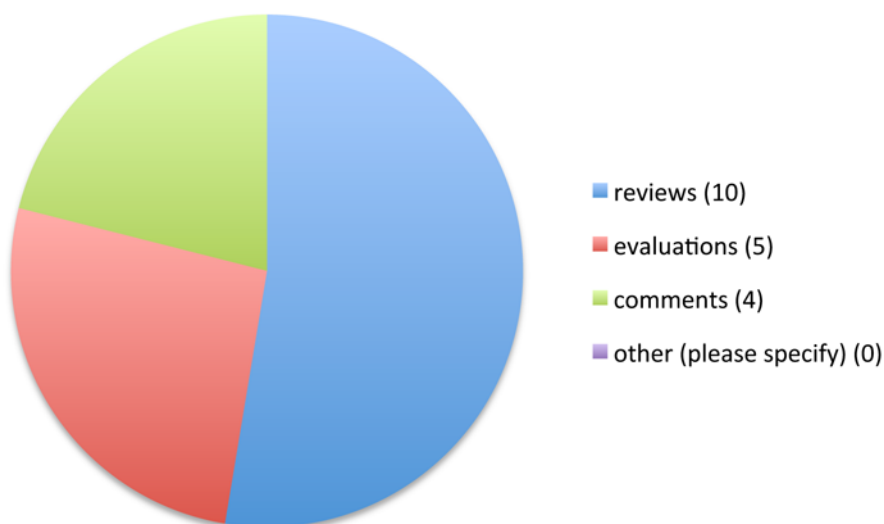


Question 4b presented the type of accompanying material, meaning the material which accompany enhanced publications. The most common types are videos, audio records, plans, models, schemes, visual documentation, and algorithms.

Following the Question 4a it seemed only logical that the most common accompanying material would be connected with measuring and experiments as the most common activities for the data creation, and so models, schemes and visual documentation would be the most common types of such materials:

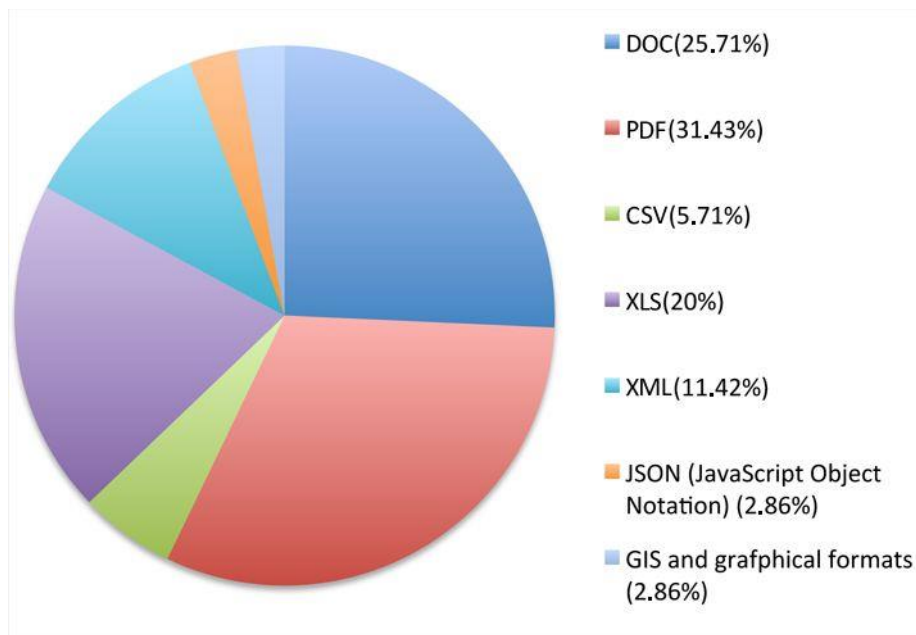


Question 4c was dedicated to the types of materials, which presented post-published data, such as reviews, evaluation or comments. The most common type(s) can be seen in the pie chart below:



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

The multiple choice question was focused on the format of saved data. To the relevant formats DOC, PDF, CSV, XLS, XML and JSON (JavaScript Object Notation) proposed by authors of the project, GIS was added by respondents. However, the most used formats are very common formats widely used also by common users outside the communities of scholars and researchers. Such formats are mainly DOC, PDF and XLS. The frequency of usage of these formats may be seen below, in a pie chart:

