conference on grey literature and repositories

the value of grey literature in repositories

proceeding 2014

Conference on Grey Literature and Repositories

Proceedings 2014

„The Value of Grey Literature in Repositories“

National Library of Technology

Prague, 2014

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

Proceedings from the 7th Conference on Grey Literature and Repositories, held on 22nd October 2014 in the National Library of Technology in Prague, Czech Republic.

English conference website (http://nrgl.techlib.cz/index.php/7th_year_of_the_seminar)

Czech conference website (http://nusl.techlib.cz/index.php/7_rocnik_seminare)

Publisher: National Library of Technology, Technická 6/2710, Praha 6, Czech Republic

Editor: Mgr. Hana Vyčítalová

ISSN: 2336-5021

Programme Committee:

PhDr. Eva Bratková, Ph.D., Charles University in Prague

Ing. Martin Lhoták, Academy of Sciences Library

Ing. Jan Mach, University of Economics, Prague

PhDr. Petra Pejšová, National Library of Technology

Mgr. Iveta Rosolová, Charles University in Prague

PhDr. Radka Římanová, Charles University in Prague

Mgr. Hana Vyčítalová, National Library of Technology

Organizing Committee

Mgr. Lenka Patoková, National Library of Technology

PhDr. Petra Pejšová, National Library of Technology

Mgr. Hana Vyčítalová, National Library of Technology

List of Reviewers:

Ing. Lukáš Budínský, Tomas Bata University in Zlín

Dr. Jan Dvořák, Charles University in Prague

Dr. Dominic Farace, Greynet

PhDr. Václava Horčáková, The Institute of History, Academy of Sciences of the Czech Republic

JUDr. Pavel Koukal, Ph.D., Masaryk University

PhDr. Hana Landová, Ph.D., Czech University of Life Sciences Prague

JUDr. Silvia Lattová, Masaryk University

Ing. Martin Lhoták, Academy of Sciences Library

Mgr. MgA. Jakub Míšek, Masaryk University

Doc. PhDr. Richard Papík, Ph.D., Charles University in Prague

Mgr. Věra Pilecká, Ph.D., Czech Technical University in Prague

Doc. JUDr. Radim Polčák, Ph.D., Masaryk University

Mgr. Pavla Rygelová, VŠB – Technical University of Ostrava

PhDr. Radka Římanová, Charles University in Prague

Mgr. et Mgr. Jan Vobořil, Iuridicum Remedium

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

Table of Contents

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

Access to all information sources including grey literature in the Surveying Library

Jiří Drozda, Veronika Synková

jiri.drozda@vugtk.cz, vugtk.vers@gmail.com

The Research Institute of Geodesy, Topography and Cartography

Abstract

The Surveying Library has used the presented version of the INVENIO system not only as its digital repository, but also as a solution for uniform information access to all its funds. Searching and accessing the library catalogue, the bibliographical articles and the digital Kramerius library and access to electronic information resources, grey literature and internal documents are provided via the INVENIO system. INVENIO serves as a comprehensive library system. Part of this article includes a presentation of the BibCirculation module which secures the library's lending services. The Surveying Library is the first organization which use this module in the Czech Republic.

Keywords

Grey Literature, INVENIO, BibCirculation, Library System, Digital Repository, The Surveying Library

The Surveying Library is part of the Research Institute of Geodesy, Topography and Cartography (VÚGTK). The library has a unique and exclusive position not only in the Czech Republic, but also on an international scale thanks to its collections and its specialization in the areas of geography, geodesy, geodynamics, cadasters and metrology. The library is involved in many international activities involving cooperation and the provision of scientific information resources from the areas of its competence. Its uniqueness has been confirmed by the registration of "The Surveying Library®" as a trade mark (hereafter simply referred to as the library) in the Register of Trademarks held by the Czech Industrial Property Office. The library cooperates with the Czech Land Survey and Cadastral Office (hereafter simply referred to as the ČÚZK) in the collection of expert travel reports for all the employees from the ČÚZK, research reports, industry regulations and standards, all resort directives, professional newsletters and journals and the proceedings of scientific conferences and meetings.

The demise of the library was threatened a few years ago. Fortunately, the library and its unique collection managed to survive. The library has had to solve the question of how to most efficiently transform itself so that it can offer the best possible services to all its users. One of the areas where the library has had to change involves the availability and accessibility of information about the library's collections and other information sources. The existing information resources in the library were fragmented. The paper card catalogue was the register of the library's collection.

Only part of the collection had been digitalized in the Tinlib library system which was already outdated and inadequate. The list of available online journals which the library periodically subscribes to was and still is only portrayed in a table on the website. Those customers interested in grey literature (research, annual reports and other materials) had to search in a variety of places and at different levels on the VÚGTK site. In addition, the library also used the Kramerius library system which stored and provided access to digitized documents. Kramerius represented another separate system where customers could search.

As well as the library's collection and the materials on the website, the VÚGTK still collects many other materials in digital form which are only available locally on the institute's network or on its employees' computers. This mainly involves the different standards of the ČÚZKC and the Surveying Office, materials from conferences, analytical and methodological materials, research data, promotional and training materials, output from projects, departmental directives, etc.

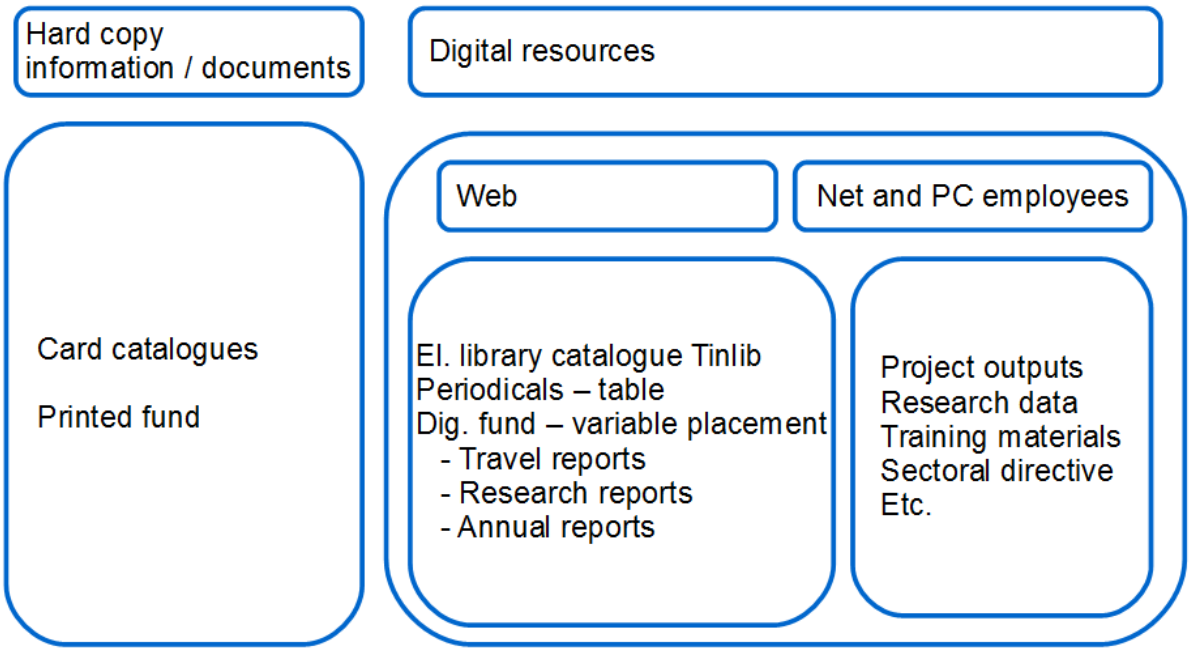


Figure 1: The original location of the information sources

The library, therefore, had to deal with two issues; it needed both a library system and a digital repository within the framework of insufficient finances and capacity building. From the perspective of maintaining the research needs of the VÚGTK, it needed to ensure the availability of a high-quality library and information resources for all its employees. The VÚGTK also wanted to act openly and to share its research output.

At this present time of permanent budget cuts, the library had difficulty in purchasing a completely new library system. This meant that purchases of books, journals and other information resources, as well as the register of readers and loans, were also limited, as was the possibility of cooperating with national portals and enabling searches using the Union Catalogue or the Uniform Information Gateway. These are mainly commercial library systems currently available in the Czech Republic and their prices are in excess of the VÚGTK budget. Open source library systems are not so common in the Czech Republic. The Evergreen [1] and Koha [2] systems are the two most popular.

Another problem which the VÚGTK had to address was how to preserve and make documents and research data available. Only a small portion of scientific output is published in journals, anthologies and monographs, while the remainder is hidden on the computers of the individual researchers or teams, including the primary data. Data back-up is often undertaken by the workers and it is usually non-systematic. The question is "how to disseminate information about your research, how to share primary data (the results of measurements, material analysis, etc.) and at the same time preserve its authorship?" The preferable solution is using an institutional repository with publishing solutions related to the auto-archiving of publishing activities and in an open access mode. As Jindra Planková stated, "At present, it is necessary to understand the access as an open access initiative, a global phenomenon" (Planková, 2007). Rosalia Ferrara mentioned that“OA should not be considered as just a set of standards and

1 http://open-ils.org/
2 http://www.koha.org/

protocols to ensure interoperability between archives and to share resources on a global scale, but as a factor of social growth“ (Ferrara, 2007).

The absence of institutional repositories often leads scientists to try to build archives and disseminate them themselves, for example on CD/DVD media, as described in the case of the personal archives of Ing. arch. Jan Moučka (Vyčítalová, 2011). Currently, the VÚGTK has resolved the problem by publishing the data on its website. Many Czech research institutions have adopted this solution in the past and continue to do so. An example of this is the analysis of the presentation of the full texts from research and technical reports on the Czech web servers (Vyčítalová, 2007). However, this solution has been unsystematic and unsustainable due to the nature of the rapid changes in webpages and problematic from the point of view of finding and sharing the information.

The library sought a solution that would provide the widest range of users with access to all of its information resources from one place. The aim was to publish information about the library's collections and products, online resources, digitized documents stored in the Kramerius system and grey literature.

In 2011, the VÚGTK was approached by the National Technical Library (the NTK) with the request for cooperation with the National Repository of Grey Literature (hereafter simply referred to as the NUŠL). In order to promote cooperation with the NUŠL, the NTK offered the INVENIO system and open source methodology for storing, archiving and making available grey literature with options for its further expansion. The INVENIO system also contains a modular library system which can be localized to the Czech environment.

This offer was accepted, because, like other research organizations, the VÚGTK is focused on the development of information technologies and processes and it is methodologically guided and supported by the work of scientific publications and data. There is a lack of the technological and methodical assistance at the VÚGTK, as well as at other public research institutes, which is provided to large institutions such as the Academy of Sciences of the Czech Republic (ASCR) through the ASCR library. Large projects designed by the ASCR library, the National Library of the Czech Republic and the Moravian Library are not optimal for these small research institutions due to their robust and technical performance. The basic requirement for these institutions is a unified and simple system for managing, storing and disseminating scientific information and data. Another requirement is technical and personal unpretentiousness of the system. The VÚGTK does not have the capacity to provide and maintain multiple systems: a library system, repository publishing, a digital library and more.

The NTK offered a free system for processing preset INVENIO grey literature as one of the forms of cooperation. (Pejšová, 2011) The pre-setup INVENIO library system can be downloaded from the NUŠL website and installed on your own hardware. Installation instructions are also available. Furthermore, the library's INVENIO system can be edited according to the library's needs. The NTK also ensured the localization of INVENIO in Czech. Another important reason involved the fact that stable development was ensured. The European Organization for Nuclear Research (CERN) in Geneva ensures that the INVENIO system has a team of developers which develops the system, regularly publishes new INVENIO versions and has a clearly defined system development for INVENIO for the future. However, the most important reason for this involved the fact that the system library already had a built-in INVENIO module that was able to provide all the library's services. The CERN library fully

activated the INVENIO system from the Aleph system in 2012 without any restriction of services.

The original intention to use the INVENIO system as a library system and a digital repository has grown to an idea to use Invenio system as the discovery system too (use it as a central search point for both the printed and digital collections). The library implemented this idea as follows. The printed catalogue of the paper collection and the article list were transferred from the obsolete TINLIB bibliography system to the INVENIO system. Already digitized titles stored in the Kramerius system have been revised and their records have been linked to the digitized texts in Kramerius. Import records for available prepaid titles of online magazines are being resolved. The records of freely available professional online magazines have been added. We started with the VÚGTK materials as grey literature and internal documents, but access to the internal materials is restricted.

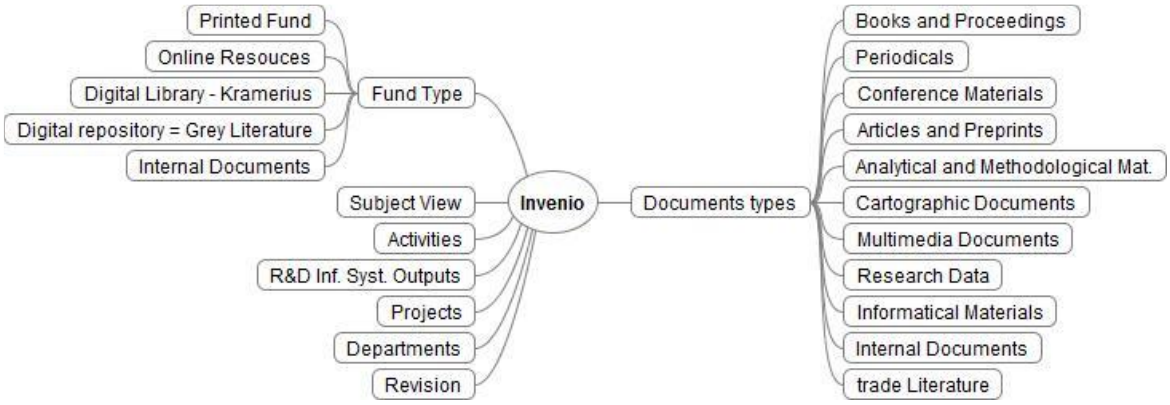


Figure 2: The schematic of the collections in Invenio

As has already been mentioned, various types of records have been recorded in INVENIO. These types are presented in detail in the following examples.

Primarily, they are records of the documents from the library. The library has carried out the conversion of the bibliographic records and/or units which were stored in the TinLib system and migrated them to the INVENIO system. Approximately 40,000 records of monographs, 3,000 records of scientific journals and 148,000 records of articles have been imported. An inspection of the imported records against the documents in hand is currently underway. We assume that the greatest number of records will be checked during the planned

reconstruction of the library. Templates for inserting new bibliographic records have been prepared.

In cases where TinLib records had connected units, those records were also imported into the INVENIO system. They are displayed alongside the record preview for the searched title in the folder marked "Holdings". If the user is logged in, he or she can enter a loan request for the selected units.

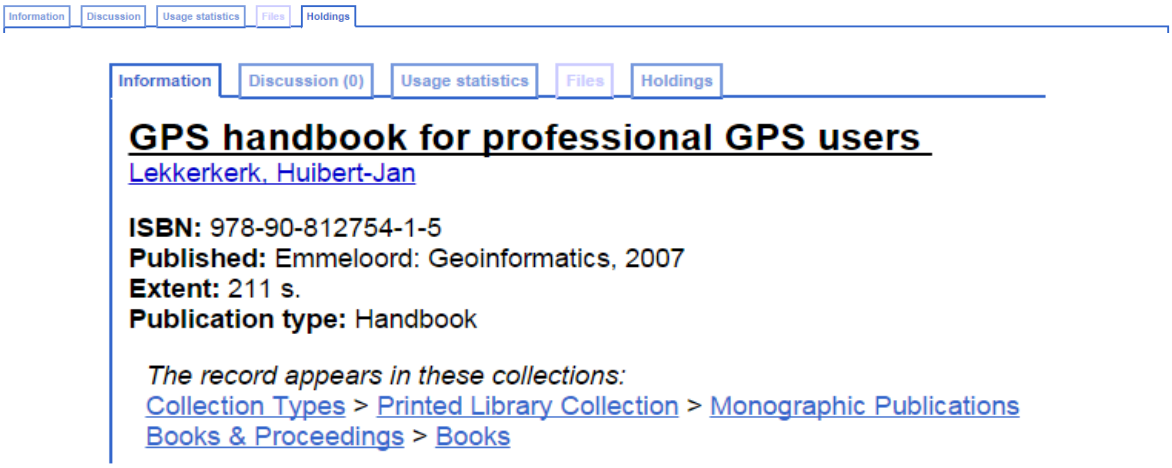


Figure 3: A preview of the imported bibliographic record for a monograph

Another type of records which is available through the INVENIO system involves online journals. Only freely available online magazines and journals have been included. We are currently looking for a convenient way of importing the online journals that the library regularly subscribes to. A link to the full text is displayed in the preview for the online magazines.

Figure 5: A record of an available online magazine

Part of the bibliographic records imported from TinLib included titles that the library had already digitalized and made available via the Kramerius system. The records have also been revised and supplemented with a link to the digitized text in the Kramerius system.

Figure 6: An example of a record with a reference to the Kramerius system

We have also started to include the grey literature in the INVENIO system. Part of the grey literature records is usually linked to full text files which can be opened directly in the INVENIO system. If it is necessary to restrict the access to the full text, this can be achieved using the system tools available at the time of the inclusion of the files. If the titles are also part of the library collection, they have the unit attached in addition to the attached file.

Figure 7: A record of a collection of grey literature with available full text

The INVENIO system is also used in the library to register internal documents. Only authorized users can access the records in these documents. After selecting the records, authorized users can see the details of the title and, if they have sufficient authorization, can also view the attached full text.

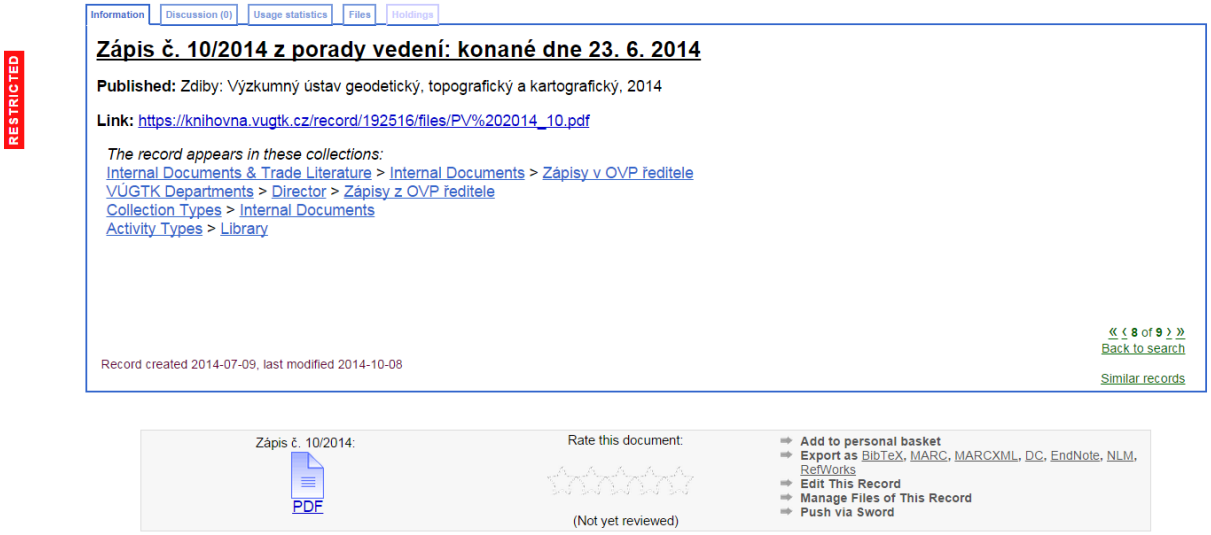


Figure 8: A record of an internal document with restricted access

The BibCirculation module has been activated in order to make full use of the capacity of the INVENIO system. All possible functions/services will be used in this module, including loan services (the register of readers), loan records (including borrowing under the interlibrary loan services), reminders, loan orders and reservations. The INVENIO system also has functions for acquisitions which the library will use in the future

After opening the BibCirculation module and finding the record of a reader, the librarian will receive basic information about the reader and a brief overview of the current number of loan requests, loans and Interlibrary loan services. The list also includes the reader's loans history and requests. Functions for the implementation of a loan, a new loan request or sending an e-mail to the reader are also available.

