

## **ENHANCED PUBLICATIONS IN THE CZECH REPUBLIC**

PETRA PEJŠOVÁ, HANA VYČÍTALOVÁ

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

The National Library of Technology, Czech Republic

### **Abstract**

The aim of this paper is to introduce the notion of enhanced publications to the Czech audience. Enhanced publications are a new type of scholarly publishing. This concerns extra materials linked with electronic publications, such as research data, models, algorithms, figures, metadata, post-publication data (e.g. commentaries) and other items. The paper includes a presentation of a survey of the state of enhanced publications in the Czech Republic.

### **Keywords**

Enhanced publications, research data, research publications, research institutions.

### **ENHANCED PUBLICATIONS**

The Czech audience is more familiar with the English term “enhanced publications” than with its Czech equivalent “přidružené publikace”. According to the definition from the DRIVER-II project, enhanced publications are publications that are enriched with three categories of information – research data, extra materials and post-publication data. (DRIVER, 2013a) Research data include for instance measuring records, experiment results etc. Extra materials means models, algorithms, images, metadata, etc. Post-publication data include commentaries, ranking and other types of information that are produced only subsequently.

This constitutes a next level of scholarly publishing, the objective of which is no longer just a publication as a text, but a work which should at the same time contain such features as underlying data of models, algorithms, etc. Adding these basic details from research to publications facilitates verification, reproduction and reuse of research results. The strength of enhanced publications rests in description of the relationships between underlying data and research output in a meaningful way, both in a readable and machine-processable form. (SURF, 2013) Reference possibilities enable researchers to link any type of object to another object. It is thus easy to link an article to lectures, a review or an interview which may be found on the Internet at various places. This way, the relationships between individual types of information can be described in a structured manner and at a single location. This is what we call enhanced publications.

The European network of cooperating scholarly repositories, DRIVER, was one of the first entities to take an interest in enhanced publications. DRIVER's mission is to optimize the electronic infrastructure used for storage of knowledge in such a manner that value is added to primary research data and information from secondary research becomes more effective.

(DRIVER 2013b) This knowledge is valuable for the industry and enhances cooperation between research and education. Thanks to its network of freely accessible digital repositories, DRIVER presently provides access to more than 3.5 million scholarly publications, articles, dissertations, lectures, reports and other types of documents from all disciplines of science. There are 295 digital repositories from 38 countries of the world participating in the Driver network.

DRIVER development is ensured thanks to the DRIVER II project. One of the primary objectives of the project is a wider coverage through enhanced publications. The study "Enhanced Publications: Object Models and Functionalities", which defines the requirements for storing and managing of enhanced publications within the DRIVER infrastructure, was produced as part of the DRIVER II project. This report can be freely downloaded at [http://www.driver-repository.eu/component/option,com\\_jdownloads/Itemid,83/task,view.download/cid,54/](http://www.driver-repository.eu/component/option,com_jdownloads/Itemid,83/task,view.download/cid,54/) .

Promoting of the concept of open enhanced publications is also one of the aims of the OpenAIRE<sup>1</sup> portal. „Creating a robust, participatory service for the cross-linking of peer-reviewed scientific publications and associated datasets is the principal goal of OpenAIREplus. As scholarly communication touches upon many disciplines, the project's horizontal outreach will facilitate collaboration across data infrastructures, providing information to scientists, non-scientists as well as to providers of value-added services. The project will establish an e-Infrastructure to harvest, enrich and store the metadata of Open Access scientific datasets. Innovative underlying technical structures will be deployed to support the management of and inter-linking between associated scientific data.