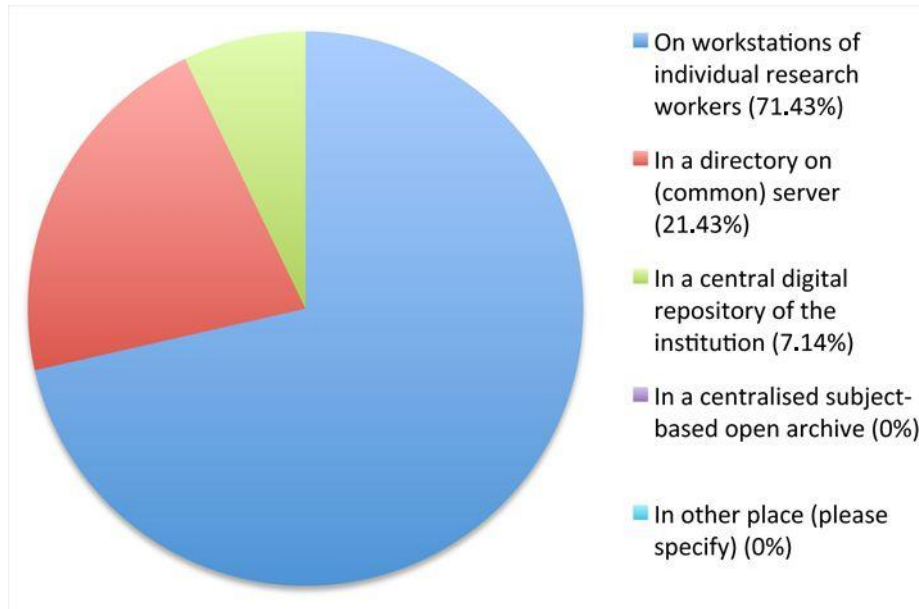


*Question 6: Where are the raw research data of your institution stored?*

From the point of open access, general accessibility, security, reliability, interconnectivity, it is important to divide the manner of storing the data into the following categories:

- on workstations of individual research workers;
- on a directory on (common) server;
- in a central digital repository of the institution;
- in a centralized subject-based open archive;

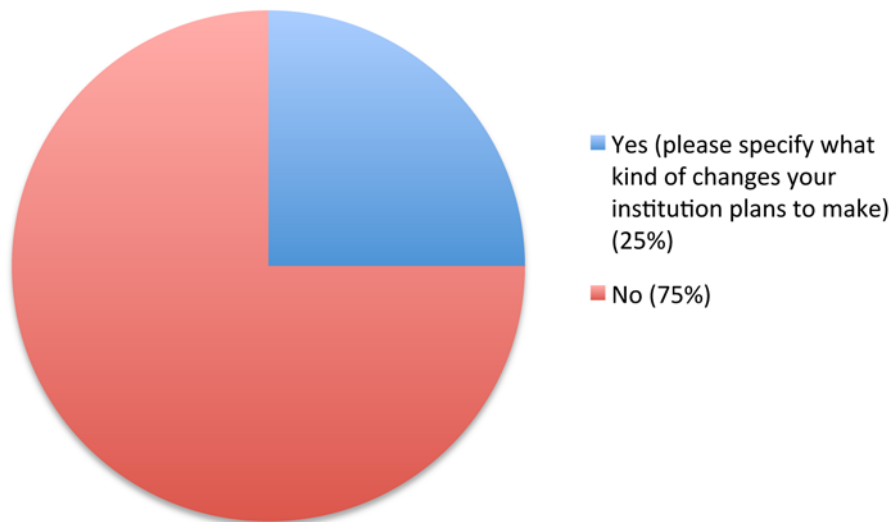
Following this and according to the survey, the Slovak institutions still miss by the institution managed centralized secured storage and in a great deal the institutions rely on researchers who keep the data on their workstations as can be seen in the chart below:



*Question 7: Does your institution plan any changes in the way of storing and archiving the research data?*

Following Question 6, this question aims to find out if there are any changes intended. In this case it would mean, if there are any plans to move the data from the researchers' workstations and devices elsewhere.