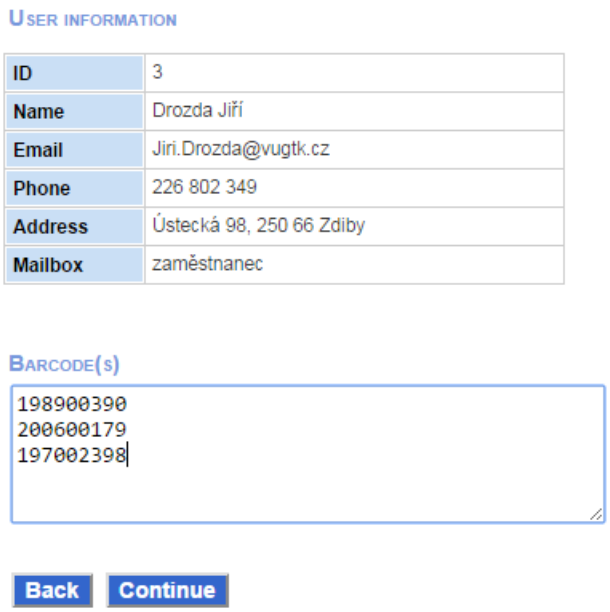


Figure 10: A loan of selected titles to readers based on the reference

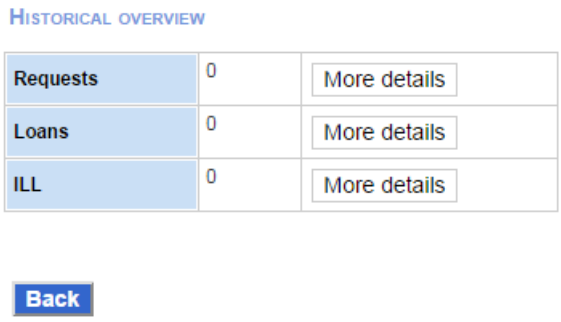
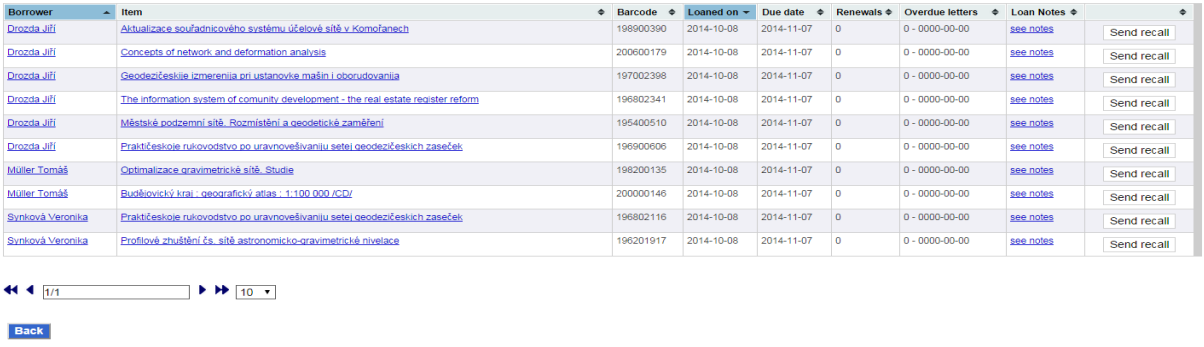


Figure 9: A record of a reader with a brief overview of his

If the reader wants to borrow certain documents, the librarian finds user´s account and performs borrowing through entering acquisition numbers of the intended units. During this process, the librarian has the option to add notes to the loan record or to verify the names of the loans. When a loan is returned, only the acquisition numbers of the returned units can be used.

Currently, the librarian must input the acquisition numbers. The library has started labeling its units with QR codes. The QR code also includes an acquisition numbers. The librarian scans the QR code and acquisition numbers is downloaded into the appropriate field in INVENIO. In addition, the QR code will contain the library's name and the reference number to the bibliographic record in the INVENIO system. The QR code will not only serve to record the borrowing and returning of loaned documents, but also as a source of basic information on the title, the author and the ownership.

The librarian can view various reports; for example, the list of current loans, the list of overdue titles, the loan order summary and reservations. The appropriate functions which can be used with the selected record are always offered.

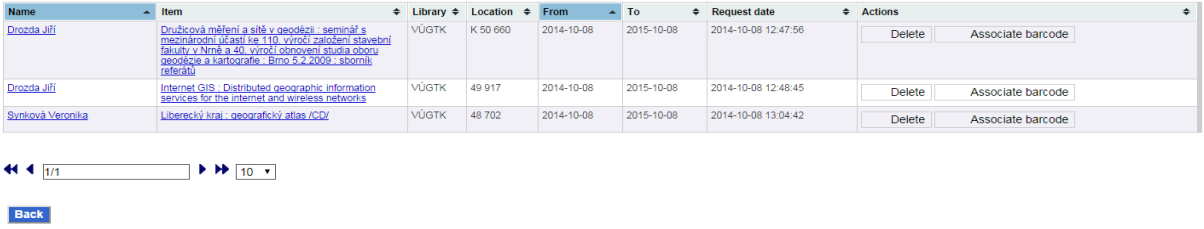


Borrower	Item	Barcode	Loaned on	Due date	Renewals	Overdue letters	Loan Notes	
Drozda Jiří	Aktualizace souřadnicového systému účelové sítě v Komořanech	198900390	2014-10-08	2014-11-07	0	0 - 0000-00-00	see notes	Send recall
Drozda Jiří	Concepts of network and deformation analysis	200600179	2014-10-08	2014-11-07	0	0 - 0000-00-00	see notes	Send recall
Drozda Jiří	Geodezičeskije izmerenija pri ustanovke mašin i oborudovanija	197002398	2014-10-08	2014-11-07	0	0 - 0000-00-00	see notes	Send recall
Drozda Jiří	The information system of comunity development - the real estate register reform	196802341	2014-10-08	2014-11-07	0	0 - 0000-00-00	see notes	Send recall
Drozda Jiří	Městské podzemní sítě. Rozmístění a geodetické zaměření	195400510	2014-10-08	2014-11-07	0	0 - 0000-00-00	see notes	Send recall
Drozda Jiří	Praktičeskoje rukovodstvo po uravnovešivaniju setej geodezičeskich zaseček	196900606	2014-10-08	2014-11-07	0	0 - 0000-00-00	see notes	Send recall
Müller Tomáš	Optimalizace gravimetrické sítě. Studie	198200135	2014-10-08	2014-11-07	0	0 - 0000-00-00	see notes	Send recall
Müller Tomáš	Budějovický kraj : geografický atlas : 1:100 000 /CD/	200000146	2014-10-08	2014-11-07	0	0 - 0000-00-00	see notes	Send recall
Synková Veronika	Praktičeskoje rukovodstvo po uravnovešivaniju setej geodezičeskich zaseček	196802116	2014-10-08	2014-11-07	0	0 - 0000-00-00	see notes	Send recall
Synková Veronika	Profilové zhuštění čs. sítě astronomicko-gravimetrické nivelace	196201917	2014-10-08	2014-11-07	0	0 - 0000-00-00	see notes	Send recall

1/1 10

Back

Figure 11: A summary of the current loans



Name	Item	Library	Location	From	To	Request date	Actions
Drozda Jiří	Družicová měření a sítě v geodézii : seminář s mezinárodní účastí ke 110. výročí založení stavební fakulty v Brně a 40. výročí obnovení studia oboru geodézie a kartografie : Brno 5.2.2009 : sborník referátů	VÚGTK	K 50 660	2014-10-08	2015-10-08	2014-10-08 12:47:56	Delete Associate barcode
Drozda Jiří	Internet GIS : Distributed geographic information services for the internet and wireless networks	VÚGTK	49 917	2014-10-08	2015-10-08	2014-10-08 12:48:45	Delete Associate barcode
Synková Veronika	Liberecký kraj : geografický atlas /CD/	VÚGTK	48 702	2014-10-08	2015-10-08	2014-10-08 13:04:42	Delete Associate barcode

1/1 10

Back

Figure 12: A list of pending requests

A user, who has also set up access to his or her reader account via the web interface, can check the status of his or her membership account online. If the loan period has not been exceeded,

the user can extend an existing loan and/or may cancel loan requirements for any titles which he or she is no longer interested in.

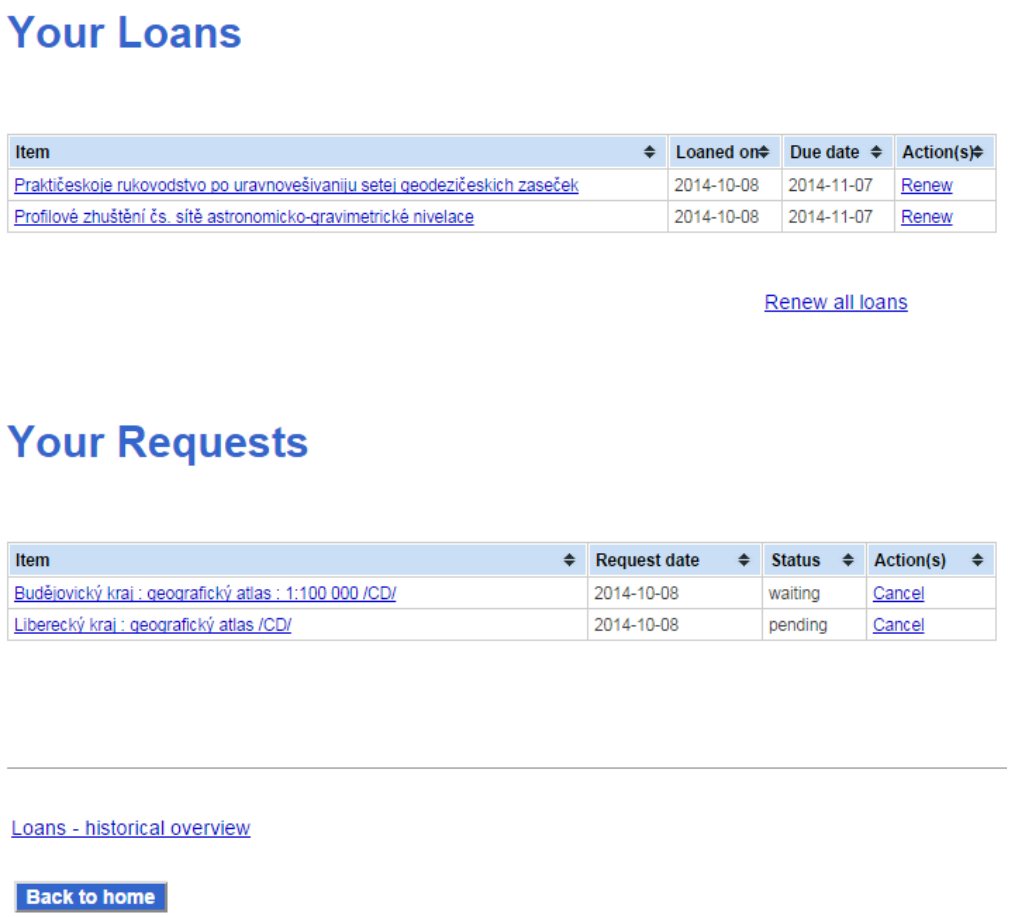


Figure 13: A current reader account with current loans and requests

If a user finds records which he or she wants to save in order to return to them later, work with them, borrow the titles and/or otherwise handle them, the user can save the records and sort them into so-called baskets (folders). The baskets are sorted into thematic groups. Each thematic group can contain multiple baskets. The user can add notes to the entire basket or to the individual titles.

Figure 14: A basket preview

If the user is logged on, the system will store the user's queries. The queries are kept in the history for the next 30 days. The user can use selected queries to set alert services. The results of a regularly set query can be sent via a specified e-mail and/or stored in the basket.

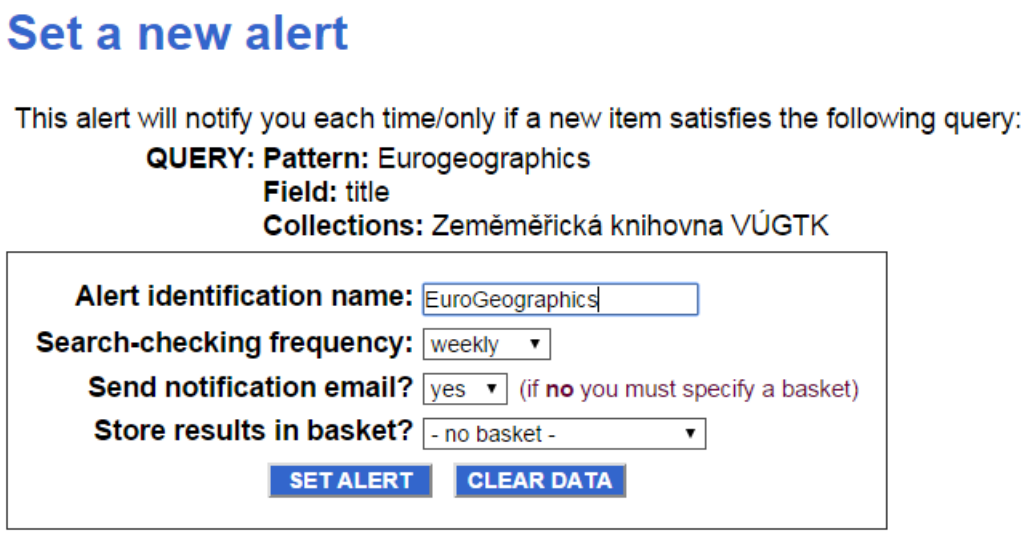


Figure 15: Setting up an alert query

Although the library has used the INVENIO system that the NTK runs for grey literature within the NUŠL project, the library has also greatly expanded its use for many other purposes. The library stores not only the grey literature but also a variety of other types of documents, including its library collection too, into the INVENIO system. Some records refer to sites outside the INVENIO system (Kramerius, the full text of available online journals and magazines).

Compared to the INVENIO version, which is run by the NTK, the library has also added some new elements such as MARC21, different templates and also the adjustment of the indexes for searching and displaying thumbnails of bibliographic records to its installation. A major extension includes the activation of the BibCirculation module and the loan service functions in this module. There has been no experience with this module in the Czech Republic to date. The activation of the BibCirculation module has brought new changes and tasks involving familiarizing the register readers with the available functions, entering orders, borrowing and returning all related instruments for the loan system. It also coincided with the need to localize the module into the Czech language. The localization has been time consuming and has still not been fully completed.

The main aim of the library is to fully satisfy both internal and external users and to always provide accessible and transparent information about the library's collection, subscribed and freely available online resources and other information materials that the library can offer all its users. The library is trying to achieve this goal by using the INVENIO library system which represents a single interface enabling searches in all the library's information sources. The use of the INVENIO system also provides access to many full texts which are attached to the bibliographic records or connected via a link to the bibliographic record. The access to the full text units naturally respects both the license rights for prepaid online resources and copyright and the internal nature of the information which is stored in the collections for grey literature and the internal institutional repositories. The access rights to the different user groups are distinguished using the INVENIO functions and features.

Another important function which the library fulfils is loan services. The librarians have all the necessary tools to the loans and the related information and information about the readers. After logging on, registered users can find a summary of their loans and orders. They are able to take advantage of the history of entered queries, to save selected records in a basket and/or to set a selected query so that it regularly informs them about new records related to the theme which they are interested in.

The library is still merging and optimizing the settings for the proposed procedures and processes within the beta phase. We believe that, after the planned reconstruction of the depositary and reading room, the library will be able to offer all the users its resources and services in the new cover. We believe, satisfied internal and external customers will be the result of these changes.

References:

FERRARA, Rosalia a Poltronieri, Elisabetta. Free on the Net: accessing digital research output with no barriers. In *INFORUM 2007: 13. annually conference* [online]. Prague: Albertina Icome, 2007, 11 p. [cit. 9. October 2014]. Available from: <http://www.inforum.cz/pdf/2007/poltronieri-elisabetta.pdf>.

PEJŠOVÁ, Petra. National Repository of Grey Literature. In *Seminar on Providing of Grey Literature 2011 : 4th annually seminar*, 25. October 2011 [online]. Prague, NTK, 2011. [cit. 9. October 2014]. Available from: <http://www.nusl.cz/ntk/nusl-97123>.

PLANKOVÁ, Jindra. Obecné souvislosti Open Access - zamyšlení nad aktuálním tématem. *Ikaros* [online]. Praha: Ikaros, 2007, vol. 11, iss. 10 [cit. 9. October 2014]. ISSN 1212-5075. Available from: <http://www.ikaros.cz/node/4340>.

VYČÍTALOVÁ, Hana. Elektronické vědeckovýzkumné a technické zprávy zpřístupňované na českých webových serverech. *Ikaros* [online]. Praha: Ikaros, 2007, vol. 11, iss. 10 [cit. 9. října 2014]. ISSN 1212-5075. Available from: <http://www.ikaros.cz/node/4341>.

VYČÍTALOVÁ, Hana. Osobní archiv Ing. arch. Jana Moučky v Národním úložišti šedé literatury (NUŠL). *Knihovna: knihovnická revue*. Praha: Národní knihovna České republiky, 2011, vol. 22, iss. 2, p. 48-58. ISSN 1801-3252.

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

Grey literature in the National Medical Library

Michal Záviška, Lenka Maixnerová

zaviska@nlk.cz

National Medical Library, Czech Republic

Abstract

National Medical Library (NML) has been engaged in the grey literature since its establishment. Types of grey literature represented in the collections of the NML. Methods of collecting, processing, storing, or digitizing and services from the funds of grey literature in the NML. Grey literature intended for the general public (patients, their family members, etc.) versus grey literature for professionals. Continuity of other databases of the NML, the processing of grey literature and cooperation with the National Repository of Grey Literature (NRGL). New options to storage of full-text scientific works by authors themselves - self-archiving.

Keywords

Grey Literature, Libraries, Self-archiving, Digital Repositories

The National Medical Library of the Czech Republic (NML) is the central national departmental library. Legally, it is a state organization managed by the Ministry of Health. The origins of the NML date back to the establishment of the *Prager medizinisches Lesemuseum* in 1841. The library has been housed in the Medical House in Sokolská street, Prague since 1931. Part of its operations are still located there to this day. The Medical Museum (founded in 1934) is also part of the NML. The *National Medical Library* has had its current name since 1992. Today, the NML has over 102,000 titles and over 416,000 library items with various kinds of documents in its collections stored in repositories located at six different addresses. The entire collection has been catalogued using the Davinci/Medvik system. The NML has used the Medical Subject Headings (MESH) thesaurus for collection indexing since 1977. In 2008, the NML began building a digital library, now in the Kramerius version 4.