Access to and deposit of linked publications via the OpenAIRE portal will be supported by a Help Desk, and OpenAIRE's collaborative networking structure will be extended to promote the concept of open enhanced publications among user communities. Liaison offices in each of the project's 31 European countries work to support the needs of researchers in Europe. The project will also actively leverage its international connections to contribute to common standards, data issues and interoperability on a global level.“ (OpenAIRE, 2011)

A significant initiator of the idea of making enhanced publications more available is the Dutch holding SURF<sup>2</sup>, which brings together Dutch research universities, universities of applied sciences and research institutions. These institutions collaborate on innovative projects with a view to improve the quality of tertiary education and research in the field of information and communication technology. In 2011, the SURFshare programme provided funding for six projects that enabled researchers to gain experience with creation of enhanced publications. These projects are described in a short film, in which the researchers explain how the links between research outputs provide a broader view, higher quality and more opportunities to develop academic research. The video is available on YouTube in the

---

<sup>1</sup> <http://www.openaire.eu/en/component/content/article/326-openaireplus-press-release>

<sup>2</sup> <http://www.surf.nl/en/themes/research/research-data-management/enhanced-publications/index.html>

*Seminar on Providing Access to Grey Literature 2013: The 6th year of the seminar focused on storage and providing access to the grey literature, 23th October 2013 [online]. Praha: National Library of Technology, 2013. Available at WWW: <http://nusl.techlib.cz/Sborniky>. ISSN 1803-6015.*

---

SichtingSURF channel under the title “Enhanced Publications: from experiment to practice”<sup>3</sup>. There are numerous other videos on enhanced publications available in the SichtingSURF channel.

Thanks to the SURF holding, various supporting materials have been created, all of them available in English. One of them is the book *Enhanced Publications: Linking Publications and Research Data in Digital Repositories*<sup>4</sup>, which provides an up-to-date overview of the basic features of enhanced publications. The main finding is that the publishers and operators of digital repositories have the tools for creation of all three types of enhanced publications, but do not use them. Publishers and operators of digital repositories should offer services and tools to add research data, extra materials and post-publication data to publications. Researchers should be responsible for the content. Another interesting book is *Emerging Standards for Enhanced Publications and Repository Technology*<sup>5</sup>, which deals with two areas – new technologies and communities and interoperability. The part devoted to new technologies and communities contains a description of networks, long-term data storage, strategies and projects concerned with the topic of enhanced publications and also the European CRIS community (Computer Research Information Systems). In the second part, the authors focus on new standards, formats and development of digital repositories, as all of this is connected with interoperability of enhanced publications. The last of interesting outputs mentioned here is a visualizer of enhanced publications, making objects in enhanced publications accessible through a pleasant, intuitive and friendly interface. The visualizer can be freely downloaded at this link <http://wiki.surf.nl/display/vp/4.3+InContext+Visualiser>.

Another noteworthy example from the Netherlands is the DANS<sup>6</sup> (Data Archiving and Networking Services) system, which takes care of a digital repository specializing in long-term archiving of research data. DANS is operated in the EASY<sup>7</sup> system and provides an opportunity for online storage and access to thousands of data files from the humanities and social sciences as well as other disciplines. (DANS, 2013) This concerns in particular scholarly datasets, e-publications and other research information. DANS is available for free for downloading and storing of files, but registration is required. DANS stores data in a sustained manner and enables them to be reused. The files are made available to other researchers under the conditions defined by the depositor. The users undertake to cite the source of data in their publications and to link the publication to data in DANS. In addition, DANS also provides training and consulting and performs research into sustained access to information. DANS is operated jointly by the Royal Netherlands Academy of Arts and Sciences (KNAW) and the Netherlands Organisation for Scientific Research (NWO). As part of its mission, DANS promotes open access, even though it acknowledges that not all

---

<sup>3</sup> <http://www.youtube.com/watch?v=FHI4J94YUUK&list=PLA936B4E394A29093&index=16>

<sup>4</sup> <http://dare.uva.nl/document/150723>

<sup>5</sup> <http://dare.uva.nl/aup/nl/record/316870>

<sup>6</sup> <http://www.dans.knaw.nl/en/content/data-archive/depositing-data>

<sup>7</sup> <https://easy.dans.knaw.nl/ui/home>

research data can be freely available without any limitation. However, even in cases when data are not freely available or access to them is limited, it is essential that research data are archived in a sustained manner. In order to ensure that archived data can be still found, made available and reused in the future, Data Seal of Approval<sup>8</sup> has been developed within DANS. This seal guarantees that a data repository complies with several clearly outlined criteria as regards quality, protection and accessibility of data. DANS also finances small data projects, such as feasibility studies and pilot solutions for data archiving in new scientific disciplines. Data stored in DANS are accessible also via the Dutch information portal of scientific information NARCIS<sup>9</sup>. NARCIS provides information on research, programmes, projects, researchers, research institutions and their profiles, including data files and enhanced publications.