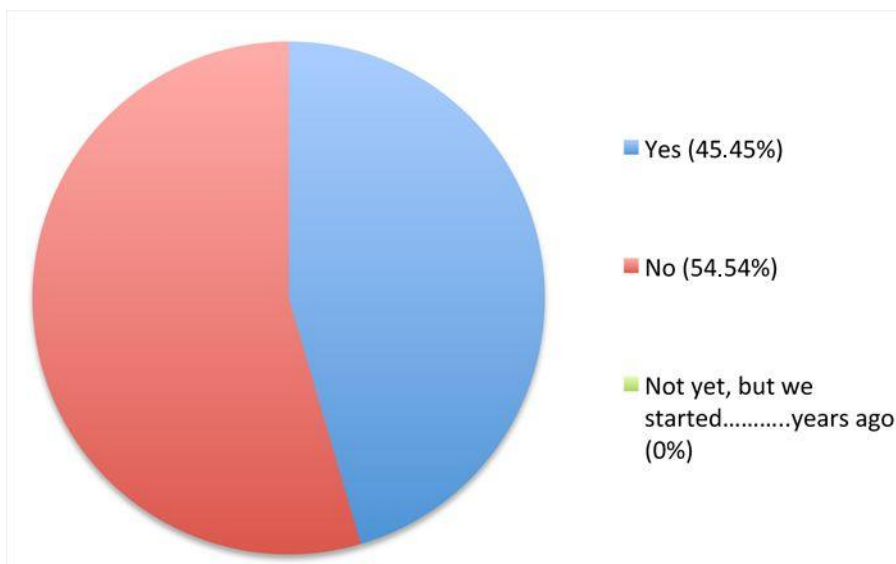
However, there seems to be no/small change intended or planned as can be seen in the chart:



However, no respondent who answered there was a plan to change the actual situation specified what change is planned so it is not possible to identify any need which would result from such intentions.

*Question 8: Are the research data archived in your institution for more than 10 years?*

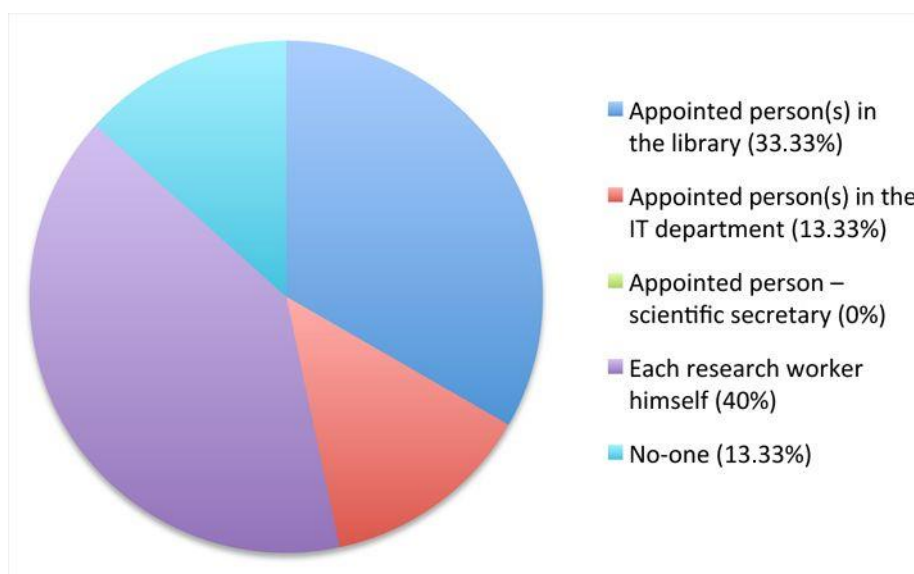
The question focuses on time spent on archiving.



The chart above shows that approximately half of respondents have archived their data for more than 10 years.

*Question 9: Who in your institution is responsible for (takes care of) storing and archiving the research data?*

This part of the survey shows if there is a person responsible for storing the data and so may be seen as a centralized system within an institution where the data are systematically stored.



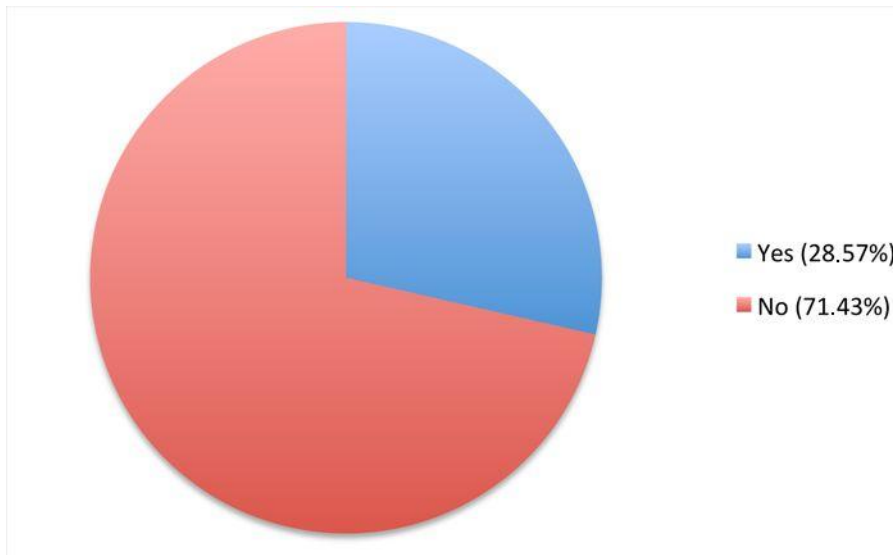
The answers seem to relate with the Question 6 “Where are the raw research data of your institution stored?” and so the highest proportion of the data is stored by individual research workers on their workstations.