The NML began collecting grey literature for its collections from the time it was established. Initially, this mainly involved university graduation work - dissertations. Currently, the NML primarily acquires conference materials and materials from various professional societies and patient associations or even pharmaceutical and healthcare companies for its collections. An important part of the collection also includes the reports from the past projects of the Internal Grant Agency at the Ministry of Health (the IGA). We are currently expecting an increase in preprint and post-print documents, as well as presentations, posters, etc. especially in their native digital form in the digital library in connection with the launch of our self-archiving service.

Figure 1: Cover of the Final Report of the IGA project

Figure 2: Cover abstracts of conference proceedings

Examples of the frequency of grey literature in the NML collections	
Final grant reports	4669
University theses and dissertations (mostly historical collection)	822
Guidelines	981
Annual reports	185 titles
Directories	140
Certification works	276
Brochures and information leaflets for patients	502
Leaflets	438
Accreditation works - degree programs accreditation management	ca 30
Conference proceedings	5781

Note: The indication of the publication type in the NML records has only been mandatory since 2006. For this reason, the occurrence of the individual types is best shown by the documents which have been processed during the last ten years.

The NML currently focuses primarily on Bohemian documents when obtaining grey literature. The NML acquires grey literature for its collections in a number of ways, but mostly as gifts. Since 1993, it has systematically and comprehensively collected all IGA final grant reports, but this cannot be said of the other types of grey literature. University theses and academic works are only collected marginally and purely at random. On the other hand, the NML tries to systematically and comprehensively collect all conference materials of Czech origin. Conferences organized mainly by professional medical associations, universities and other organizers of conferences and professional meetings are used as sources. Recently, an increasing number of conference proceedings have been published only electronically. These publications are included in our digital library. Other types of grey literature in the NML collections (which we try to acquire comprehensively) include documents intended for the general public, whether they are patients, their families, friends or any other laymen. We obtain this type of literature from professional societies, but especially from patient organizations.

Grey literature is treated at the NML in the same way as normally published literature. Each document is nominally catalogued according to AACR2. The authors are linked to the personal authorities or personal authority records are created for them first. Corporations and events are also linked to authority records. Factual indexing is done on four levels. Basic indexing is done using Conspectus. The next level involves the classification of the subjects according to the field at the NML and UDC. The last and most detailed level involves the allocation of descriptors from the international thesaurus MESH. The log also contains documents publication type indexing. After that, the documents are filed in the standard collection. With

regard to the document format, we distinguish between the conventional book format and so-called "small print". This mainly involves single-page publications or small brochures which may have suffered damage when placed on a warehouse shelf. If this involves collections of specialized articles or abstracts, they are also included in the analytical process for the Bibliographia medica Čechoslovaca (BMČ).

Bibliographic records are sent each month to the Union Catalogue of the Czech Republic. Records of the final grant reports are also sent annually to the National Repository of Grey Literature - NRGL.

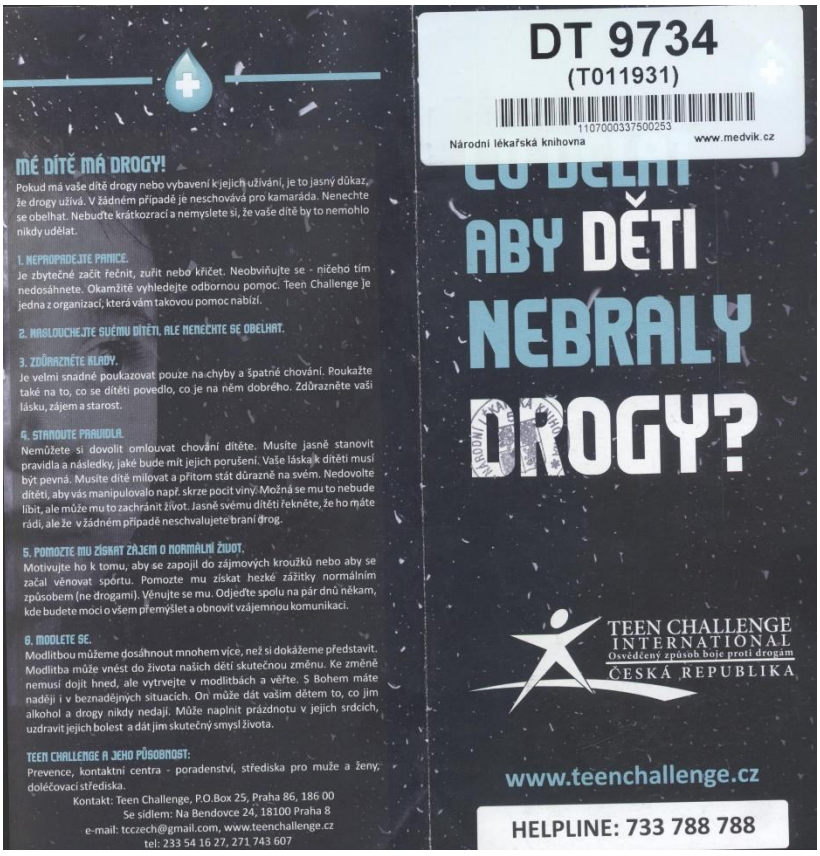


Figure 3: Support company flyer

Figure 4: Cover for brochure with information for patients from the endowment fund

Figure 5: Ministry's information brochure

Figure 6: Information booklet for patients professional society

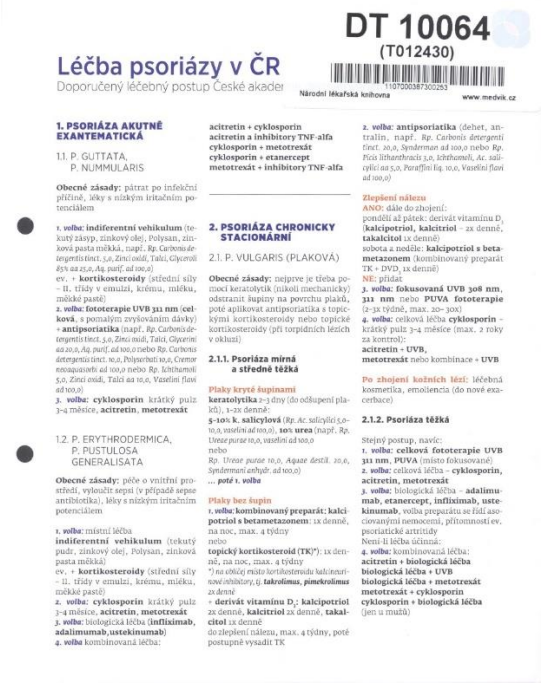


Figure 7: Recommended treatment procedure

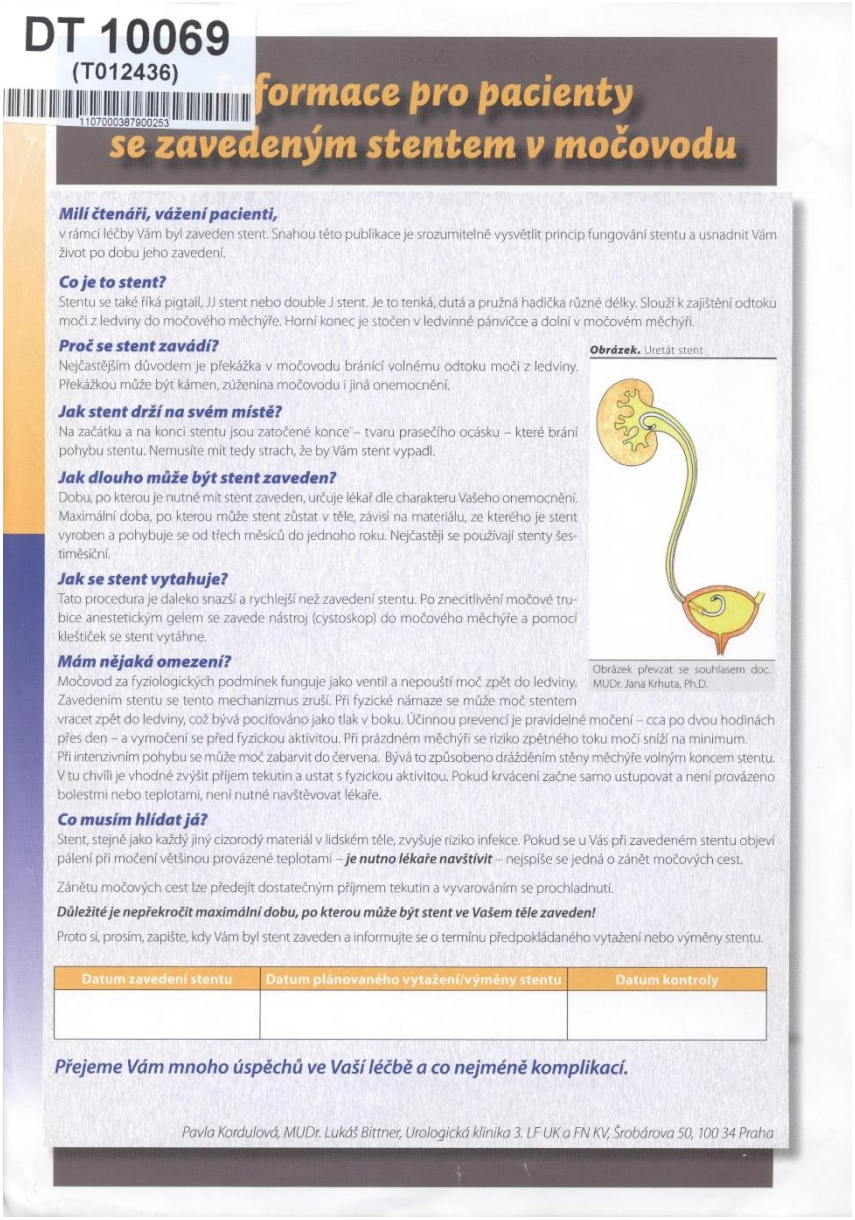


Figure 8: Information leaflets dispensed medical facilities

The definition of grey literature provided by TDKIV *(Documents which are not published in the usual manner and are therefore not available in the current book market (i.e. theses and dissertations, research reports, internal documents, official publications, etc.).)* and the description of grey literature on Wikipedia *(Grey literature is a term for documents that are not formally published. They do not undergo the standard editorial process. They include, for example, reports (annual reports, research, technical reports, findings, reports on the state), university theses (bachelor's, master's and doctoral theses), labor and materials, conference materials (proceedings from conferences, presentations, posters, programs) and corporate literature. They can be available to a broad range of people, yet they are often difficult to find.)* do not directly mention whether grey literature should be strictly professional or if it can also be non-professional. In the experience of the NML, the importance of grey literature designed specifically for the general public is readily apparent. Such documents are generally intended for patients, their family members and other people who come across them in everyday life, for

example at school or work. This often merely involves small prints or brochures, but with some unique content. These types of grey literature are mostly produced by various patient organizations, professional medical societies or health insurance companies and most are available in doctors' offices or at healthcare facilities in general. The equal significance of both professional and non-professional grey literature in the collections of the NML can be documented using the recent borrowing statistics for individual titles. There is no significant difference in the percentage of borrowing for both individual titles and individual units for titles acquired after 1st January 2012 in the statistics for last year. In both cases, borrowing stands between 20% and 25%.

The latest way in which the NML, respectively its digital library can acquire grey literature is using the self-archiving facility. In 2013, the NML developed a web application allowing the storage of full-text scientific works by their authors using the Medvik portal in the NML Digital Library. The self-archiving facility is available in its test mode at http://www.medvik.cz/. Once it goes live, registered NML users will be able to use the self-archiving facility. The service will be provided free of charge. The author selects the copyright license arrangement (Creative Commons), which determines how other users may handle the document, during the self-archiving process. Currently, it is only possible to archive documents that have a bibliographical record in the BMČ. In the near future, however, there will also be an archiving option to upload works not included in the BMČ. The pilot operation of the self-archiving service this year has demonstrated the applicability of the basic types of published works, mainly journal articles and contributions to anthologies. It is necessary to incorporate other document types in order to ensure its wider use by the professional community. In short, self-archiving in institutional or disciplinary repositories is an emerging trend in the world of scientific literature publishing.

A registered user with a national authority identifier stored in their account can log onto their library account via the Medvik portal. There, they can easily browse through publication details with which they are linked as the author or co-author. The account environment easily shows the authority record, the list of the documents linked to the particular authority record, as well as the list of the documents already recorded and the full text of the documents recorded by the co-authors.

If authors wish to archive documents, they choose a record from the list of available documents whose full text they want to archive and select the archive button. The documents can be uploaded in the PDF format. If the authors are not sure whether the document can be archived and made accessible, they can use the wizard (manual).

Figure 9: Article archiving

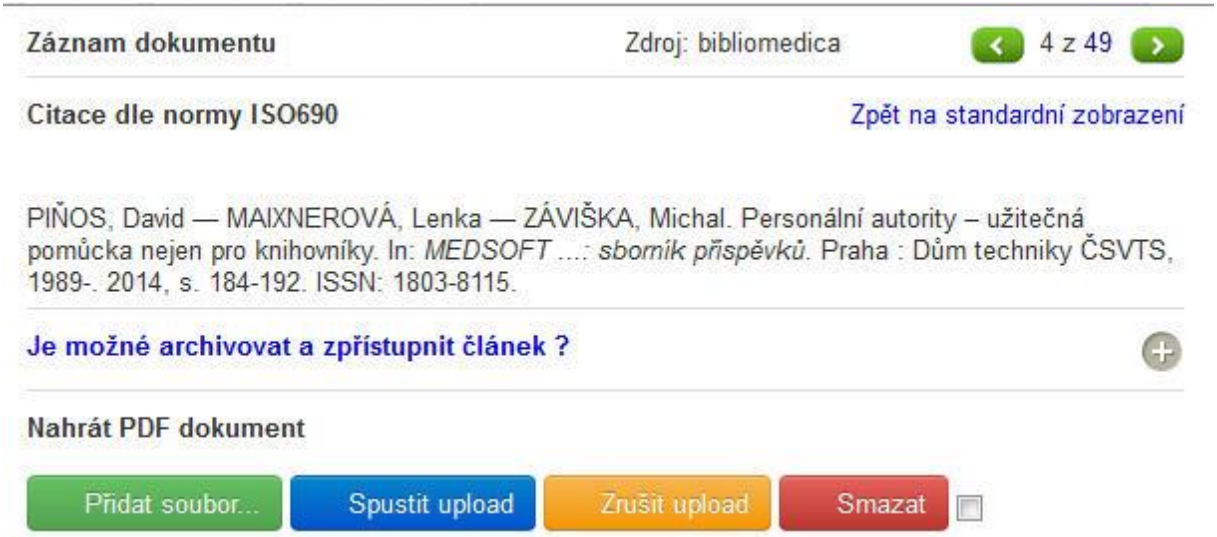


Figure 10: Document upload

The NML self-archiving service is intended to operate in accordance with the Czech Republic's current legislation. License, subsection 3. Special provisions for the publisher's license agreement:

"(1) By means of the publishing license contract, the author provides the licensee a license to reproduce and distribute copyright works of literary, musical and dramatic or musical, visual or photographic, or expressed in a manner similar to that of photography, unless the copyright owner in the implementation of the performers.

(2) Unless the contract concluded in writing expressly agrees upon a non-exclusive license, it shall be regarded as an exclusive license; this does not apply in case of the reproduction and distribution of copyright works in a periodical publication. "[cit. 1]

It can be assumed from the aforementioned that, if there is no agreement in writing, the author does not need to worry about archiving articles and making them available via the Medvik

portal, but in the case of the chapters of monographs and monographs themselves the author is not entitled to this, if the rights have not been treated differently in the contract with the publisher. In those cases where the publisher/editor has gone out of business, all the rights are transferred to the author. The right to archive and access a publisher's PDF (in the form of articles which have appeared in a magazine) is still not clear. It is especially necessary to verify this information for foreign titles, i.e. using the SHERPA/RoMEO service. "SHERPA / RoMEO provides information on access to publishers' (auto) archiving of electronic preprints and post-prints in a digital repository. The database now includes over 1,000 publishers (and nearly 20,000 journals), primarily Anglo-American in origin. It distinguishes between the individual publishers' approach to self-archiving using the following color distinction:

• Green – it is possible to archive a preprint and post-print or a publisher's PDF

• Blue – it is possible to archive a post-print or publisher's PDF

• Yellow – it is possible to archive a preprint

• White - archiving not formally supported "[cit. 2]

A manual has been created in order to facilitate the determination of whether the author may or may not archive the document. This manual shows the ways or archiving and publishing the articles and it makes use of the information from the SHERPA/RoMEO register, as well as references to relevant magazine websites (as long as these references are available in the Medvik catalogue). During the course of the archiving process, the author selects the document type: preprint, post-print, publisher's version or poster and selects the public license type preferred in the Medvik portal. One bibliographic record allows the uploading of several document types, i.e. abstracts, full-text lectures and presentations. The author can see a list of the uploaded documents, including the list of documents archived by co-authors, in the account.