The European Union is in its policies not directly concerned with enhanced publications, but the policy of open access in scholarly publishing, whose aim is to facilitate development of education, promote innovations and creative reuse of the results of European research, applies to them. The framework EU programme for research and innovations Horizon 2020 has laid down the obligation to store research data necessary for verification of the results presented in scholarly publications that form the output of projects in open access mode, see clause 29.2 (a) of the handbook Multi-beneficiary General Model Grant Agreement: Version 1.0<sup>10</sup>. Under clause 29.3, the obligation to store research data in open access mode is imposed only on recipients of support within the activities of "Research Data Pilot: Regarding the digital research data generated in the action"

## **SURVEY OF ENHANCED PUBLICATIONS**

The state of enhanced publications has not been mapped in the Czech Republic yet. Therefore, the National Library of Technology has carried out a survey this year. The aim of the survey was to find out what research data are produced and archived by research institutions. Furthermore, we were also interested in the manner of data storing and archiving and, most importantly – whether data are linked with research publications and we can therefore talk about enhanced publications in the Czech Republic.

The survey was targeted on public research institutions, private research institutions and other institutions engaged in research. We addressed the representatives of these institutions that were expected to have a comprehensive knowledge of the research activities of the institution concerned; we chose persons holding the posts of science officers, science secretaries or deputy directors for research or science. In cases when it was not possible to determine who holds such post in the institution or no such posts have been set up by the institution, representatives from the institution's management were selected.

---

<sup>8</sup> <http://www.datasealofapproval.org/en/>

<sup>9</sup> <http://www.narcis.nl/>

<sup>10</sup> [http://ec.europa.eu/research/participants/data/ref/h2020/mga/gga/h2020-mga-gga-multi\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/mga/gga/h2020-mga-gga-multi_en.pdf)

The survey was performed through online questionnaire, consisting of nineteen questions. This was a structured questionnaire; most of the questions were closed with one or more answers possible. Some questions followed in the previous ones, depending on the answer chosen for the preceding question. One of online applications on the web was used to disseminate the questionnaire.

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

The survey was performed from 11 September till 7 October 2013. First, we received only a small portion of completed questionnaires, therefore we once again requested the respondents to fill in the questionnaire over the phone. In total, we addressed 113 institutions, or rather their representatives, and obtained 66 completed questionnaires. Four institutions refused to or were not able to complete the questionnaire. The reason for the refusal was research information secrecy; the representative of the given institution did not want to disclose any information. Another reason stated was the fact that the institutions did not have any research data as they directly formulate the summary of findings from the research in published output. One institution stated that it is not directly engaged in research, but rather in preparation of supporting materials for research. A detailed summary of the institutions addressed and obtained answers is provided in the following table and chart.

Institution type	Institutes of Academy of Sciences	of Public research institutions	Private research institutions	Other research institutions	In total
Number of addresses	54	21	34	4	113
Number of responses	33	14	16	2	65
Negative answer	2	0	1	1	4

Table 1 Number of responses

First, it was necessary to get an overview of what research data are produced by research institutions. Individual types of data that can be usually attached within enhanced publications to research publications were listed in the questionnaire. In addition, the respondents had an option to indicate other possibilities. The respondents mostly selected more possibilities. 48% of the results were formed by various types of research data (from measuring, testing, trials...), 42% was formed by accompanying material (visual documentation, videos, models, diagrams...) and 10% were formed by post-publication data (reviews, ranking).