**Page 3: 11 questions on Dissemination and Reuse of the Research Data**

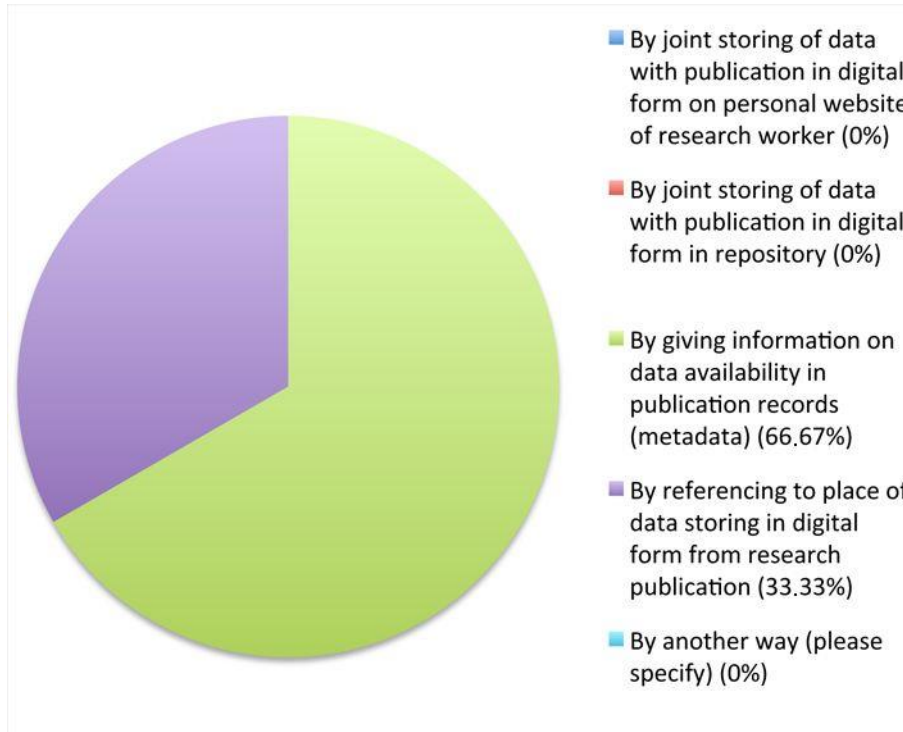
*Question 10: The main question is a yes/no question “Are the research data produced in your institution linked with the research papers?” In case of “yes” answer, the question “How are the research data linked with research papers?” followed. In case of “No”, the question “Are you planning to link data with research publications in your institution?” followed.*

The chart below shows that there was a smaller proportion of “Yes” answers:



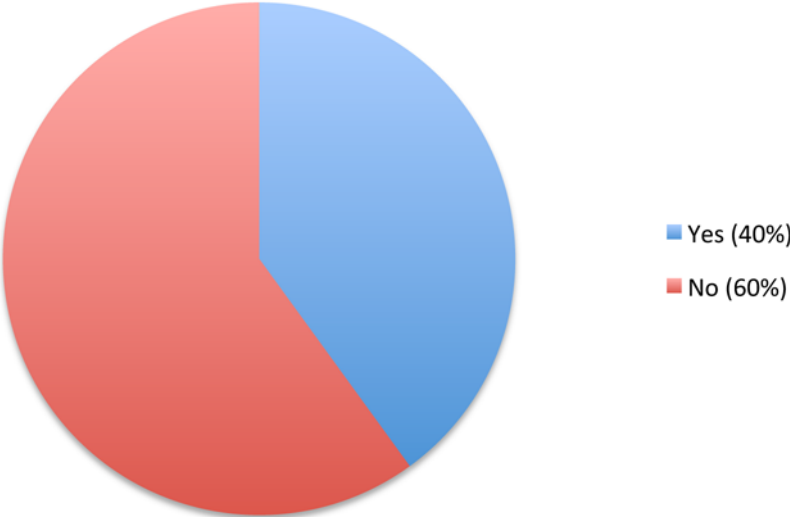


Therefore, the connected Question 10a “If yes, how are the research data linked with research papers?” had just a few answers. The “by giving information on data availability in publication records (metadata)” answer was the most popular one in case the produced data are linked with an enhanced publication.