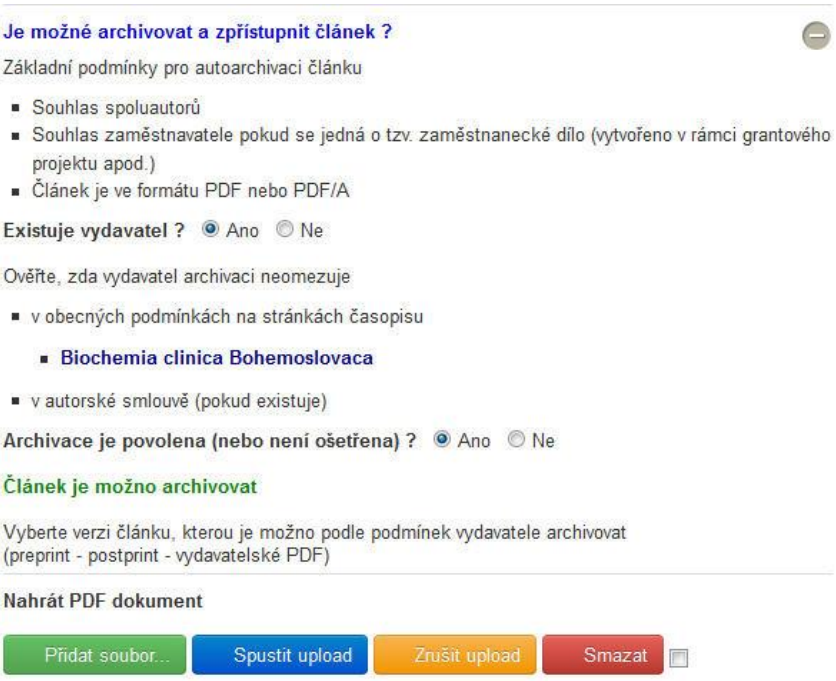


Figure 11: Archiving Wizard (manual)

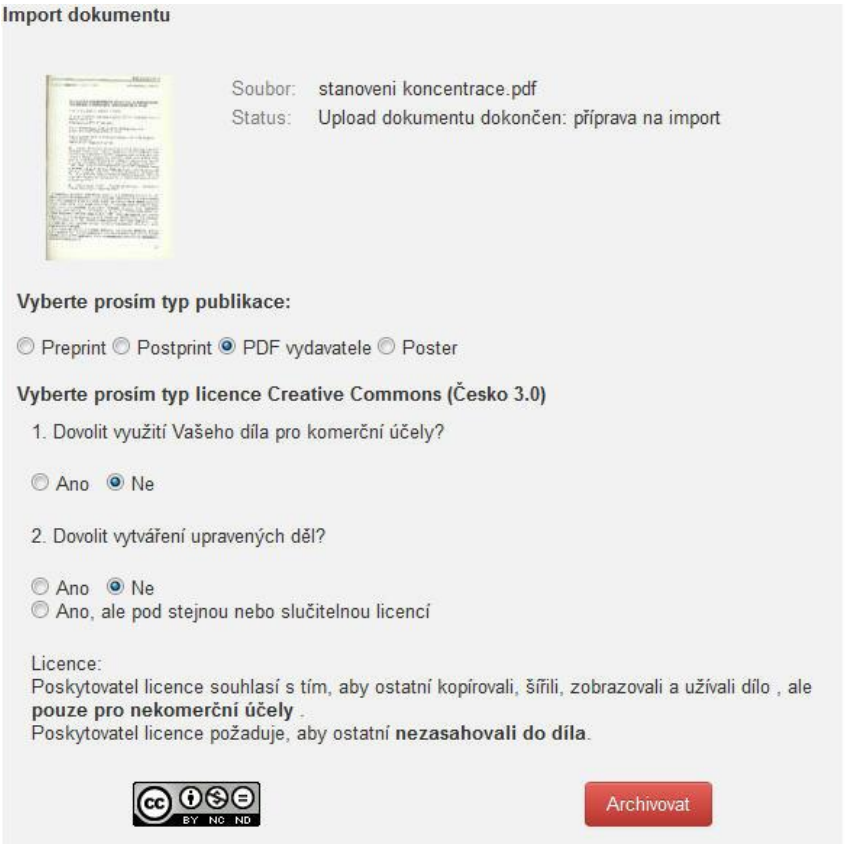


Figure 12: Document import

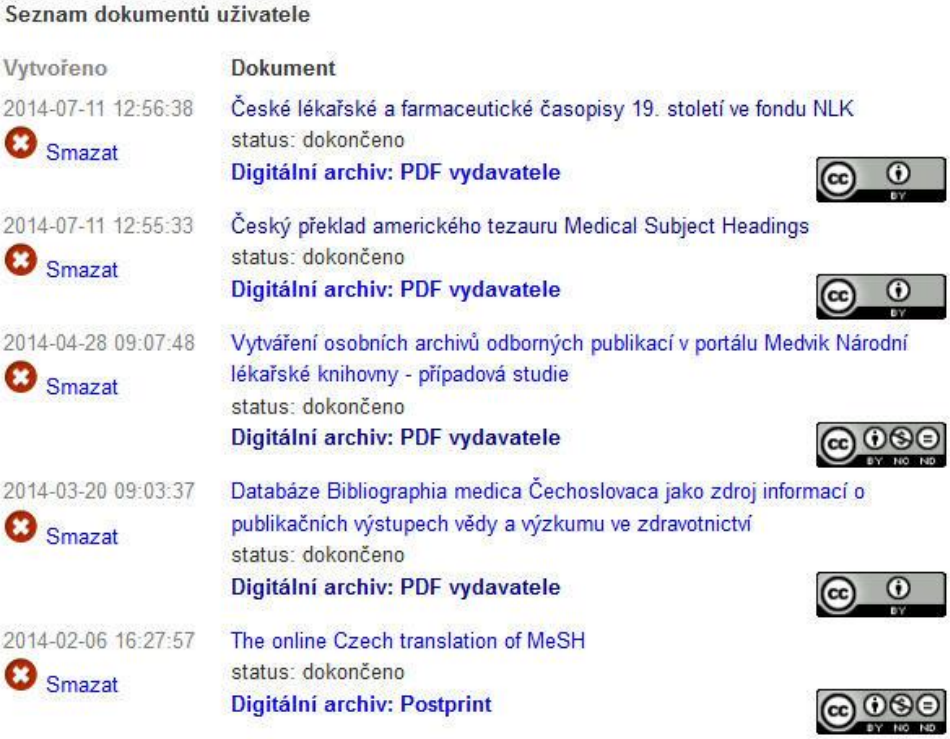


Figure 13: The list of archived documents

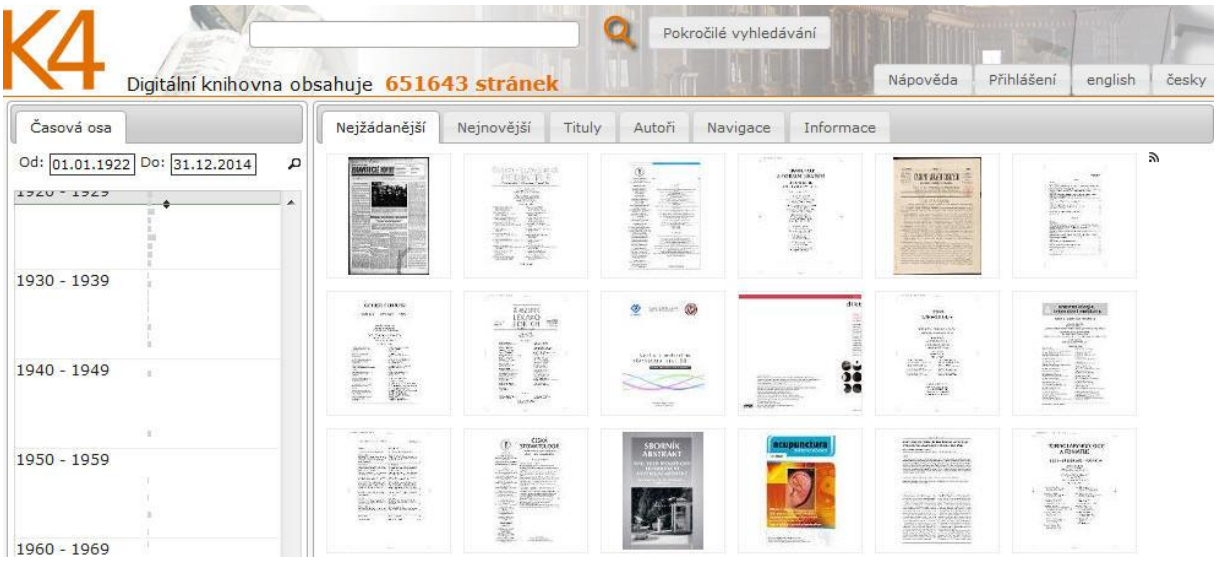


Figure 14: Kramerius version 4

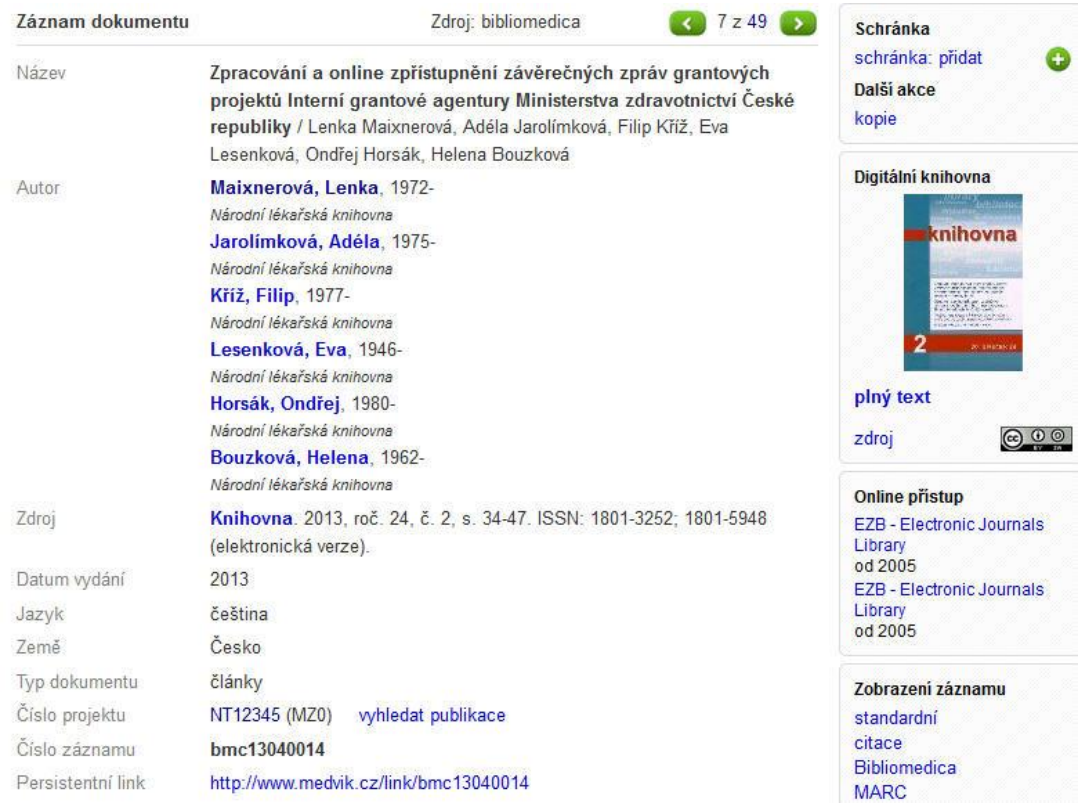


Figure 15: Detailed view of the bibliographic record of archived full text

One of the sources of grey literature in the NML collections is its own grey literature. Currently, the NML is preparing an internal directive on the retention of its own staff's publishing activities. This directive will apply to items such as posters, the preservation of which has been problematic in the past. According to this directive, all publishing activities will be archived in

the NML Digital Library and a standard bibliographic record will be created for each individual document. Older publications, including various project reports, have also been re-digitized.

In conclusion, we can say that grey literature has long been an integral and essential part of the collections and services at the NML, regardless of the orientation of the document towards the professional or lay public. The auto-archiving facility at the NML Digital Library constitutes a promising way to gathering, recording and disclosing grey literature. At the same time, the NML Digital Library is also the most appropriate method of recording and storing its own employees' publications.

References

Česko. Zákon č. 89 ze dne 3. února 2012: Občanský zákoník. In: Sb.zákonů č.89/2012. 2012.

BUREŠOVÁ, I. Sherpa/RoMEO, institucionální repozitář AV ČR a otevřený přístup (open access). *Informace* [online]. 2012, č. 1 [cit. 2014-02-19]. Available from: <http://www.lib.cas.cz/casopis-informace/sherparomeo-institucionalni-repozitar-av-cr-a-otevreny-pristup-open-access/>.

MAIXNEROVÁ, L. et al: Vytváření osobních archivů odborných publikací: případová studie. In: *MEDSOFT 2014*. Praha: Creativ Connections, 2014, s. 154-164. ISSN 1803-8115.

Národní lékařská knihovna. [online]. [cit. 2014-10-02]. Available from: <http://www.nlk.cz/>.

Wikipedia [online]. [cit. 2014-10-07]. Available from: <https://cs.wikipedia.org/>.

TDKIV [online]. [cit. 2014-10-07]. Available from: <http://www.old.nkp.cz/o_knihovnach/Slovnik/index.htm>.

Medvik [online]. [cit. 2014-11-10]. Available from: <http://www.medvik.cz/bmc/>.

Tezaurus Medical Subject Headings. [online]. [cit. 2014-11-10]. Available from: <http://www.nlk.cz/informace-o-nlk/odborne-cinnosti/tezaurus-medical-subject-headings>.

Digitální knihovna NLK [online]. [cit. 2014-11-10]. Available from: <http://kramerius.medvik.cz/search/>.

Bibliographia medica Čechoslovaca [online]. [cit. 2014-11-10]. Available from: <http://www.nlk.cz/informace-o-nlk/odborne-cinnosti/bmc>.

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

The GreyGuide Repository and Web-Access Portal: GreyNet's Response to the Pisa Declaration

Dominic Farace, GreyNet International

info@greynet.org

Grey Literature Network Service

Abstract

In December 2013, the GreyGuide was launched and in early 2014 GreyNet welcomed far reaching developments in its resource management. These developments are in line with the "Pisa Declaration on Policy Development for Grey Literature Resources" published on May 16, 2014. In compliance with this 15-point roadmap, a selection of GreyNet's web-based content is currently in the process of migrating to the GreyGuide, thus allowing for seamless browse, search, and retrieval across its collections.

Keywords:

GreyGuide, Pisa Declaration, Policy Development, Grey Literature, Open Access

Introduction

I'd like to ask you a question. How many towers are there on this earth that do not have the fame of Pisa? Its leaning and scientists and technical engineers have for centuries sought solutions to shore-up and preserve this world heritage. Scientists and technicians since the mid-20th century have also been engaged with yet another phenomenon "how on earth to deal with grey literature"[3]? In April of this year, information professionals from Europe's finest academic and research centers met in Pisa to address issues dealing with the policy and management of grey literature resources[i]. This meeting resulted in the formulation of the Pisa declaration on grey literature – a 15 point roadmap[ii], which serves as a guide for organizations involved in the production, publication, access and use of grey literature well into this 21st century.

Such an initiative is truly innovative – for it has found a new way to deal with a growing wealth of knowledge across disciplines and fields of industry and government that impact our global information society. What was not possible a couple of decades ago when the First International Conference on Grey Literature met in Amsterdam, is today manifest.

The Pisa Declaration much like the Budapest[iii], Bethesda[iv], and Berlin[v] Declarations are neither inventions of the mind nor improvements in doing the same things better. Instead, they offer a different approach in dealing with information and data to meet the market needs and requirements of today.

Policy Development for Grey Literature

Until now, the problem was the lack of cooperation and coordination between and among organizations dealing with grey literature. However, the time of going it alone is now ended. The Pisa Declaration marks the close of an era of *ad hoc* policy and decision making with regard to grey literature resources. In fact, the term Declaration says it – We call for, We endorse, We the international grey literature community, the authors and researchers, the information practitioners and professionals working now for years in this field of library and information science.

If I may take a few moments to recap the main points set out in the Pisa Declaration by addressing them in five categories:

First, the organizational commitment to open access, to further cooperation and coordination between and among grey literature communities, sharing open data standards.

Second, the commitment to research and education, where recognition and reward is associated with quality grey literature, and where attention is given to good practices in the field.

Third, the commitment to address and safeguard legal issues inherent to grey literature by exploring the various types of licensing agreements now available and by fostering constructive relations with commercial publishers.

Fourth, a commitment to sustainability linked to a financial prerequisite. Identifying funding and grants for special collections and repositories, commitment to long term preservation, and investments in new technologies.

[3] "Grey Literature is a field in library and Information science that deals with the production, distribution, and access to multiple document types produced on all levels of government, academics, business, and organization in electronic and print formats not controlled by commercial publishing i.e. where publishing is not the primary activity of the producing body." Luxembourg, 1997 - New York, 2004 – Prague, 2010

And Fifth, a firm technical commitment, a guarantee to continued online services and further crosslinking of textual and non-textual content. A commitment, which ranges from tackling broken links to facilitating interoperability - regardless of the system or portal in which grey literature and its accompanying data are housed.

GreyNet Compliance

It is in this way that the Pisa Declaration can revel in the strengths and opportunities that grey literature offers, while at the same time exposing the weaknesses and threats facing our community. No longer are we resigned that grey literature is hard to find, but instead how can we best search and access it. No longer hold in question its worth and value, but instead set out the review process it has undergone. And, no longer hesitate as to whether it is published or not, but instead cite and reference grey literature – make it openly public – isn't that what published means?

Colleagues, I could not stand before you here today as a representative of GreyNet – one of the already more than 100 signatories[vi] to this document - if our own organization were not prepared to implement the points set out in the Pisa Declaration. In the remainder of this paper, I would like to focus on the organizational and technical commitments GreyNet has laid out for itself.

Regarding GreyNet's organizational commitment, a committee established earlier this year seeks to further coordinate GreyNet's collections in the OpenGrey Repository[vii], the DANS Data Archive[viii], and the GreyGuide Repository[ix]. And, as to its technical commitment - *certainly the most ambitious at the moment* - GreyNet has reached agreement with the National Research Center in Pisa, the service provider for the GreyGuide Repository. GreyNet collections and content that are as yet only accessible either via its website[x] or allied conference site[xi] will now migrate to the GreyGuide Repository. In this way, the GreyGuide will function as GreyNet's web access portal. This will allow for combined search, browse, and retrieval capability - where standardized metadata and full-text can be online harvested.

GreyNet Collections via GreyGuide alongside Good Practices

Alongside the collections of published[xii] and proposed[xiii] good practices in grey literature, five of GreyNet's in-house collections have been identified and are in the process of migrating to the GreyGuide Repository. These include:

BIO Collection (Who is in Grey Literature)

WHOIS[xiv] in the field of Grey Literature is a compilation of over 200 biographical records provided by authors in the International Conference Series on Grey Literature. This online resource is maintained by TextRelease, the GL Program and Conference Bureau. Records in this directory appear in alphabetical order by last name of author and each record contains a current email address. In early October 2014, the metadata record format[xv] used in the migration of content to the GreyGuide became operational and the authors in the GL-Series are requested to complete their own online record entry.

GL16 Collection (Conference Proposals)

Participants who seek to present a paper at GL16 were invited to submit an English abstract between 350-400 words. The abstract is required to deal with the problem/goal, the research method/procedure, an indication of costs related to the project, as well as the anticipated results

of the research. The abstract should likewise include the title of the proposed paper, theme(s) most suited to the paper, name(s) of the author(s), and full address information. Abstracts are the only tangible source that allows the Program Committee to guarantee the content and balance in the conference program. In previous conferences in the GL-Series these abstracts were accessible on the conference site, however this content is now deposited in a standardized metadata format[xvi], where it will remain housed in the GreyGuide.

GSI Collection (GreySource Index)

The GreySource Index[xvii] comprises a collection of over 70 web based resources in grey literature. This index was originally intended to expose grey literature to the average net-user and in so doing profile organizations responsible for their production and/or processing. Only web-based resources that explicitly refer to the term grey literature (or its equivalent in any language) are listed. GreySource identifies the hyperlink directly embedded in a resource. The migration of content for this collection in the GreyGuide began in the 3rd quarter of 2014 with the first release of its standardized metadata format[xviii].

GFS Collection (GreyForum Series)

The GreyForum[xix] is a series of onsite and online courses, seminars, and workshops where grey literature provides common ground for information professionals in the process of knowledge transfer. The migration of content to the GreyGuide is scheduled to commence in the 1st quarter of 2015.

IDGL Collection (International Directory of Organizations in Grey Literature)

The International Directory of Organizations in Grey Literature[xx] provides a list of over 200 organizations in 36 countries worldwide that are currently associated with GreyNet either via partnership, membership, sponsorship, or authorship in the field of grey literature. Entries are alphabetical by country and each entry has an embedded link to the corresponding organization's website. GreyNet International is proud to serve a global grey literature community and welcomes additions and revisions to this Directory. The migration of content to the GreyGuide is likewise scheduled to commence in the 1st quarter of 2015.

In Closing

This new approach in channeling GreyNet's diverse web-based content will no doubt better meet the needs and requirements of today's international grey literature community. A community just like other knowledge based communities, where users choose to rely ever more on one point of entry for relevant information and data.

Colleagues, you can provide a significant contribution to this initiative in two ways: first of all, by contributing content to one or more of the collections in the GreyGuide Repository; and secondly by endorsing the Pisa Declaration scheduled for ratification during the Sixteenth International Conference on Grey Literature[xxi] this December at the Library of Congress. Thank you for your attention, hopeful for your endorsement[xxii].