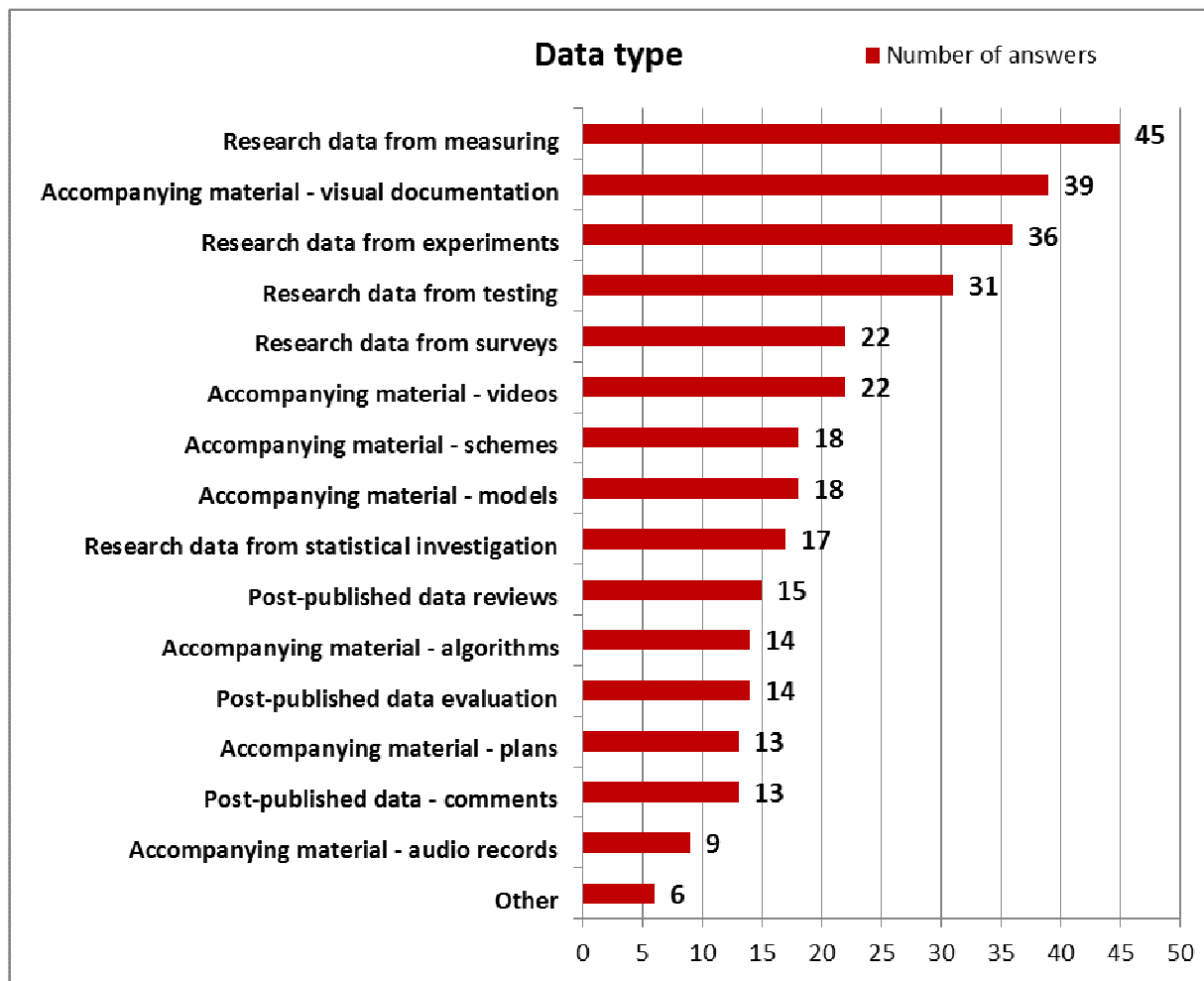


Figure 1 Types of research data

Data stored and archived in PDF, DOC and XLS formats were most frequent. Other data formats were not included so frequently. XML, CSV or image formats (JPEG, TIFF etc.) belonged to the more common formats too. Less common were SAV, CIF formats and various formats used in cartography.