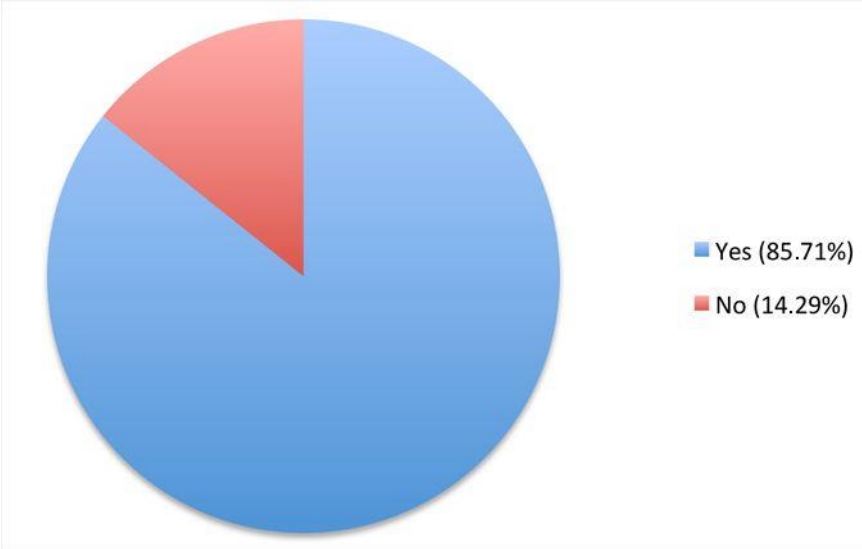
Most respondents who answered “No” to Question 10 also replied that their institution is not planning to link the data with the enhanced publications to question 10b: “Are you

planning to link data with research publications in your institution?" as can be seen in the chart:



*Question 11: Are the data produced in your institution reused in other research projects and by other research workers?*

The answers to this question showed that the majority of the responding institutions provide the data to other researchers and that the data are used by them or the data are used again in other projects. The proportion can be seen in the pie chart below:



*Question 12: Are the data produced during the research work in your institution available for researchers from other institutions?*

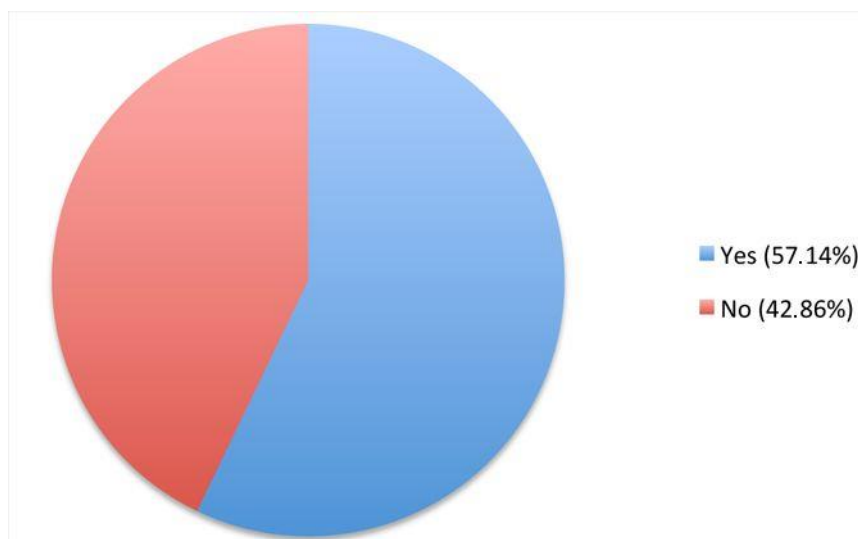
The “Yes” answer was followed by question 12a: “How does your institution provide the research data for researchers from other institutions?”