Web References

i. <http://www.greynet.org/greyforumseries/policydevelopment.html>

ii. <http://greyguide.isti.cnr.it//include/pisadeclarationmay2014.pdf>

iii. <http://www.budapestopenaccessinitiative.org/>

iv. <http://en.wikipedia.org/wiki/Bethesda_Statement_on_Open_Access_Publishing

v. <http://openaccess.mpg.de/Berlin-Declaration>

vi. <http://greyguide.isti.cnr.it//pisadecla/listaiscritti.php>

vii. <http://www.opengrey.eu/search/request?q=greynet>

viii. <https://easy.dans.knaw.nl/ui/?wicket:bookmarkablePage=:nl.knaw.dans.easy.web.search.pages.PublicSearchResultPage&q=greynet>

ix. <http://greyguide.isti.cnr.it/>

x. <http://www.greynet.org/>

xi. <http://www.textrelease.com/>

xii. <http://greyguide.isti.cnr.it/listtitoli.php?authority=GreyGuide&collection=Published&langver=en&RighePag=100>

xiii. <http://greyguide.isti.cnr.it/listtitoli.php?authority=GreyGuide&collection=Proposed&langver=en&RighePag=100>

xiv. <http://www.textrelease.com/whois2014.html>

xv. <http://greyguide.isti.cnr.it/linkdoc.php?idcode=2014-B01-001&authority=WhoIs&collection=BIO&langver=en>

xvi. <http://greyguide.isti.cnr.it/listtitoli.php?authority=GLConference&collection=GL16&langver=en&RighePag=100>

xvii. <http://www.greynet.org/greysourceindex.html>

xviii. <http://greyguide.isti.cnr.it/listtitoli.php?authority=GreySource&collection=GSI&langver=en&RighePag=100>

xix. <http://www.greynet.org/greyforumseries.html>

xx. <http://www.greynet.org/internationaldirectory.html>

xxi. <http://www.textrelease.com/gl16conference.html>

xxii. <http://greyguide.isti.cnr.it//pisadecla/iscrivi.php>

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

Providing access to research data, publications and current research information at Data Archiving and Networked Services - DANS

Elly Dijk, Peter Doorn

elly.dijk@dans.knaw.nl, peter.doorn@dans.knaw.nl

Data Archiving and Networked Services – DANS, the Netherlands

Abstract

Data Archiving and Networked Services (DANS) promotes sustained access to digital research data in the Netherlands. Researchers can deposit their data through the online archiving system EASY. Via the portal NARCIS the research data are shown in context, namely in relation to e-publications, and other research information. Both EASY and NARCIS contain grey literature like archaeological reports, data documentation, doctoral thesis, conference papers, patents, and technical documentation of universities and other research institutes.

Keywords

Research Data, Archiving Systems, E-publications, Research Projects, Research Institutions, Researchers, Enhanced Publications, NARCIS, EASY, Dutch Dataverse Network

Introduction

The mission of the Dutch institute Data Archiving and Networked Services (DANS)[4] is to promote sustainable access to digital research data (such as databases, spreadsheets, text, images, audio, video, multimedia), research information (information on research, research organisations and researchers) and electronic publications (including preprints, reports).

To carry out its mission, DANS offers services for long-term archiving and re-use of completed research data, mainly in the humanities and the social sciences, and it supports data management during ongoing research (research data management). Additionally, DANS offers training and consultancy.

DANS supports its services with research into sustainable access to digital information. Research (Research & Innovation) is a derivative of the main objective and is focused on innovation and improving service to users.

DANS is an institute of the Royal Netherlands Academy of Arts and Sciences (KNAW)[5] and the Netherlands Organisation for Scientific Research (NWO)[6] since 2005, but the first predecessors of DANS date back to 1964, the Steinmetz Foundation for the social sciences, and to 1989, the Historical Data Archive.

The emergence of institutes for data archiving

In the sixties of the 20th century we saw the emergence of data archives in the social sciences, like the Dutch Steinmetz Archive and the British UKDA. Later followed by text archives for linguistics and literary studies in the seventies (e.g. Oxford Text Archive), historical archives in the eighties and nineties (e.g. NHDA, HDS, IPUMS), and archaeological data archives at the beginning of this century (e.g. ADS, EDNA). About ten years ago the universities started to develop university repositories for publications and research data. General data sharing facilities, such as Zenodo, Figshare and the EUDAT B2-tools are even more recent.

In the past few years there is growing attention for the value of (big) data, data sharing and proper research data management. Data sharing has become an important issue among research funders, policy makers, in the scholarly world itself, among publishers, and even for the general public. In 2010 the influential EU report, "Riding the Wave"[7] about the future of scientific data was published. In 2011 Neelie Kroes, Vice-President of the European Commission responsible for the Digital Agenda said that "Data is the new gold"[8]. In the EU programme Horizon 2020 one of the goals is "open research data". This has a positive effect on the willingness of researchers and on their institutes to share research data.

DANS offers services for data curation, archiving and data management in the social sciences and humanities. The institute has been concerned with the awareness of the value of preserving data for re-use, so one can use the data for validating the results of earlier research. Also it is possible to use the data for comparative analysis, and researchers can use the data for secondary analysis, that is to say answering new research questions with existing data.

[4] http://www.dans.knaw.nl/en
[5] http://www.knaw.nl/en
[6] http://www.nwo.nl/en
[7] Riding the wave: How Europe can gain from the rising tide of scientific data. Final report of the High Level Expert Group on Scientific Data. A submission to the European Commission, October 2010, 40 p. See: http://cordis.europa.eu/fp7/ict/e-infrastructure/docs/hlg-sdi-report.pdf
[8] http://europa.eu/rapid/press-release_SPEECH-11-872_en.htm?locale=en

Federated data infrastructure in the Netherlands

On the initiative of Research Data Netherlands (RDNL)[9], a partnership between DANS, 3TU.Datacentrum and SURFsara, a federated data infrastructure is being developed, with services, roles and responsibilities for various parties in the Netherlands. The "Collaborative Data Infrastructure", a framework for a future data infrastructure from EU report "Riding the Wave", formed the inspiration for this model. RDNL elaborates this framework in four instead of three layers: the layer Common Data Services from the 'Riding the Wave' report is split into a technical basic infrastructure and back-office data services (see Fig.1).

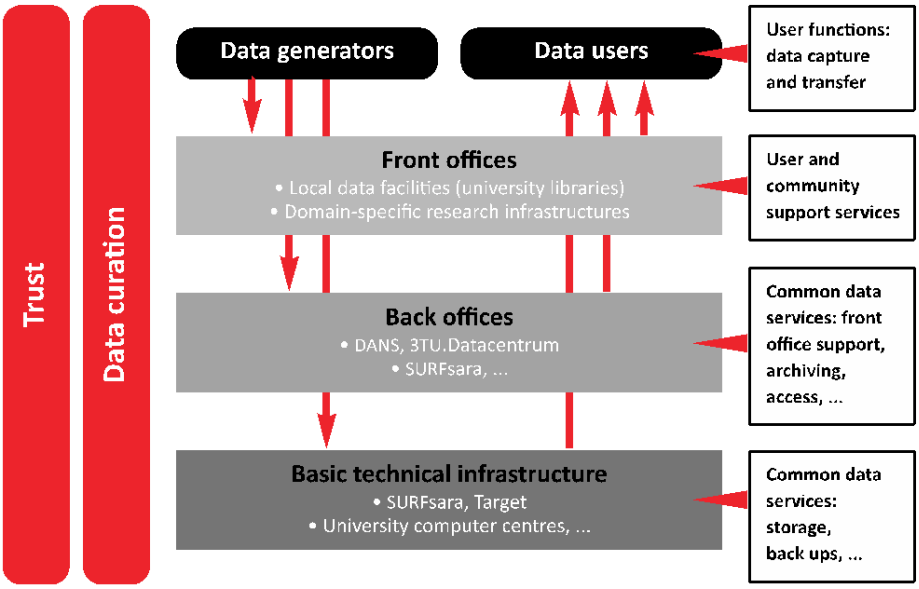


Fig. 1. Federated data infrastructure in the Netherlands

The foundation of the federated model is formed by the technical basic infrastructure, that are the computer centres which take care of data storage and backup facilities. Above this is a layer of back-office data services, which provides facilities and support for the long-term archiving and accessibility of the research data in trusted digital repositories, like the DANS archiving system EASY. Another level up one can find the front-office services, which are performed by university libraries as the first contacts with producers and users of research data. The European Research infrastructures such as CLARIN[10], DARIAH[11] and CESSDA[12] also perform front-office functions in the domains they represent. The vertical bar "Trust" in the figure refers to the importance of integrity, transparency and/or certification at all levels. "Data curation", or the care that data needs to remain available for the long term covers all phases of the life cycle of research data and it is a task to which all levels of the model contribute[13].

[9] http://www.researchdata.nl/en/

[10] http://clarin.eu/

[11] http://www.dariah.eu/

[12] http://www.cessda.net/

[13] See: http://www.dans.knaw.nl/sites/default/files/file/DANS%20FO-BO-model%20UK%20V1%20DEF%20web.pdf
(An update will be published later this year)

The services of DANS

In order to support the research community with research data management DANS offers different services (more about these services in the underlying sections):

- EASY[14]: electronic archiving system for self-deposit of research data. Researchers and institutions in the social sciences and humanities can submit their data in EASY. A majority of the datasets are under open access licences available for reuse.
- Dutch Dataverse Network (DDN)[15]: researchers can use this service to store scholarly data during the research process in a safe and sustainable manner in an online environment.
- NARCIS[16]: provides access to scholarly information, including (open access) publications from the repositories of all the Dutch universities, the Academy, the national research council (NWO), and a number of research institutes as well as datasets from some data archives (e.g. DANS). NARCIS also contains descriptions of research projects, researchers, and research institutes.
- Data Seal of Approval (DSA)[17]: DANS has developed minimum requirements for (scholarly) data archives.

Electronic archiving system EASY

Contents

EASY's collection currently (October 2014) consists of 28,130 datasets. These datasets are distributed among the following disciplines (a dataset can belong to more than one discipline): humanities – 25,933 datasets; socio-cultural sciences – 3,623 datasets; social sciences – 2,554 datasets; behavioural sciences – 993 datasets; life sciences and medicine – 276 datasets; and geospatial sciences – 40 datasets.

The collection Humanities consists primarily of a historical collection and an archaeological collection. The historical collection contains a broad range of historical datasets, which relate to the Netherlands and the former Dutch colonies. The archaeological collection, that forms a large part of the Humanities collection, is also known as the e-Depot for Dutch Archaeology, which consists of archived excavation reports and related datasets of archaeological excavation and exploration projects. There is also a substantial oral history and qualitative social science collection, consisting of recorded and/or transcribed interviews. Oral history is the collection and study of historical information using sound recordings of interviews with people having personal knowledge of past events, like the Interview project Dutch Veterans.

The social and behavioural sciences data collection consists of a large number of social survey data relating to the Netherlands from the late 1950s to the present, including longitudinal surveys of the Netherlands Institute for Social Research (SCP). The collection also includes National Election Studies (*Nationale Kiezersonderzoeken*) and protected microdata files of Statistics Netherlands (CBS).

[14] See: https://easy.dans.knaw.nl/ui/home

[15] See: http://www.dataverse.nl

[16] http://www.narcis.nl

[17] http://www.datasealofapproval.org

The Geospatial Sciences collection mainly consists of files from the Dutch Land Registry Office (Kadaster).

Reuse of data

In 2013, about 30,000 datasets were downloaded. The reviews of the datasets show that users highly appreciate the reuse of data[18].

Besides the download of the datasets deposited in EASY, there are many national and international projects where research data sharing concerns collaborative work. DANS is involved for example in international European infrastructure projects like Holocaust studies (European Holocaust Research Infrastructure - EHRI[19)], ARIADNE (for integration of archaeological datasets)[20], Dendrochronology (Digital Collaboratory for Cultural Dendrochronology – DCCD)[21], Global inequality (Clio Infra)[22] and Dutch collaborations for instance on Dutch Census Data 1795-1971 (a cooperation of four institutes)[23] and Historical Shipping (bringing together shipping records)[24]. (See also chapter 4).

Grey literature in EASY

EASY not only contains datasets but also grey literature. These publications, around 14,000, are connected to the datasets. One can find in EASY archaeological reports (in which the excavation results are described and documented), publications belonging to research datasets, and data documentation.

Dutch Dataverse Network - DDN

In May 2014 DANS took over the hosting of the Dutch Dataverse Network (DDN) from Utrecht University, where it was installed in 2010. The Dataverse Network is a scholarly data storage and sharing platform developed by the IQSS at Harvard University and is used worldwide. Researchers and lecturers from universities and other academic institutions can store and share their data in DDN during the course of their research. It also makes possible persistent data citation, and it enables reproducible research. With this service DANS supports researchers with research data management (RDM). Since May the number of users was extended from 7 to 10 institutions. Most of the 13 Dutch universities are currently formulating and implementing RDM policies. We expect this will have a boosting effect on the use of Dataverse by Dutch scholars.

The scholarly portal NARCIS

Since 2011 the portal NARCIS is a service of DANS. DANS took over this portal with the idea that it is not only important to provide open access to data and publications, but that it is also necessary to connect them to research information. Research information can be described as information about researchers, research institutes and research projects. With this connection the research data and publications are shown in their context.

[18] http://datareviews.dans.knaw.nl/index.php
[19] http://www.ehri-project.eu/
[20] http://www.ariadne-infrastructure.eu/
[21] http://dendro.dans.knaw.nl/
[22] http://www.clio-infra.eu/
[23] http://www.volkstelling.nl/
[24] https://easy.dans.knaw.nl/ui/datasets/id/easy-dataset:33898

NARCIS can be seen as the national aggregator of the Current Research Information Systems and repositories of the Dutch science system. It harvests the metadata of 7 repositories with datasets, for example EASY, and 33 repositories with (open access) publications. Besides that NARCIS has a controlled database with information about (current) research, researchers, and research institutes. The information in NARCIS is connected via the Digital Author Identifier (DAI) that is given to researchers, authors and depositors of datasets. The DAI is a unique national identifier assigned to every author working at a Dutch university or research institute. (In the Netherlands there are plans to replace the DAI by the international author identifiers ORCID and/or ISNI). We aim to present the information in an increasingly integrated and interconnected way.

The metadata of the publications and datasets within NARCIS, such as title or author's name, are also available via other services[25], like Google Scholar[26], DART-Europe E-theses Portal[27], OpenAire[28], and WorldWideScience.org[29].

Contents

In October 2014 NARCIS contains over 900,000 publications, of which 41% are open access publications. It features nearly 175,000 datasets, over 51,000 researchers including 8,500 professors and associate professors, almost 60,000 research projects (25% current research) and 2,900 research institutes[30].

In addition NARCIS contains more than 1,800 enhanced publications. Dr. Leen Breure, researcher at DANS and Utrecht University, has made a typology of 80 types of enhanced publications. This was made in the context of the Xpos're, research project on scholarship and multimedia[31]. Enhanced publications contain additional materials that have been attached to the full text version, such as underlying research data, models, images, conference papers, ranking data, and comments. The enhanced publications contained in NARCIS originate from various projects and cover a wide range of research fields, like archaeology, humanities, and social sciences. The presentation of these enhanced publications is still a beta version. DANS has however plans to change the way enhanced publications will be displayed. We will also soon launch a Data Journal in which data depositors can publish data papers, which describe their datasets.

Grey literature in NARCIS

Over 300.000 (open access) publications in NARCIS (33%) are grey literature: reports, doctoral theses, conference papers, contributions to periodicals, book reviews, working papers, annotations, patents, lectures et cetera.

[25] See for the overview of these services: http://www.narcis.nl/dataprovider/Language/en
[26] http://scholar.google.com/
[27] http://www.dart-europe.eu/basic-search.php
[28] https://www.openaire.eu/
[29] http://worldwidescience.org/
[30] This number is including all the parts of the 13 main universities in the Netherlands.
[31] See: http://xposre.nl

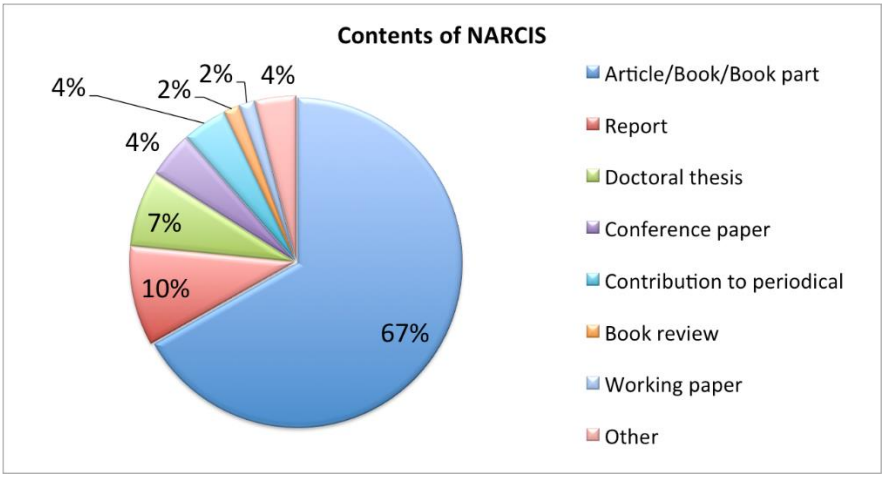


Figure 2: Contents of NARCIS, October 2014

It depends on the sort of publication if the publication is available as open access or not. For example 82% of the doctoral theses is open access, 83% of the patents, and 58% of the conference papers. On the other hand "only" 37% of the contributions to periodicals are openly available.

Data Seal of Approval (DSA)

For researchers, who want to share or want to use data, it is essential to know that the quality and integrity of the data is guaranteed. Since 2005 DANS has developed the so-called Data Seal of Approval (DSA)[32]; in 2009 DSA was handed over to an international board. DSA prescribes 16 minimum requirements concerning the creation, storage, use and reuse of digital data, that data archives have to meet. Data archives can perform a self-assessment through an online tool on the DSA website. These self-evaluations are reviewed by the DSA board, before the DSA seal label is issued. In October 2014 there are 36 repositories with DSA seal, including EASY. Furthermore, there are another 35 DSA seals requested.

DSA is part of the European Framework for Audit and Certification of Digital Repositories. The Data Seal of Approval offers a basic certification, the next level is extended certification DIN 3164433 and the third level is the formal certification according to ISO 1636334.

DANS in a national and international context

National context

Research Data Netherlands (RDNL) is a coalition between DANS, 3TU.Datacentrum and SURFsara. The mission of this cooperation is the promotion of sustained access and responsible re-use of digital research data. RDNL provides training for librarians (Essentials 4 Data Support training) and awards every two years the Dutch Data Award to researchers who gather and document data, and make it accessible to others.

With NARCIS, DANS cooperates with all the Dutch universities, the Academy, the National Research Council, the National Library, and a number of research institutes to make datasets,

[32] See for an overview of DSA: http://datasealofapproval.org/media/filer_public/2014/10/03/20141003_dsa_overview_defweb.pdf
[33] http://www.langzeitarchivierung.de/Subsites/nestor/EN/nestor-Siegel/siegel_node.htm
[34] http://www.iso16363.org/

open access publications and an overview of research projects, researchers and research institutes accessible for the academic world, the media, policy makers et cetera.

In the Netherlands DANS contributes to the research infrastructures of the national roadmap CLARIN.NL[35], European Social Survey (ESS)[36], CLARIAH[37] and The Language Archive[38].

International context

The aim of DANS is to be a leading building block of the data infrastructure in Europe. In October 2014 DANS participates in data activities of 21 international and European research infrastructures, and is an active contributor to a substantial number of European research infrastructure proposals for Horizon 2020.

Examples of the international research infrastructures in which DANS is involved are CESSDA, CLARIN ERIC and DARIAH ERIC. European projects in which DANS takes part are Data without Boundaries, EHRI, DASISH, APARSEN, OpenAirePlus, NeDiMAH, EUDAT, ARIADNE, Europeana Cloud, 4C, EuroCRIS, and IMPACT-EV[39].