We were also interested in the total volume of data. We asked the respondents at least for a rough estimate; the answer was not obligatory. Nonetheless, it seems that this question was difficult to answer as only a small number of respondents provided an answer to it. The total volume of research data is probably also related to the field in which the institution is engaged. The usual answer was in the order of tens or hundreds of gigabytes; several answers reached the order of terabytes.

Another significant area to be investigated was the question where and how research data are stored and archived. The respondents were again offered several options and could choose one or more of them or add their own one. The answers confirmed our hypothesis that most often data are not stored centrally in a shared repository, but that they remain on workstations (computers) of individual researchers. This answer was selected 52x. Other frequent manners

of storage were a shared directory on a disc (29 answers) and central data repository of the institution – this answer was selected by 18 respondents. Some of the respondents indicated e.g. two or three of the provided options, which means that the conception of data storage is clearly not entirely resolved and unified in many of the institutions. However, most of the institutions are not going to make any changes for the time being, which was the focus of the following questionnaire question.

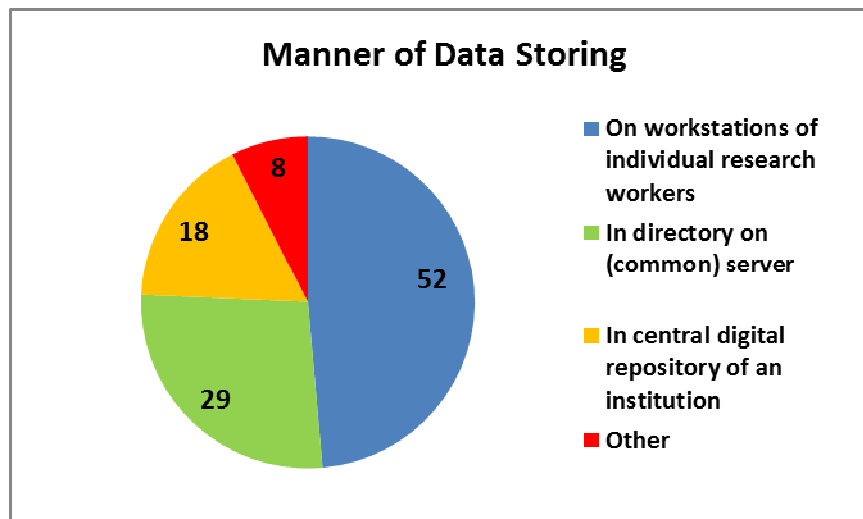


Figure 2 Manner of data storing

Also the question as to whether the institutions archive their data for a period of more than ten years rendered interesting findings. 54% of respondents answered yes, 31 % did not know, but 15% (10 respondents) answered that they do not archive data for such a long time, which is quite a large number. Yet, the majority of the respondents at the same time indicated that they reuse data in further research.

Data are taken care of mostly by researchers themselves; this was the answer of 53 respondents. But respondents could choose more than one option for this question as well. Taking care of the data is further often the responsibility of authorized persons in the library, archive, IT department or of science secretaries or officers or managers.

The core of the survey consisted in the question as to whether we can actually talk about enhanced publications in the Czech Republic, i.e. whether research publications are linked with research data. The questionnaire outlined several answers as to how such a linkage can function in practice. Apart from the ideal method of linkage – joint storage of the publication in electronic form together with the relevant research data in a digital repository – several other alternative options were suggested. It was a relatively big surprise that 58% of respondents stated that they link research publications and data in one way or another. 5 representatives of institutions that do not link publications and data at the moment indicated that they would like to change the situation.