The “No” answer was followed by

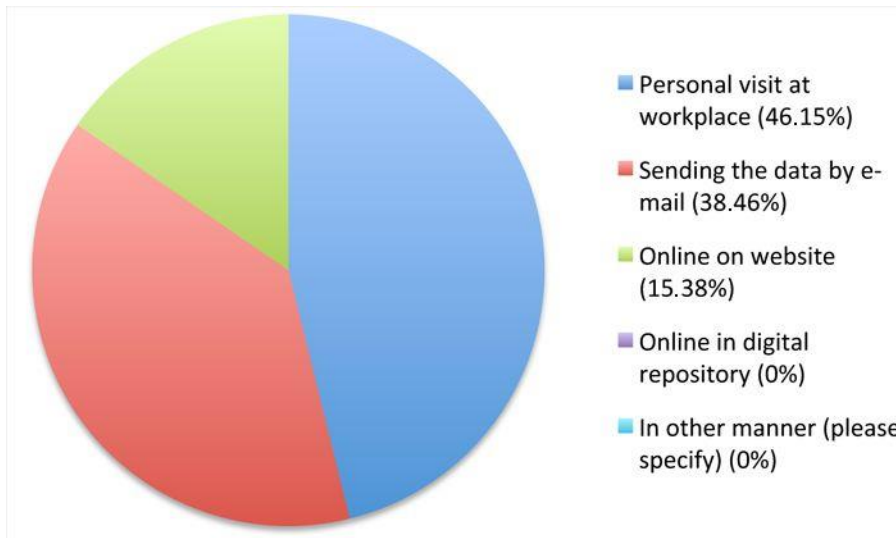
- question 12b: “Why are the research data not available for the researchers from other institutions?”

- question 12c: “Are you planning to provide data stored by your institution for the outside researchers in the near future?”

Answers to the main question showed that higher percentage of institutions do make their data available for another institutions or their researchers:

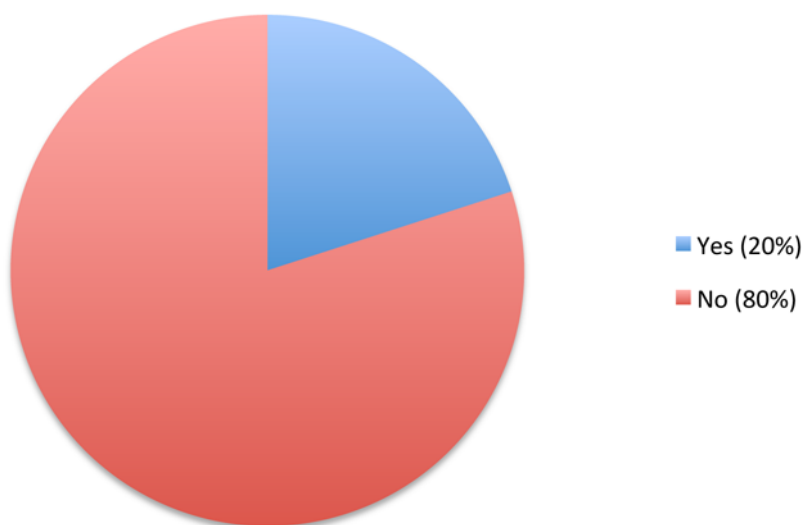


The answers to the following question 12a showed that most often this happens during the personal visit of the researchers at the hosting institution or via email. No responding institution provides the data via an online digital repository.



Question 12b following the “No” answer to the main question 12 had unfortunately only two answers, one being “I am not sure” and other “Copyright issues”. However, such a low response amount would require further investigation into what exactly lies behind the unavailability of the data.

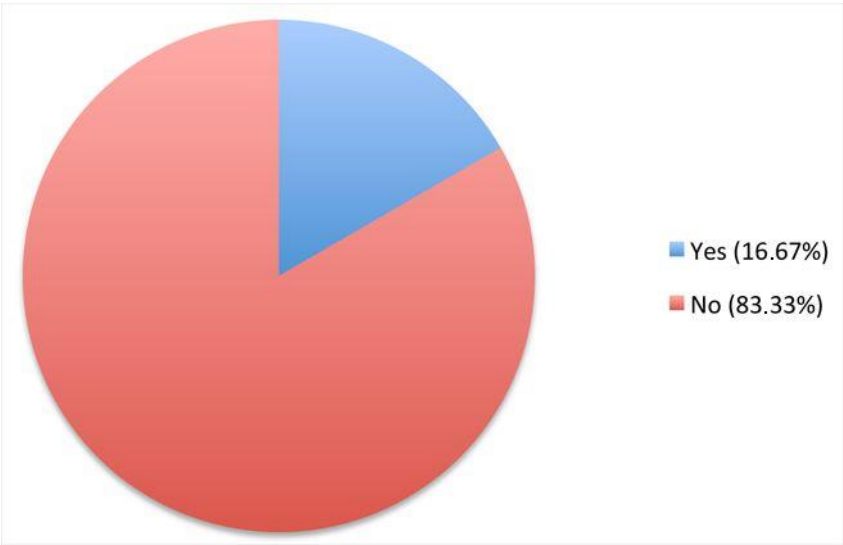
Question 12c, also following the “No” answer to the main question 12 brought more insight into plans of institutions and showed that institutions are mostly not planning to change the current situation or that there are no plans/projects to make the data available. The difference between the proportion of institutions planning a change and institution not planning any change or having no project can be seen in the pie chart below:



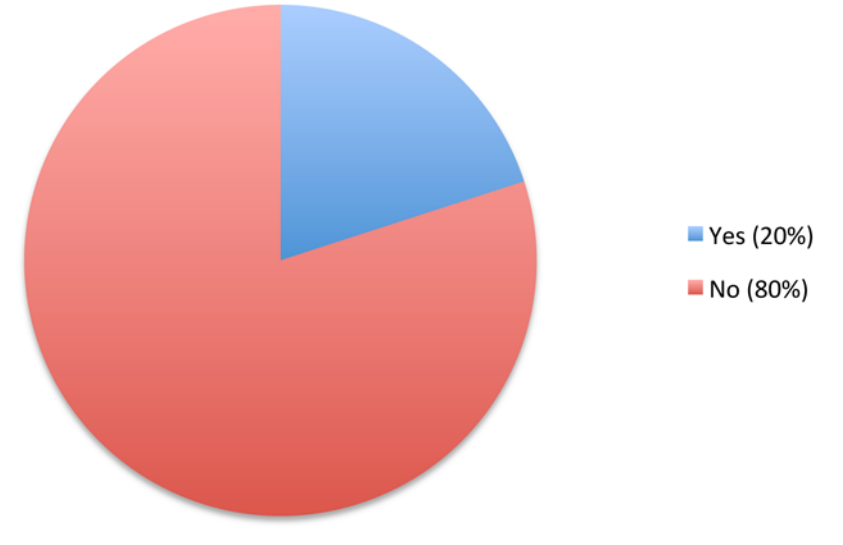
*Question 13: Do you use public licenses at your institution for providing and marking the research data?*

In case the answer was negative, the main yes/no question was followed by question 13a: “Are you planning to use public licenses in your institution in the near future?”

The main question “Do you use public licenses in your institution for providing and marking the research data?” showed quite rare usage of public licenses:



Similarly to the questions 12, there are no plans or projects focused on change of the situation and starting of usage of such licenses:



*Question 14: If there was a project to build a long-term central interinstitutional repository for research data, would your institution be interested in participating in it?*

The last question was focused on the will to participate in projects aimed at changing the current situation and participating in building a common repository where the data would be stored. However, this question was not answered by more than 50% of the respondents. There may be various reasons why there were so few answers (for example, incompetence to answer, hypothetical question, unclear status of participation, etc.).

On the other hand, 100% of the respondents responded in a positive manner.