Furthermore, DANS is member of the Research Data Alliance (RDA)40 and organised the last RDA Plenary in September in Amsterdam. The goal of the RDA is that researchers around the world share and use research data without barriers. The Research Data Alliance wants to accelerate international data-driven innovation and discovery by facilitating research data sharing and exchange, use and re-use, standards harmonization, and discoverability.

References

Data Archiving and Networked Services - DANS. *A federated data infrastructure for the Netherlands: The front office - back office model*. Den Haag: Data Archiving and Networked Services, October 2013. 7 p.

High Level Expert Group on Scientific Data. R*iding the wave: How Europe can gain from the rising tide of scientific data*. Final report of the High Level Expert Group on Scientific Data. A submission to the European Commission. European Union: October 2010. 40 p.

DILLO, Ingrid and DE LEEUW, Lisa. Het Data Seal of Approval: keurmerk voor duurzame en betrouwbare databewaarplaatsen. In *Handboek Informatiewetenschap en in de IWA-database*. Alphen aan de Rijn: B+B Vakmedianet, 2014. 29 p. Translated in English: Data Seal of Approval: Certification for sustainable and trusted data repositories. The Hague: DANS, 2014. 20 p.

[35] http://www.clarin.nl/
[36] http://www.europeansocialsurvey.org/
[37] http://www.clariah.nl/
[38] https://tla.mpi.nl/
[39] See for an overview of projects: http://www.dans.knaw.nl/content/projecten (in Dutch)
[40] http://rd-alliance.org/

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

Enhanced publications in V4 countries

Edith Görögh, the University of Debrecen

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

Jozef Dzivak, the Slovak University of Technology, Chemical Library

Petra Pejšová, Hana Vyčítalová, the National Library of Technology

Abstract

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

Keywords

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

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

This paper contains five parts:

- An introduction to Open Access in scholarly communication, research data and projects
- The report from the survey of Hungarian universities
- The report from the Survey of Polish Scientific and Research-Development Units
- The report from the survey of Slovak universities and scholarly institutions
- The report from the survey of Czech research institutions and universities

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

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

Edith Görögh,

editg@lib.unideb.hu

The University of Debrecen

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

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

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

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

Open Access in scholarly communication

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

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

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

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

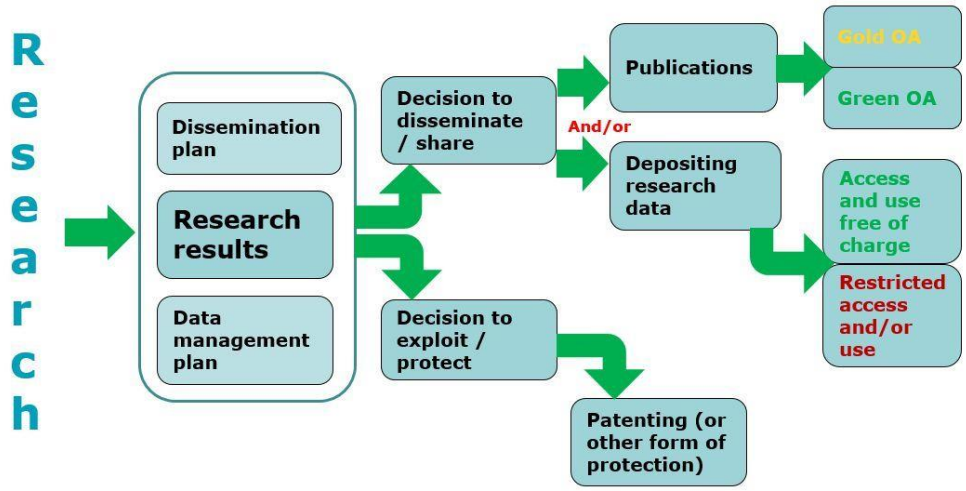


Figure1: Open Access to scientific publication and research data in the wider context of dissemination and exploitation [41]

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

Research data publication

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

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

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

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

The IVF project

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

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

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

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

The survey

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

The NTK initially prepared a survey in 2013 and tested it on research organizations in the Czech Republic. This survey subsequently served as the basis for the joint survey of all V4 institutions

in 2014. Due to the fact that the other three participating institutions did not have prior experience in open research data, we felt it would be wise to include some questions on Open Access publishing in general. This way, an attitude on Open Access in general could be estimated alongside the more detailed examination of data management practices.

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

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

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

The report from the survey of Hungarian universities

Edith Görögh,

editg@lib.unideb.hu

The University of Debrecen

Introduction

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

Survey description

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

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

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

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

Questions about Open Access publishing

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

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

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

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

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

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

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

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

Questions about research data management

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

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

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

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

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

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

Conclusion

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

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

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

Edyta Kędzierska, Natalia Kavalchuk, Jolanta Stepniak

E.Kedzierska@bg.pw.edu.pl, N.Kavalchuk@bg.pw.edu.pl, J.Stepniak@bg.pw.edu.pl

Warsaw University of Technology

Introduction

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

Survey description

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

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

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

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

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

Questions about Open Access publishing

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

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

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

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

Opinions about Open Access publishing

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

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

Questions about research data management

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

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

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

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

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

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

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

Conclusion

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

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

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

The report from the survey of Slovak universities and scholarly institutions

Jozef Dzivak

jozef.dzivak@stuba.sk

The Slovak University of Technology, the Chemical Library

Introduction

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

Survey description

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

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

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

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

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

Questions about Open Access publishing

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

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

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

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

Questions about research data management

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

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

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

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

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

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

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

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

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

Conclusion:

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

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

The report from the survey of Czech research institutions and universities

Petra Pejšová, Hana Vyčítalová

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

The National Library of Technology

Introduction

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

Survey description

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

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

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

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

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

Questions about Open Access publishing

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

Questions about research data management

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

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

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

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

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

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

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

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

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

Conclusion

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

References:

AYRIS, Paul. Open Research Data: Statement. *LERU* [online]. 2012 [cit. 2014-10-16]. Available from: <http://www.leru.org/files/general/Open%20Access%20to%20Research%20Data-FINALdocx.pdf>.

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

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

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

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

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

Appendix

Enhanced Publications Survey in V4 Countries

Questionnaire for research organizations and universities

Open access

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

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

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

1a. If it is obligatory – since when …..

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

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

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

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

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

- Lack of OA journals/repositories in the appropriate field
- Low prestige of OA journals/repositories
- Small influence of OA journals/repositories on the development of a scientific discipline

- Quality of the published papers (not peer-reviewed papers) is not proven
- Quality of the published papers in OA journals is not confirmed (lack of peer-view)
- Additional costs (required fees)
- Narrow readership for research papers published in OA journals
- Objections of traditional publishers
- Uneven copyright issue
- Additional work
- Other (please specify)…

Raw research data

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

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

Research data produced:

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

Accompanying materials

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

Post-published data materials:

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

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

- doc
- pdf
- csv
- xls
- xml

- JSON (JavaScript Object Notation)
- other (please specify)…

6. Where are the raw research data of your institution stored?
   - On the workstations of individual research workers
   - In a directory on a (common) server
   - In a central digital repository of the institution
   - In a centralized subject-based open archive
   - In other place (please specify)…

7. Does your institution plan any changes to the way of storing and archiving the research data?
   - Yes (please specify what kind of changes your institution plans to make)
   - No

8. Is the research data archived in your institution for more than 10 years?
   - Yes
   - No
   - Not yet, but we started……..years ago

9. Who in your institution is responsible (takes care) for storing and archiving the research data?
   - An appointed person(s) in the library
   - An appointed person(s) in the IT department
   - An appointed person – scientific secretary
   - Each research worker himslef
   - Other person(s) (please specify)…

Dissemination and re-use of the research data

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

10. Is the research data produced at your institution linked to the research papers?
    - Yes – go to question 10a
    - No – go to question 10b

10a. If yes, how is the research data linked to the research papers?

- By joint storing of data with publication in digital form on personal website of research worker
- By joint storing of the data with publication in a digital form in the repository
- By giving information on the data availability in the publication records (metadata)
- By referencing to place of data storing in digital form from the research publication
- In another way (please specify)….

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

- Yes
- No

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

- Yes
- No

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

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

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

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

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

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

- Yes
- No

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

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

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

- Yes
- No

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

- Yes
- No

Identification information

Name of your institution:

Contact e-mail:

Type of institution:

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

Field of specialization:

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

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

Privacy and anonymization in repositories of grey literature

Michal Koščík

michalkoscik@gmail.com

Masaryk University, the Faculty of Law, the Faculty of Medicine, the Czech Republic

Abstract

Recent case law from the Court of justice of the European Union, such as the case of Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González indicates that the repositories of grey literature may be qualified as institutions processing the personal data of the subjects mentioned in the documents stored in repositories. These repositories may face requests for the anonymization or even the removal of their documents. The purpose of this paper is to outline the legal framework and suggest procedures to approach this issue in compliance with the EU legislation.

Keywords

Privacy, Personal Data, The Right to be Forgotten

Privacy and anonymization in grey literature repositories

Most grey literature repositories process documents containing random information which directly or indirectly relates to living or deceased individuals. Even if these repositories are not operated for the purpose of collecting specific information on the individuals who are mentioned in the archived documents, the recent case-law of the Court of Justice of the European Union (hereafter simply referred to as the CJEU), such as Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González[42] (hereafter simply referred to as Google Spain) indicates that such repositories may be considered to be data controllers and have to comply with data protection regulation. The case-law also suggests that EU regulation could also apply to repositories and search indexes operated from outside the EU, most notably from the USA.

The case of Google Spain and its consequences for grey literature

The case of Google Spain involved the "google search" service, whose primary purpose is to index information contained in the websites. The purpose of the "google search" is not to collect the personal data of any individuals coincidentally contained in the indexed websites. Nevertheless the CJEU was asked to decide, whether the Google was or was not the controller of personal data under the EU regulations. The facts of the Google Spain case can be summarized as follows:

"In 2010, a Spanish citizen lodged a complaint against a Spanish newspaper with the national Data Protection Agency and against Google Spain and Google Inc. The citizen complained that an auction notice of his repossessed home on Google's search results infringed his privacy rights because the proceedings concerning him had been fully resolved for a number of years and hence the reference to these was entirely irrelevant. He requested, firstly, that the newspaper be required either to remove or alter the pages in question so that the personal data relating to him no longer appeared; and secondly, that Google Spain or Google Inc. be required to remove the personal data relating to him, so that it no longer appeared in the search results.[43]"

The Spanish court which dealt with the case initiated a preliminary ruling procedure[44] at the CJEU. The CJEU was asked with question, whether the "google search[45]" internet service can be qualified as an activity that falls under the definition of personal data processing as per the

[42] See: Case C-131/12: Judgment of the Court (Grand Chamber) of 13 May 2014. Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González.

[43] See: Factsheet on the right to be forgotten ruling: (C-131/12). EUROPEAN COMMISSION. Europa.eu [online]. [cit. 2014-10-25]. Available at: http://ec.europa.eu/justice/data-protection/files/factsheets/factsheet_data_protection_en.pdf

[44] The preliminary ruling procedure has been explained by the Court of Justice as follows: The national courts in each EU country are responsible for ensuring that EU law is properly applied in that country. But there is a risk that courts in different countries might interpret EU law in different ways. To prevent this happening, there is a 'preliminary ruling procedure'. If a national court is in doubt about the interpretation or validity of an EU law, it may – and sometimes must – ask the Court of Justice for advice. This advice is called a 'preliminary ruling'. See: Preliminary ruling procedure, EUROPEAN COUR OF JUSTICE, [cit. 2014-10-25]. Available at http://europa.eu/about-eu/institutions-bodies/court-justice/index_en.htm

[45] This activity was described as a provider of content, consisting of locating information published or included on the net by third parties, indexing it automatically, storing it temporarily and finally making it available to internet users according to a particular order of preference, when that information contains the personal data of third parties, see paragraph 20 of the Google Spain ruling

data protection directive[46]. A positive answer to this question would mean that Google is a "data controller" and as such has to comply with EU regulations and is obliged to grant certain individuals the "right to be forgotten[47]" i.e. to erase any processed information on request.

In a surprise to many, the CJEU eventually ruled that the activities of a search engine consisting of finding information published or placed on the internet by third parties, automatically indexing it, temporarily storing it and making it available to internet users according to a particular order of preference must be classified as 'personal data processing' in accordance with the Data Protection Directive when that information contains personal data and, secondly, the operator of the search engine must be regarded as the 'controller' with regard to said processing[48]. In order to comply with the rights laid down in the directive, the operator of a search engine is obliged to remove links to web pages published by third parties and containing information relating to a certain person from the list of results displayed following a search made on the basis of the given person's name,[49] if the information is inadequate, irrelevant or excessive in relation to the purpose of the processing.

The main practical consequence for grey literature repositories is that the activities which comprise the indexing and storing documents for the purpose of making them available on the internet are qualified as personal data processing, if the documents contain any personal information. Given that the repositories typically categorize and index their documents, enable full text searches within documents and make the documents available online, they share similarities to the functionality of the google search engine and the conclusions in Google Spain case would most likely apply to them in any eventual dispute with a data subject. Hence, the operators of the repositories have to acquaint themselves with the data protection rules and adapt their internal policies accordingly.

The basic legal framework

The general legal framework for data protection

The Charter of Fundamental Rights of the European Union guarantees every EU citizen and resident the "*right to respect for his or her private and family life home and communications*," and the right to the *"protection of personal data concerning him or her."* Personal data must be: "*processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified.*

More detailed rules on the protection of personal data are contained in the secondary EU legislation. The European Union has harmonized its legal framework for data protection in

[46] DIRECTIVE 95/46/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data (OJ L 281, 23.11.1995, p.31) as amended by Regulation (EC) No 1882/2003 of the European Parliament and of the Council of 29 September 2003

[47] For further reference on right to be forgotten see BENNETT, Steven C. Right to Be Forgotten: Reconciling EU and US Perspectives, The. *Berkeley J. Int'l L.*, 2012, 30: 161. ROSEN, Jeffrey. The right to be forgotten. *Stanford law review online*, 2012, 64: 88. Ausloos, Jef. "The 'Right to be Forgotten'–Worth remembering?" Computer Law & Security Review 28.2 (2012): 143-152.

[48] See the final ruling of the Case C-131/12: Judgment of the Court (Grand Chamber) of 13 May 2014. Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González.

[49] See Ibid.

the Data Protection Directive[50]. The rules outlined by the directive are subsequently implemented by individual member states within their respective national laws and are enforced by their national data protection agencies. Therefore, there is not a single European law on the protection of personal data, there are 28 national laws enforced by 28 national agencies. The national laws are based on the same concepts due to harmonization, but they are not identical among member countries.

The Data Protection Directive defines the rights and duties of the "controllers[51]" and "processors[52]" of personal data and the rights of the "data subject" i.e. the individuals to whom the processed personal data relates. The definition of personal data is very broad and comprises any information relating to identified or identifiable natural (i.e. not legal) person. The directive sets forth the conditions under which the processing of personal data is **legitimate**. Personal data can only be processed, if the data subject has unambiguously given his or her consent or under the exemptions provided by the directive, such as the performance of a contract, compliance with a legal obligation, a task carried out in the public interest or when pursuing other legitimate interests. Conversely, the data subjects have the right to withdraw their consent, to be informed of the extent of the processing of the data and the right to object to the processing of data or even to demand its erasure.

As was mentioned above, one of the main conclusions of the Google Spain case is that even the indexing of documents which contain personal data for the purpose of creating a search engine constitutes "data processing" and an entity which operates such a search engine or stores such documents is indeed a "data controller". The legal status of the "data controller" and the "data processor" is associated with specific duties prescribed by both European and national laws. These duties especially relate to the proper administration of the data, the legitimate use of the data and the relations towards the "data subjects"[53].

The duties related to the proper administration of the data would not be too burdensome for a repository which is operated in a professional manner. These include the duties related to data quality, such as the duty of specifying the purpose of the processing data or the duty of processing adequate, relevant, not excessive[54], accurate and, if possible, up-to-date data[55] and the duties related to data security (i.e. the appropriate organizational and technical measures). Repositories usually comply with these rules by complying with general professional standards without having specific data protection directives in mind.

Administering and monitoring the legitimacy of data use can be much more complex and challenging. The general rule, as set out in the Data Protection Directive is that personal data can be processed solely on the basis of data subject's consent or on the basis of the data controller's statutory right.

[50] DIRECTIVE 95/46/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data (OJ L 281, 23.11.1995, p.31) as amended by Regulation (EC) No 1882/2003 of the European Parliament and of the Council of 29 September 2003

[51] The controller is the person who determines the purpose and means of processing of personal data

[52] The processor is the person who processes the personal data on behalf of the controller

[53] i.e. an identified or identifiable natural person See Art. 2 a) of the Data Protection Directive

[54] See Art. 6 par. 1. c) of the Data Protection Directive

[55] See Art. 6 par. 1. d) of the Data Protection Directive

The legal rules concerning the "right to be forgotten"

It has to be said that the "right to be forgotten" is a rather popular term which can be found in newspapers or scholarly articles more often than in the legislation of case-law. The Data Protection Directive does not mention the "right to be forgotten" explicitly. Even the Google Spain ruling (which is sometimes referred to as "the Right to be Forgotten Ruling[56]") uses this phrase only twice, and even then only when paraphrasing the complainant's argument.

From the perspective of the European Data Protection Directive, it is possible to define the right to be forgotten as the aggregate of several rights granted by the European Commission. Most notably the rights granted by: Article 12(b) - Every data subject has the right to obtain from the controller, *as appropriate, the rectification, erasure or blocking of data, the processing of which does not comply with the provisions of this Directive, in particular because of the incomplete or inaccurate nature of the data;* Article 14(b) – The data subject has right to object at *any time on compelling legitimate grounds relating to his particular situation to the processing of data relating to him. Where there is a justified objection, the processing instigated by the controller may no longer involve those data;*

Grey literature repositories may therefore face requests for the erasure of certain data or documents contained in their collections.

Possible approaches to the data protection

A rigorous pro-active approach

Considering the fact that repositories usually gather and collect their documents from third parties, the repositories might face a very complicated task in obtaining proof as to the fact that the uploaded documents have been created with the due consent of all the data subjects mentioned within the documents. The *most rigorous approach* to compliance with the personal data legislation would require:

1) the examination of all the documents contained in the repository,
2) the identification of which documents contain personal information,
3) the evaluation of which personal information may be processed without the consent of a data subject
4) the processing of the personal data identified in point 3
5) the selection of any personal data which can be processed with the consent of the data subject
6) the identification and contacting of the data subjects in order to receive their approval
7) processing the data for which consent has been obtained and erasing the data for which consent was not or could not be obtained

The aforementioned approach to the personal data is time consuming and expensive. The costs of this approach would not be justifiable for many repository operators, especially in cases where the repository stores large amounts of documents which contain only coincidental and random personal data.

[56] See: Factsheet on the right to be forgotten ruling: (C-131/12). EUROPEAN COMMISSION. Europa.eu [online]. [cit. 2014-10-25]. Available at: http://ec.europa.eu/justice/data-protection/files/factsheets/factsheet_data_protection_en.pdf

The repository operator might find itself in a paradoxical situation when it discovers that it stores or indexes documents with personal data towards which it does not have any legitimate entitlement. The natural reaction would be to anonymize or de-identify these documents or even to delete them completely. However, such actions would fall under the definition of data processing[57]. This leads to a formal loophole where the data controller is not entitled to either store the information or delete it. In these cases, we believe that the documents could (and should) still be de-identified[58], anonymized[59] or deleted, because the effort to terminate the unlawful status can be justified by Article 7 in the same directive which enables the processing of personal information in a way which is necessary for compliance with a legal obligation to which the controller is subject.