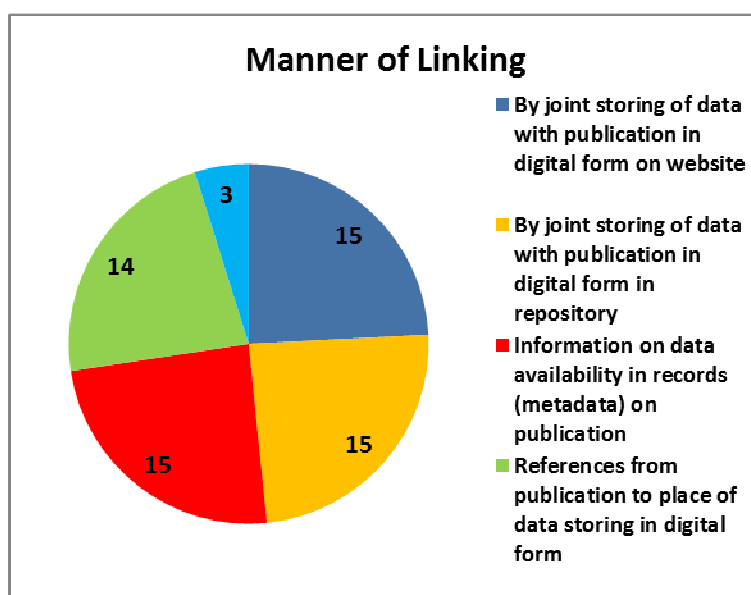


Figure 3 Manner of Linking

Another fact that we were interested in was whether and how the institutions were willing to provide their research data to colleagues from other institutions. Forty-four respondents stated that their data may be provided in some way. However, those interested usually have to visit the research institution in person and collect or request the data. 68% of respondents (44 institutions), thereof 37 public research institutions and 7 private research institutions, would be willing to provide data in this way. 12 public research institutions and 9 private research institutions are not willing to provide data. There are 24 institutions willing to provide data using the ideal method (online on web or online in digital repository). In some cases, the institution must fulfil certain conditions, e.g. obtain consent of the author or research data owner.

If institutions cannot or are not willing to make their data available to interested parties, they were asked to give a reason. The position is clear in the question of commercial research – tailored research. In such case, the data are owned by the client that paid for the research. The client would have to agree with disclosure of the data. Relatively frequent reason was the risk of misuse of the data. In some institutions, the data as such are deemed intellectual property of the author (researcher) who would have to agree with the disclosure of the data. It is also often the case that only the data that are part of final research output, i.e. those that were published in specialist publications or at conferences, are deemed public. The question of patent protection was also mentioned.