**Page 4: 4 questions on identification**

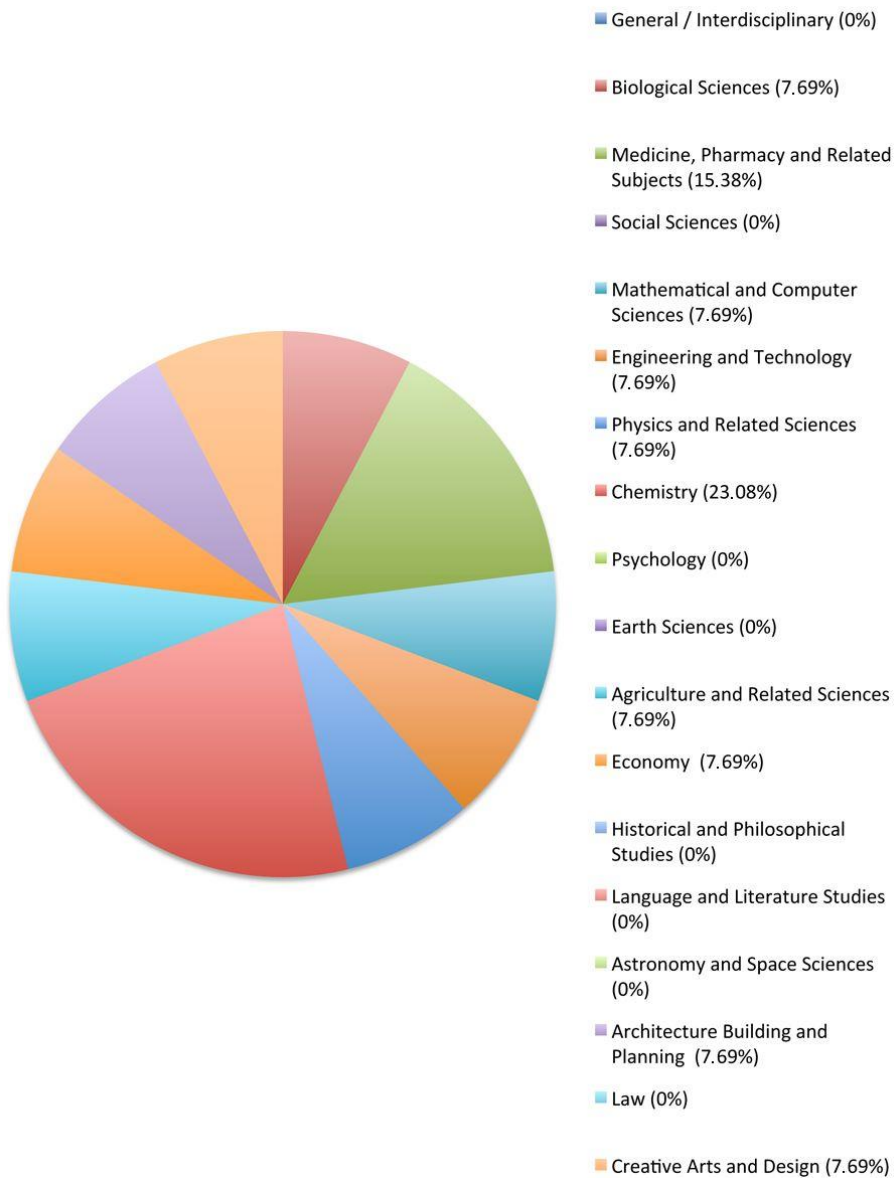
*Question 15 and 16: included personal data of respondents to have a clearer picture of which institution they represent.*

*Question 17: Type of institution (shown in diagram):*



- Higher education institution (100%)
- Research institute (0%)
- Academy of Science (0%)
- Auxiliary scientific unit (0%)

*Question 18: Field of specialization:*



## Conclusion

The survey demonstrated that the questions of Open Access, research data, their availability and storage, licenses and other important issues are only beginning to be open or in some cases not even open yet. Many institutions do not have the online repositories nor work on their availability. The institutions showed willingness to cooperate on such repositories; however, they miss the concept, the systematic plans to open their data, which they often share with other researchers. Also they may miss the necessary staff, know-how or leadership.



## References

Beall, Jeffrey. *Predatory' Open-Access Scholarly Publishers*. **The Charleston Advisor**, April 2010.

Boston University Libraries. *What is "Research Data"?* **Research Data Management**.  
<http://www.bu.edu/datamanagement/background/whatisdata/>

Brandon University. <http://libguides.brandonu.ca/content.php?pid=223976&sid=1857707>

DCC. **How to Develop a Data Management and Sharing Plan**.  
<http://www.dcc.ac.uk/resources/how-guides/develop-data-plan#Why%20develop>

Hogenaar, Arjan. **A short introduction to enhanced publications**.  
<https://www.openaire.eu/en/component/content/article/76-highlights/344-a-short-introduction-to-enhanced-publications>

Hoorn, Esther and Maurits van der Graaf. **Copyright Issues in Open Access Research Journals**. D-Lib Magazine, February 2006  
<http://www.dlib.org/dlib/february06/vandergraaf/02vandergraaf.html>

**Horizon 2020**. Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020  
[http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi-oa-pilot-guide\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf)

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

Morrison, Heather. <http://poeticeconomics.blogspot.hu/2013/04/elseviers-twist-on-open-access-and.html>

Steven, Harnad. <http://users.ecs.soton.ac.uk/harnad/Hypermail/Amsci/3151.html>

Suber, Peter. **Open Access**. MIT Press, June 2012.

Trochim, M.K. William. *Types of Data*. **Research Methods Knowledge Base**.  
<http://www.socialresearchmethods.net/kb/datatype.php>

University of Rhode Island. **OA journals: quality and legitimacy**.  
<http://uri.libguides.com/oafund/oajournals>