A reactive approach – a privacy policy

Grey literature repositories which make available large amounts of documents containing little, random and coincidental personal data produced by third parties are unlikely to have enough resources (neither financial nor personal) to adopt the rigorous approach described above. The necessary investments for a rigorous approach might not appear cost effective in relation to the actual risk of harming someone's rights.

In these cases, we recommend the formulation and publishing of a specific privacy policy that would transparently set out the process of how to indicate any potential infringements of the personal data rules and how the institution would proceed after being notified as a minimum standard. Such a policy should contain the exact identification of the institution that is responsible for operating the registry and indicate the contact person (or designated department) where the requests for the removal of displayed personal data can be forwarded to. We advise the definition of a specific request format or the provision of a form or electronic tool for users to report alleged data infringements. It is highly advised that the institution should indicate why it operates the repository (or search tool) and any public interest that lies behind making such documents available. This information will be useful when justifying the display of any such information, if the institution decides not to comply with a request to delete or anonymize any personal data contained in the repository.

The administration of the requests to be forgotten

Justified grounds for refusal

Neither the European personal data legislation nor the CJEU case-law can be interpreted as meaning that the repository is obliged to remove or anonymize any personal information contained in its repository. Whenever a repository operator receives a request to remove certain documents or personal information displayed online, the operator should consider whether or not it can justifiably refuse such request.

The justification for a refusal is not the same category as the legitimate basis for personal data processing. For example, in the case of Google Spain, the data controller processed data which

[57] Under the Article 2 of the Data Protection Directive, personal data processing ('processing') means any operation or set of operations which is performed upon personal data, whether or not by automatic means, such as collection, recording, organization, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, blocking, erasure or destruction;

[58] By de-identification, we mean the complete removal of a subject's personal information

[59] For the purposes of the article, anonymization is perceived to be a process of encrypting data in a form which enables the re-identification of data

it had gained lawfully, but it was still obliged to remove the personal information upon request, because it could not justify processing the personal information after receiving the data subject's request.

Therefore, dealing with a request to remove personal information contains two steps. **In the first step**, the operator decides, whether it is legally entitled to process such information. **In the second step**, the operator decides, whether the data subject's objection is justified or whether the data controller can justify a refusal of the request. The data controller considers whether the extent and manner of the processing of such information is truly relevant and proportionate (i.e. justifiable) to the purpose of the processing. The relevance and proportionality of this data may change over time, as the CJEU pointed out, *even the initially lawful processing of accurate data may, in the course of time, become incompatible with the directive where that data is no longer necessary in the light of the purposes for which it was collected or processed. That is particularly so where it appears to be inadequate, irrelevant or no longer relevant or excessive in relation to those purposes and in the light of the time that has elapsed.*[60]

The court also ruled that the denial of the erasure of personal data by the data controller cannot be justified merely by the data controller's economic interests or merely by the interests of the general public in having access to that information based on a search relating to the data subject's name[61], unless there is a particular reason for displaying such data because of the role of the individual in the public life. The question remains, what other grounds can justify a refusal to remove personal data. We believe that a grey literature repository has the right to display documents which contain specific personal information, if the processing of such personal information is necessary for the performance of a task carried out in the public interest[62,63], especially if it stores documents which have been created in the course of research into or the administration of public matters. We hold the opinion that the repositories cannot justify a refusal to grant a request with the argument that the data has been made available legally online by a third party. This argument might justify the processing of personal data under some jurisdictions, but it is not valid enough to justify the refusal of the data subjects' request to remove his or her personal information. The mere fact, that the information is available from other sources does not clarify whether the processing of such data serves a justifiable purpose.

Complying with the request

If the operator decides to comply with the request, it may remove the document completely, if the document is completely excess for the purposes of the repository. If the document itself is relevant, but the personal information contained therein is inaccurate or irrelevant for the purposes for which the document has been made available, it is possible to merely de-identify

[60] See the final ruling of the Case C-131/12: Judgment of the Court (Grand Chamber) of 13 May 2014. Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González

[61] See the final ruling of the Case C-131/12: Judgment of the Court (Grand Chamber) of 13 May 2014. Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González

[62] See TAN, Domingo R. Personal privacy in the information age: Comparison of internet data protection regulations in the United States and European Union. *Loy. LA Int'l & Comp. LJ*, 1999, 21: 661. NARAYANAN, Arvind, et al. A critical look at decentralized personal data architectures. *arXiv preprint arXiv:1202.4503*, 2012. MILLARD, Christopher; HON, W. Kuan. Defining 'personal data' in e-social science. *Information, Communication & Society*, 2012, 15.1: 66-84.

[63] See Art. 7 (e) of the Data protection directive

the document or anonymize certain information. If the institution operates a search engine, it may be enough to remove the information as a keyword from the search site so that the document still remains indexed, but cannot not found, if the user enters the name of the data subject in the query.

References

AUSLOOS, Jef. The 'Right to be Forgotten'–Worth remembering? *Computer Law & Security Review*, 2012, 28.2: 143-152. Factsheet on the right to be forgotten.

BENNETT, Steven C. Right to Be Forgotten: Reconciling EU and US Perspectives, The. *Berkeley J. Int'l L.*, 2012, 30: 161.

NARAYANAN, Arvind, et al. A critical look at decentralized personal data architectures. *arXiv preprint arXiv:1202.4503*, 2012.

HON, W. Kuan. Defining 'personal data' in e-social science. *Information, Communication & Society*, 2012, 15.1: 66-84.

ROSEN, Jeffrey. The right to be forgotten. *Stanford law review online*, 2012, 64: 88.

TAN, Domingo R. Personal privacy in the information age: Comparison of internet data protection regulations in the United States and European Union. *Loy. LA Int'l & Comp. LJ*, 1999, 21.

USTARAN, Eduardo, The wider effect of the 'right to be forgotten' case, *Privacy and data protection Journal*, vol. 14, Issue 8 (online) [cit. 20-11-2014]. Available from <http://www.hldataprotection.com/files/2014/09/The-wider-effect-of-the-right-to-be-forgotten-case-Eduardo-Ustaran-Hogan-Lovells.pdf>.

EUROPEAN COMMISSION. *Europa.eu* [online]. [cit. 2014-10-25]. Available from: <http://ec.europa.eu/justice/data-protection/files/factsheets/factsheet_data_protection_en.pdf>.

Case C-131/12: Judgment of the Court (Grand Chamber) of 13 May 2014. Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González.

7th Conference on Grey Literature and Repositories: proceedings [online]. Prague: National Library of Technology, 2014 [cit. 2014-12-15]. Available from: http://nrgl.techlib.cz/index.php/Proceedings. ISSN 2336-5021.

The new Creative Commons 4.0 licenses

Matěj Myška

matej.myska@law.muni.cz

Masaryk University, School of Law, Institute of Law and Technology

Abstract

This short paper focuses on the new Creative Commons 4.0 licensing suite in the context of enabling open access to grey literature and the re-use thereof. Firstly, the new licenses are presented with regard to any differences in comparison with version 3.0. The limits of the Creative Commons 4.0 licenses are then also identified. Finally, the licenses are briefly assessed from the point of view of the Czech law. Drawing from this analysis, the paper ends with a concise recommendation of what CC 4.0 users should specifically take into account in order to make grey literature more open.

Keywords

Copyright, Sui Generis Database Rights, Public Licenses, Creative Commons 4.0

This paper is an outcome of the "Free Licenses Integration Project" – registration no. P408/12/2210 that has been financially supported by the Czech Science Foundation.

Introduction

The launch of the new Creative Commons 4.0 Public Licenses (hereafter referred to as "CC 4.0") marks yet another important milestone for the development of the Open Content Movement. Introduced by the Creative Commons on 25th November 2013[64], the new licensing suite aims to be an international tool that fosters the development of the semi-commons.[65] Consequently, nothing has changed in the new version with regard to its basic ideology. CC 4.0 continues to enable authors to allow others to share and use their work, albeit with the option of implementing some restrictions on the granted rights. As the idea, scope and functioning of the Creative Commons have been excellently presented elsewhere[66], this short paper merely focuses on pinpointing the newly adopted elements in the licensing suite (part 1). It then highlights the fundamental problematic issues in CC 4.0 (part 2). Even though it identifies several limits to CC licenses within Czech law, such as the limited enforceability of no warranties & liability limiting clauses and the rights outside copyright (part 3) by following a simple protocol, CC 4.0 licenses are deemed to be viable tool for opening up grey literature (part 4).

Brave new licenses

The main goals and objectives of the Creative Commons Public Licenses 4.0 have been summed up as internationalization, interoperability and longevity.[67] The specific request of the Data/PSI/Science/Education communities will be addressed later. On a practical level,[68] these uplifting proclamations have been implemented by means of the introduction of a number of major changes regarding the scope of the rights granted and the conditions under which these rights are granted.

The most important **change** involves the significant widening of the scope of the license. Unlike version 3.0, the new licenses cover (in every variant) *sui generis database rights*. Thus therestrictive licensing element prohibiting commercial use and/or the creation of derivative works may also apply to the database as a whole, if the database creator so chooses (Section 4 of CC 4.0). In the context of grey literature, this means that a more unified regimen concerning the potentially covered rights is now feasible. Not only the parts (i.e. the grey documents itself) of a grey repository may be licensed, but also the entire repository itself. This also further increases the possible future areas of application for these public licenses. Up until version 4.0,

[64] CREATIVE COMMONS. *4.0* [online]. [Accessed 8 October 2014]. Available from: https://wiki.creativecommons.org/4.0.

[65] It must be noted that not all the CC licenses fulfill the revamped Open Definition 2.0 ("*Open means anyone can freely access, use, modify and share for any purpose (subject, at most, to requirements that preserve provenance and openness"*). The restrictive licensing elements prohibiting the creation of derivative works (ND) and commercial use (NC) effectively negate the aforementioned definition and as such these licenses cannot be deemed to be "open". These considerations shall be taken into account by the potential licensors which would like to make grey literature "open" or to license a directory as open.

[66] LESSIG, Lawrence. The Creative Commons. *Montana Law Review* [online]. 1 January 2004. Vol. 65, no. 1. [Accessed 8 October 2014]. Available from: http://scholarship.law.umt.edu/mlr/vol65/iss1/1. LESSIG, Lawrence. *The Future of Ideas: The Fate of the Commons in a Connected World*. 1st. New York: Random House. ISBN 0-375-50578-4. LESSIG, Lawrence. *Free Culture: How Big Media Uses Technology and the Law to Lock Down Culture and Control Creativity*. New York : Penguin Press. ISBN 1594200068.

[67] CREATIVE COMMONS. *What's New in 4.0 - Creative Commons* [online]. [Accessed 8 October 2014]. Available from: https://creativecommons.org/version4.

[68] The formal changes include the clarification and improvement of the general legibility of the document.

the Open Data Agenda in Europe lay outside the remit of these licenses. The newest version of the Creative Commons is therefore a viable alternative to the Open Data Commons licenses.[69]

A further modification includes the more natural *attribution* of the author that follows a simple rule of thumb, which could be abbreviated as TASL. Accordingly the (re)user must provide information about the work's Title, Author, Source and License, i.e. the link to the variation of the used License. All of this information should be provided in a manner which is reasonable in relation to the means, medium and context. Furthermore, any modifications (i.e. the cropping of a photograph, the translation of a text) must also be indicated. A simple link with all the necessary information can also fulfill this condition (Section 3(a)(2)).[70]

The new CC 4.0 licenses also provide for a specific request for *anonymization* from the author (Sec. 3(a)(3)). Logically, this is not a clause with an automatic effect like the others and it must be specifically invoked by the author against a user who is making further use of the author's work.

A promising institution inspired by Open Source Software licenses involves the automatic *reinstatement* of the rights to use the licensed material as of the date that any violation has been rectified, provided it has been rectified within 30 days of the discovery of the violation by the licensor (Section 6(b)). However, possible claims arising from any infringements in the non-compliance period are not automatically mitigated and could be brought before the respective courts.

With regard to **unchanged** issues, the definition of noncommercial (NC) remains the same. Noncommercial use is therefore understood to be use that is not primarily intended for or directed towards commercial advantage or monetary compensation. This disputed[71] restrictive condition of the CC licenses has recently been at least partially clarified by the interpretational guidelines published by the Creative Commons.[72] Even though they are not binding, they do attempt to provide at least a basic outline of what should be considered to constitute (non)commercial use. Most importantly, it is not the nature of the subject using CC 4.0, but the nature of the use itself that will be decisive. Therefore, even commercial entities can use works that have been licensed for noncommercial use only. Furthermore, the NC clause does not limit the scope of the limitations and exceptions provided by the respective applicable law. The licensor is also not limited to use the work commercially, i.e. to make use of the dual-licensing. Unfortunately, court rulings on cases related to this condition have not fully

[69] See: OPEN KNOWLEDGE FOUNDATION. *Licenses | Open Data Commons*. [online], [Accessed 8 October 2014]. Available from: http://opendatacommons.org/licenses/.

[70] The Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International Public License is further used as the reference variant of the license. Available from: https://creativecommons.org/licenses/by-nc-sa/4.0/legalcode [Accessed 8 October 2014].

71 See e.g. FREE CULTURE FOUNDATION. *Stop the inclusion: n of proprietary licenses in Creative Commons* 4.0. [online]. [Accessed 8 October 2014]. Available from: http://freeculture.org/blog/2012/08/27/stop-the-inclusion-of-proprietary-licenses-in-creative-commons-4-0/; HAGEDORN, Gregor, MIETCHEN, Daniel, MORRIS, Robert, AGOSTI, Donat, PENEV, Lyubomir, BERENDSOHN, Walter and HOBERN, Donald, 2011, Creative Commons licenses and the non-commercial condition: Implications for the re-use of biodiversity information. *ZooKeys* [online]. 28 November 2011. Vol. 150, no. 0, p. 127. [Accessed 8 October 2014]. DOI 10.3897/zookeys.150.2189. Available from: http://www.pensoft.net/journals/zookeys/article/2189/abstract/creative-commons-licenses-and-the-non-commercial-condition-implications-for-the-re-use-of-biodiversity-information.

[72] CREATIVE COMMONS. *NonCommercial interpretation - CC Wiki.* [online]. [Accessed 8 October 2014]. Available from: https://wiki.creativecommons.org/NonCommercial_interpretation.

comprehended the proper functioning of this clause. In the *Curry v Audax*[73] case, the Amsterdam District Court did not award any damages to Mr. Curry. Interestingly, the court stated that the CC licensed photos had no commercial value. In another Creative Commons related case, *Deutschlandradio,*[74] the German District Court in Cologne deemed that the NC clause was not specific enough and explained it in accordance with the "Zweckübertragungslehre" as "only for private use".[75] However, this decision is not final and it has been already appealed.

Finally, the functioning of the ShareAlike clause also remains unchanged. This condition is therefore only triggered when licensed work is adapted in a manner that requires permission from the original right holder. However, no matter how generally formulated this clause is, it will still be interpreted according to the national law as will be shown in the following section.

The Limits of the Creative Commons 4.0

Even though CC 4.0 strives to operate internationally in the case of a dispute, the licenses will ultimately be subject to the interpretation of a national court. The licenses therefore do not operate on a transnational level, but will be applied within national legislations due to the principle of the territoriality of copyright. However, CC 4.0 does not include a choice of jurisdiction/law clause as the local "ported" versions 3.0 of the licenses did. The creation of ported (tailored to the national law) versions of CC 4.0 is not foreseen, so the result is to be decided under the rules of international private law. The first issue at hand is what court will have the jurisdiction to decide the dispute and secondly what law will actually be applicable. The issue becomes even more complicated, if we take into account the fact that public licenses could give rise to copyright infringement issues as well as to contractual issues.[76] In the first case, the competent courts to decide the dispute will be determined[77] on the basis of the general rules set out in Article 2 of the Brussels I Regulation[78] (Article 4 Brussels I bis)[79] Accordingly, the infringer should primarily be sued in the country of its domicile. The right holder may also make use of the special jurisdiction as stipulated in Article 5(3) of Brussels I (or Article 7(2) of Brussels I bis). Consequently, the infringer may be sued in the country where the harmful event occurred or may occur. The substantive body of case law[80] from the Court of Justice of the European Union has further "specified" the right holder's option of bringing an action for

[73] For a detailed discussion of this case, see: WEST, Ashley. Little Victories: Promoting Artistic Progress through the Enforcement of Creative Commons Attribution and Share-Alike Licenses. *Florida State University Law Review*. 2009. Vol. 36, p. 903–930.

[74] Judgment of the District Court in Cologne [Landgericht Köln], 5. 3. 2013, file ref. 28 O 232/13.

[75] According to section 31(5) of the German Copyright Act, in the case of doubts or unclear arrangements it is the presumed that the rights have not been granted and remain with the author.

[76] The legal qualification of public licenses (such as CC 4.0) is not undisputed. See: GUADAMUZ, Andres, 2009, The License/Contract Dichotomy in Open Licenses. *University of La Verne Law Review* [online]. 2009. Vol. 30, no. 2, p. 296–311. [Accessed 1 July 2014]. Available from: http://law.laverne.edu/wp-content/uploads/2010/04/dichotomy-in-open-licenses296.pdf.

[77] For the sake of conciseness I shall further focus on the overview of the EU law in this area. For details see: ROSENKRANZ, Timo, 2011, *Open Contents: eine Untersuchung der Rechtsfragen beim Einsatz "freier" Urheberrechtslizenzmodelle*. Tübingen : Mohr Siebeck. p. 155-231.

[78] Regulation (EC) No 44/2001 of 22 December 2000 on jurisdiction and the recognition and enforcement of judgments in civil and commercial matters

[79] Regulation (EU) No 1215/2012 of the European Parliament and of the Council of 12 December 2012 on jurisdiction and the recognition and enforcement of judgments in civil and commercial matters.

[80] Compare: KYSELOVSKÁ, Tereza. Procesní a kolizní problematika práv k duševnímu vlastnictví se zaměřením na judikaturu Soudního dvora EU. *Revue pro právo a technologie.* 2013, vol. 4, n. 8, p. 19-24;

damages in a different jurisdiction. Thus the alleged infringer may be sued for the whole amount in the state where the event giving rise to it occurred or for the locally limited damages in the country where the infringement had its effect (pursuant to the "mosaic theory" formulated in the *Shevill*[81] case). Recently, the Court of Justice of the European Union broadened the criterion in the criticized[82] *Pinckney*[83] decision to include the mere accessibility of a website offering the infringing works. Thus, the legitimate right holder may claim local damages basically in any EU state.[84] The applicable law should be determined according to Article 8 of the Rome II Regulation[85] as the law of the country in which protection is claimed *(lex loci protectionis).* In dealing with contractual issues, the court would also be determined on the basis of the rules laid out in Brussels I (Brussels I bis). The competent courts should be the ones located in the place of performance of the given obligation. The applicable law will be determined pursuant to the rules of the Rome I Regulation,[86] i.e. as the law of the country *"where the party required to effect the characteristic performance of the contract has its habitual residence"*.