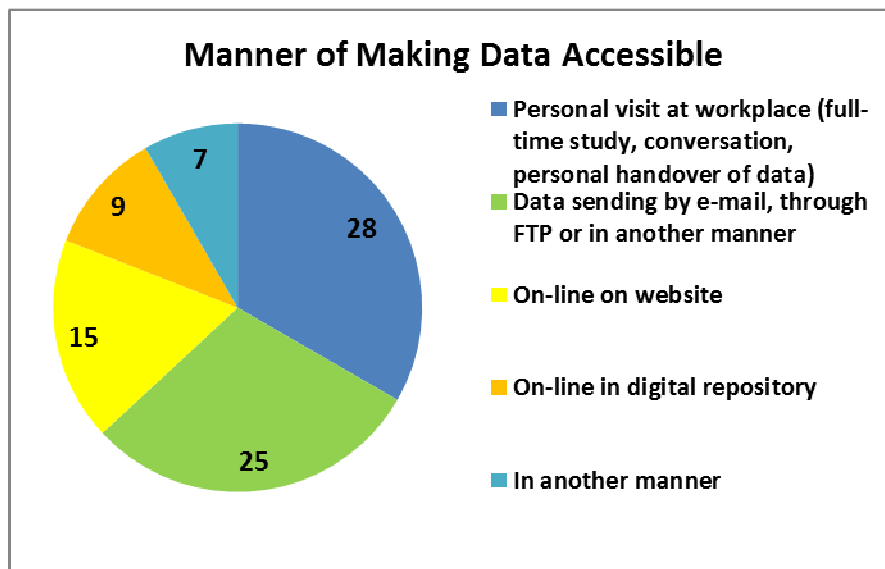


Figure 4 Manner of Making Data Accessible

The last area on which the survey focused was familiarity with and use of public licenses (specifically Creative Commons). Under public licenses the author may provide their work or data to the public under specific conditions. A user may use or distribute the work under the same conditions as determined by the author by means of the selected type of license. Only six percent (4 respondents) stated that they use licenses to designate and thus also make research data available. Two other respondents confirmed that they would like to start to use public licenses.

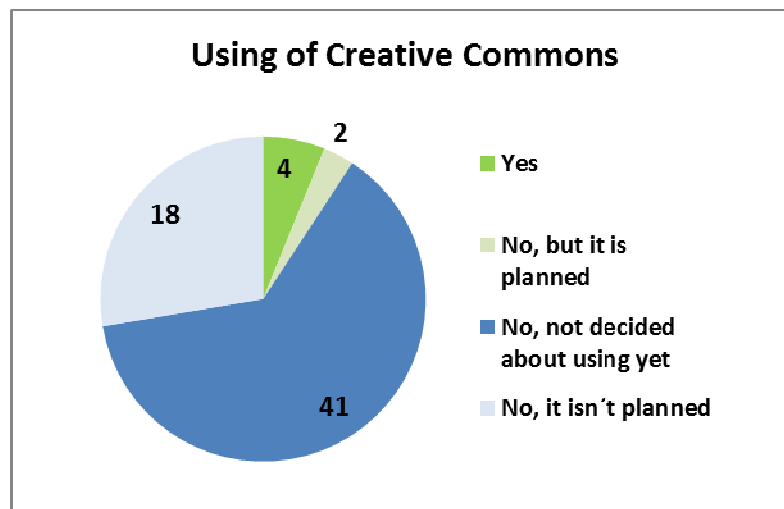


Figure 5 Using of Creative Commons

## CONCLUSION

It follows from the information obtained that a more widespread existence of enhanced publications in the Czech context is not excluded. At the moment, actual enhanced publications, i.e. research publications linked with research data available ideally online, are

at the very beginning. The manner of storing and archiving of research data at Czech research institutions would need improvement. It would be advisable to have a clear concept with a central repository for the entire institution and also greater technical support to researchers.

Improved attitude to the technical aspect of data storage would also facilitate the process of making data progressively available online. Naturally, the copyright question will always be present and individual institutions will need to consider in what mode and under what conditions their data could be made available. Greater public education in the field of copyright and public licenses would be also helpful to ease the concern of both individuals and institutions about data misuse.

## **References**

- About DANS. In: DANS: Data Archiving and Networked Services [online]. 2013 [cit. 2014-01-02]. Available from: <http://www.dans.knaw.nl/en/content/about-dans>.
- DRIVER: Networking European Scientific Repositories. In: DRIVER – Digital Repository Infrastructure Vision for European Research [online]. 2013b, 02 June 2011 [cit. 2014-01-02]. Available from: <http://www.driver-repository.eu/>.
- Enhanced Publications. In: DRIVER – Digital Repository Infrastructure Vision for European Research [online]. 2013a, 18 February 2009 [cit. 2014-01-02]. Available from: <http://www.driver-repository.eu/Enhanced-Publications.html>.
- Enhanced publications. In: SURF: Collaborative organisation for ICT in Dutch higher education and research [online]. 2013, 26-11-2013 [cit. 2014-01-02]. Available from: <http://www.surf.nl/en/themes/research/research-data-management/enhanced-publications/index.html>.
- Paving the way to an open scientific information space: OpenAIREplus – linking peer-reviewed literature to associated data. In: OpenAIRE: Open Access Infrastructure for Research in Europe [online]. 2013, 15 December 2011 [cit. 2014-01-02]. Available from: <http://www.openaire.eu/en/component/content/article/326-openaireplus-press-release>.