On the basis of the aforementioned rules, disputes regarding CC enforceability may end up before different courts and be assessed within different bodies of laws. Despite the global nature of CC licenses, the respective national law will ultimately define the limits of the granted license. However, no irregularities or deviations from the standard principles have yet been observed in analyses of national decisions dealing with the enforceability of CC licenses. The competent courts have been determined pursuant to the general rule (*auctor sequitur forum rei*) and the applicable law has been the *lex loci protectionis*. Moreover, the courts in Germany have also not found the Creative Commons licenses to be invalid merely because of the simple fact that they were available to the infringing party only in English, even though it was not said party's mother tongue.[87]

The aforementioned issues have a significant impact on another general limitation of CC 4.0 in general, i.e. the maximum possible absence of any warranties and representations whatsoever,

[81] C-68/93, Shevill and Others v Presse Alliance, ECLI:EU:C:1995:61.

[82] See: HUSOVEC, Martin. Comment on "Pinckney": Council Regulation (EC) No. 44/2001 of 22 December 2000 on Jurisdiction and the Recognition and Enforcement of Judgments in Civil and Commercial Matters, Art. 5(3) – Peter Pinckney v. KDG Mediatech AG. *IIC - International Review of Intellectual Property and Competition Law* [online]. May 2014. Vol. 45, no. 3, p. 370–374. [Accessed 8 October 2014]. DOI 10.1007/s40319-014-0192-0. Available from: http://link.springer.com/10.1007/s40319-014-0192-0.

[83] C-170/12 – Pinckney, ECLI:EU:C:2013:635.

[84] As Husovec notes, such an approach raises serious questions as regards the abuse of forum shopping. Even minor claims could also eventually lead to substantive costs on the part of the alleged infringer. See: HUSOVEC, Martin. Comment on "Pinckney": Council Regulation (EC) No. 44/2001 of 22 December 2000 on Jurisdiction and the Recognition and Enforcement of Judgments in Civil and Commercial Matters, Art. 5(3) – Peter Pinckney v. KDG Mediatech AG. *IIC - International Review of Intellectual Property and Competition Law* [online]. May 2014. Vol. 45, no. 3, p. 370–374. [Accessed 8 October 2014]. DOI 10.1007/s40319-014-0192-0. Available from: http://link.springer.com/10.1007/s40319-014-0192-0.

[85] Regulation (EC) No 864/2007 of the European Parliament and of the Council of 11 July 2007 on the law applicable to non-contractual obligations (Rome II).

[86] Regulation (EC) No 593/2008 of the European Parliament and of the Council of 17 June 2008 on the law applicable to contractual obligations (Rome I).

[87] Decision of the Landgericht Berlin, 8. 10. 2010, Gerlach vs. DVU, 16 O 458/10. Judgment of the District Court in Cologne [Landgericht Köln], 5. 3. 2013, file ref. 28 O 232/13.

as well as maximal limitation of liability. In such a setting, Leeuw[88] aptly remarks that, from the point of view of the licensee, Creative Commons Public Licenses are entirely built upon trust between the parties involved. Unfortunately, there has not yet been any case law dealing with the issue of misrepresentation and the exclusion of liability in relation to Creative Commons Public Licenses; for the time being, the discussion is merely academic.[89] However, as copyright does not take into account the *bona fides* in cases of infringement, such an application leads to the problem that has been aptly characterized by Hietanen, namely that *"the legal right holder can sue anyone in the chain of infringement."*[90] In the next section we will present a brief overview of selected CC 4.0 issues in the context of Czech Law.

Creative Commons 4.0 and Czech civil law

Czech Law has acknowledged Creative Commons Public Licenses since 2006. As such, a licensor can make an offer to conclude a licensing agreement with an unspecified circle of users and these users will then enter into the contract as soon as they use the granted rights. Recently, however, the legal nature of "public licenses" has been disputed in the Czech legal doctrine. The debate does not fall within the scope of this paper, however, and CC 4.0 are treated as gratuitous informal (i.e. not written) contracts.[91] The one simple argument that suggests this conclusion is the fact that the Czech Copyright Act[92] only allows the use of copyrighted work by another person upon authorization on a contractual basis.[93]

As a necessary precondition for the seamless functioning of CC 4.0, the licensor must not violate the *"nemo plus iuris ad alium transferre potest quam ipse habet" maxim, i.e. the licensor must not license more rights than it is actually capable of. If this should happen, Section 5 should shield the licensor from any arising claims regarding the deceived licensee. However this full limitation of liability does not completely apply in the Czech Republic. Firstly, the possible waiver of the licensee's claims arising from possible defects in the work (Section 1916 of Act no. 89/2012 Coll., the Czech Civil Code – hereafter simply referred to as the "CzCC") is only valid, if made in writing. Further misrepresentation could be considered to be a legal defect. This does not mean, however, that the licensor cannot be held liable by the legal right holder as noted above. However, section 2065 of the CzCC could also be applied analogously. This section stipulates that the donor will compensate the donee for any damages arising from the willful donation of another party's property and also if the damages resulted*

[88] LEEUW, Emerald De, 2012, ID 2328989: *The Creative Commons License: Is Validity Enough? A Review of Liability and Enforceability in Creative Commons Licensing* [online]. SSRN Scholarly Paper. Rochester, NY : Social Science Research Network. [Accessed 31 July 2014]. Available from: http://papers.ssrn.com/abstract=2328989.

[89] Cf. DULONG DE ROSNAY, Melanie. Open Content Licenses Without Representation: Can You Give Away More Rights Than You Have? *European Journal of Law and Technology* [online]. 17 September 2013. Vol. 4, no. 3. Available from: http://ejlt.org//article/view/237.

[90] HIETANEN, Herkko. *The Pursuit of Efficient Copyright Licensing: How Some Rights Reserved Attempts to Solve the Problems of All Right Reserved*. Lappeenranta : Lappeenrannan Teknillinen Yliopisto. ISBN 9789522146557. P. 123.

[91] For an overview of the national debate cf. TELEC, Ivo. Souhlas, nebo licenční závazek? *Právní rozhledy*. 2013, vol. 21, n. 13/14, p. 457-462. ISSN 1210-6410. HUSOVEC, Martin. Souhlas, nebo licenční závazek? *Revue pro právo a technologie* [online]. 2013. Vol. 4, no. 8, p. 3–8. Available from: http://revue.law.muni.cz/dokumenty/25809. MYŠKA, Matěj. Vybrané právní aspekty otevřeného přístupu k vědeckým publikacím. *Právní rozhledy*. 2014. Vol. 22, no. 18, p. 611–619.

[92] Act no. 121/2000 Sb., Copyright Act, as amended.

[93] As is similarly the case in Germany. See: ROSENKRANZ, Timo, 2011, *Open Contents: eine Untersuchung der Rechtsfragen beim Einsatz "freier" Urheberrechtslizenzmodelle*. Tübingen : Mohr Siebeck. P. 61.

have from a fault (even a legal one) in the donated property that was known to the donor, if the donor had not warned the donee beforehand. The potential licensor will usually be well aware that it is either using or fraudulently licensing the works of others in the creative process. Therefore, the result is that the user of CC 4.0 licensed works can be sued for copyright infringement, but the user can then sue the "licensor" for compensation of any damages paid.

Finally, Czech licensees should bear in mind that CC 4.0 does not automatically grant all the potential rights needed for the use of the respective work. Prominent examples of such lacking rights include, for example, the personality rights regarding any image of a natural person or protected personal data.

Creative Commons 4.0: handle with care

The previous sections have expressed a rather skeptical view of CC 4.0 licenses.[94] However, with a little care exercised by both the licensor and licensee, they may well provide a relatively easy way of opening up grey literature to the public.

Firstly, the licensor should be aware of the consequences of applying the License. The licensor must bear in mind that CC 4.0 as such is an attempt to simplify the incomprehensible legal language and that only the full legal code of the licenses is binding[95]. As such, the licensor must not rely on the commons deed (the shortened, easily readable version of the licenses) or even the icons indicating the chosen variant of the license. In order to be able to grant the selected license, the licensor must adequately secure all the necessary rights and obtain the consent to do so from all the potentially involved parties. These especially involve subjects other than the author (the licensor), which may exercise the rights such as those pertaining to an employer or co-authors.

The aforementioned remarks regarding the limited ability to communicate the entire contents of the licenses apply to the licensee *mutatis mutandis*. A diligent preliminary search is advised, especially when using works licensed under CC 4.0 with the restrictive NC element.[96] The CC licensing system is ultimately built upon trust due to the limited warranties and representations. As such, in the case of really substantive and potentially exposed usage, CC 4.0 may ultimately only function in practical use[97] as information about whom to contact in order to truly check the actual status of the granted usage rights. Consequently, CC 4.0 may become practically

[94] Due to the limited scope, not all of the problematic issues of the CC licenses have been discussed. For further discussion cf. O'SULLIVAN, Maureen. Creative Commons and contemporary copyright: A fitting shoe or "a load of old cobblers"? *First Monday* [online]. 13 January 2008. Vol. 13, no. 1. [Accessed 8 October 2014]. Available from: http://firstmonday.org/ojs/index.php/fm/article/view/2087. DULONG DE ROSNAY, Melanie, 2009. *Creative Commons Licenses Legal Pitfalls: Incompatibilities and Solutions* [online]. 2009. [Accessed 1 July 2014]. Available from: http://halshs.archives-ouvertes.fr/halshs-00671622.

[95] However imprecise simplification may undermine the necessary legal certainty. Cf. MYŠKA, Matěj, SMEJKALOVÁ, Terezie, ŠAVELKA, Jaromír and ŠKOP, Martin. Creative Commons and Grand Challenge to Make Legal Language Simple. In : *AI Approaches to the Complexity of Legal Systems. Models and Ethical Challenges for Legal Systems, Legal Language and Legal Ontologies, Argumentation and Software Agents* [online]. Springer Berlin Heidelberg. p. 271–285. Lecture Notes in Computer Science, 7639. [Accessed 8 November 2014]. ISBN 978-3-642-35730-5, 978-3-642-35731-2. Available from: http://link.springer.com/chapter/10.1007/978-3-642-35731-2_19

[96] Especially because the interpretation of the NC definition may vary before the respective competent national court.

[97] From the legal point of view, the user of the allegedly CC 4.0 licensed work may sue the fraudulent licensor for the damages that are claimed by the legal right holder. However, such an approach could be considered too risky and burdensome.

equal to another alternative tool for dealing with copyright, namely the Konomark.[98] This mark indicates that most rights are sharable and that the licensor is ready to license them for free, but that a request is necessary. Such an interpretation could be regarded, however, as overly protective and in fact goes directly against the basic idea of Creative Commons, i.e. the reduction of transaction costs. In day-to-day practice, the potential licensees should therefore take the risk and trust that the licensor is entitled to issue the licenses. In the Czech Republic, this approach has been reinforced by the provisions of section 7 of the CzCC that presumes that the party that has acted in a certain way has done so honestly and in good faith. Nevertheless they should be ready to face potential infringement claims from the legitimate right holders.

[98] Konomark was developed Eric E. Johnson as an alternative to direct licensing. For details see: http://www.konomark.org/ [Accessed 8 October 2014].

7th Conference on Grey Literature and Repositories: proceedings [online]. Prague: National Library of Technology, 2014 [cit. 2014-12-15]. Available from: http://nrgl.techlib.cz/index.php/Proceedings. ISSN 2336-5021.

Using public licenses for publishing a database of metadata

Michal Reinöhl

michal@reinohl.cz

Masaryk University, Faculty of Law, Czech Republic

Abstract

The paper focuses its attention on the legal and practical possibilities when utilizing public licenses as a tool in making metadata available to the wider public in a form that is as open as possible, as well as on the obstacles that need to be overcome in doing so, in particular with regard to the rights of third parties. The paper will draw on the experience gained from publishing the database of NRGL metadata and will present generally applicable conclusions based on this.

Keywords

Public Licenses, Metadata, Legal Aspects, Databases, Open Data

Introduction

Public licenses are a wonderful tool, which can help spread and share data freely or with minimal restrictions. However, before publishing a database of metadata under the terms of a public license, one should consider a number of issues, some of which may not be immediately obvious, while others may seem discouraging at first glance. These issues deserve to be given some attention and that is the purpose of this paper.

If we intend to publish such a database, we should first consider which rights we hold in relation to the database and how we can dispose with these rights. Then, we should consider which other parties might also hold rights relating to our database or its contents, and the parties on whose rights we might infringe by our actions. Finally, we should select the public license that best fits our needs.

Legal protection of databases

In Czech law, a database is defined in the Copyright Act (Act no. 121/2000 Coll., on Copyright and Rights Related to Copyright[99]) as "a collection of independent works, data, or other items arranged in a systematic or methodical manner and individually accessible by electronic or other means, irrespective of the form of the expression thereof."[100]

Such a database can primarily be protected against unlawful use by means of copyright and the Sui Generis Database Right (SGDR). Both of these protection regimes are set out in the Copyright Act.

A database can also be protected as a trade secret, through unfair competition legislation[101] or by means of contractual conditions[102] etc. However, these forms of protection are not so relevant to the object of this paper.

Copyright

Based on the Berne Convention for the Protection of Literary and Artistic Works and the Database Directive (Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases), the Copyright Act provides copyright protection to any database, "*which by the way of the selection or arrangement of its content is the author's own intellectual creation, and in which the individual parts are arranged in a systematic or methodical way and are individually accessible by electronic or other means.*"[103] This means that in order to enjoy copyright protection, databases only need to meet

[99]An English translation of an outdated version of the Act can be found on the website of the Ministry of Culture: http://www.mkcr.cz/assets/autorske-pravo/12-AZ_2006_v_AJ.pdf

[100]Art. 88 of the Copyright Act

[101]CONNELLY KOHUTOVÁ, Radka. Databáze ve věku informační společnosti a jejich právní ochrana. Vyd. 1. V Praze: C.H. Beck, 2013, s. 129. ISBN 978-80-7400-493-3.

[102]Although some contractual restrictions on the use of a database might not be permitted under EU law, even in relation to databases which are not protected by copyright or the SGDR – the CJEU is to decide on this issue in case C-30/14.

CJEU case on 'screen-scraping' has potential to affect business models, says expert. In: *Out-Law.com* [online]. 2014 [cit. 2014-11-20]. Available from: http://www.out-law.com/en/articles/2014/november/cjeu-case-on-screen-scraping-has-potential-to-affect-business-models-says-expert/.

[103]Art. 2(2) of the Copyright Act; see also Art. 2(5) of the *Berne Convention for the Protection of Literary and Artistic Works and Art. 3(1) of the Database Directive.*

a lower standard of originality than other works, which need to constitute "*a unique outcome of the creative activity of the author.*"[104]

However, some amount of originality and creative work is required even in databases. A database of metadata will usually be very much determined by technical considerations, rules and constraints, and therefore will not be protected by copyright.[105]

As far as the contents are concerned, metadata mostly only involve just facts as such, which are not protected by copyright.[106] There may, however, be some copyright-protected works contained among the data, such as abstracts in bibliographical data. The consequences of this will be discussed later on.

Sui Generis Database Right

The Sui Generis Database Right (SGDR) is a separate protection regime introduced by the Database Directive. According to the Copyright Act, the SGDR *"pertain to the maker of the database, provided that the formation, verification or presentation of the content of the database represents an investment, which is substantial in terms of quality or quantity, irrespective of whether the database or the contents thereof are subject to copyright protection or any other type of protection."*[107] As the SGDR is independent of copyright, both rights can apply to the same database simultaneously and each can even be held by a different person.[108]

The law requires the investment to be qualitatively and/or quantitatively substantial – that means that the investment need not be quantifiable (like e.g. a monetary investment) and as such a substantial investment of effort will be sufficient to warrant protection.[109] [110] This protection is therefore independent of creativity or originality, but it protects the (mostly

[104]Art. 2(1) of the Copyright Act

[105]Judgment of the Court (Third Chamber) of 1 March 2012, Football Dataco Ltd and Others v Yahoo! UK Ltd and Others. Case C-604/10 paras 37 – 39: "*the notion of the author's own intellectual creation refers to the criterion of originality (...) that criterion of originality is satisfied when, through the selection or arrangement of the data which it contains, its author expresses his creative ability in an original manner by making free and creative choices (...) and thus stamps his 'personal touch' (...) By contrast, that criterion is not satisfied when the setting up of the database is dictated by technical considerations, rules or constraints which leave no room for creative freedom.*"

[106]Art. 2(6) of the Copyright Act

[107]Art. 88a(1) of the Copyright Act

[108]This will mostly occur because copyright is held by the author who has to be a natural person, as opposed to the SGDR, which is held by the maker which can be a legal entity.

[109]Recital 40 of the Database Directive

[110]Caution! An investment in the creation of the content is not relevant for SGDR protection – the aforementioned *obtaining* must be interpreted as the finding, acquisition and collection of existing items:

Judgment of the Court of European Union of 9 November 2004, Fixtures Marketing Ltd v Oy Veikkaus Ab. Case C−46/02.: "*The expression 'investment in ... the obtaining ... of the contents' of a database in Article 7(1) of Directive 96/9 on the legal protection of databases must be understood to refer to investment in the creation of that database. It thus refers to the resources used to seek out existing materials and collect them in the database but does not cover the resources used for the creation of materials which make up the contents of a database.*"

economic) investment in its creation.[111] These characteristics make SGDR protection somewhat similar to the Anglo-American *sweat of the brow* doctrine[112] (which is no longer used).[113]

If a database is protected by SGDR, then "*the maker of the database shall have the right to extraction*[114] *or re-utilisation*[115] *of the entire content of the database or of its part substantial in terms of quality or quantity, and the right to grant to another person the authorisation to execute such a right.*"[116]

In principle, the copying (extraction) and/or making public (re-utilization) of a substantial part of a database protected by the SGDR is only allowed with the permission of the maker or when a gratuitous compulsory license applies, such as personal use or use for (non-gainful) scientific or educational purposes.[117] Note that the "*Repeated and systematic extraction or re-utilization of insubstantial parts of the content of the database*"[118] is also disallowed. It is important to understand, however, that the SGDR does not protect the individual contents themselves, but rather the database as a whole.

Because the SGDR was a completely new concept when it was introduced by the Database Directive, a lot of uncertainty arose around it. The ECJ has played an instrumental role in shaping and clarifying the scope and boundaries of the SGDR. Its most important decisions primarily concerned the term *database* and the s*ubstantial investment* required to obtain SGDR protection,[119] and the meanings of the terms *extraction*[120] and *re-utilization*.[121]

Although the SGDR can be transferred, protected databases are more commonly made available via a contractual license[122] and there are standard public licenses suitable for this purpose, which we will discuss later.

---

[111]Recitals 39 – 40 of the Database Directive

[112]LEŠČINSKÝ, Jan. Právní ochrana databází. Vyd. 1. Praha: C.H. Beck, 2003, s. 33. ISBN 8071798339.

[113]The doctrine was definitively rejected in the US by the Supreme Court in *Feist Publications, Inc., v. Rural Telephone Service Co.*, 499 U.S. 340 (1991).

[114]Art. 90(2) of the Copyright Act defines *extraction* as "*the permanent or temporary transfer of all the content of a database or a substantial part thereof to another medium by whatever means or in whatever form*".

[115]Art. 90(3) of the Copyright Act defines *re-utilisation* as "*any form of making available to the public all the content of a database or a substantial part thereof by the distribution of copies, by rental or by the on-line or any other forms of transmission.*"

[116]Art. 90(1) of the Copyright Act

[117]See Art. 92 of the Copyright Act

[118]Art. 90(5) of the Copyright Act

[119]Like the Fixtures Marketing Cases (C-46/02, C-338/02, C-444/02) and the British Horseracing Board Case (C 203/02).

[120]Judgment of the Court (Fourth Chamber) of 9 October 2008, Directmedia Publishing GmbH v Albert-Ludwigs-Universität Freiburg. Case C-304/07.

Judgment of the Court (Fourth Chamber) of 5 March 2009, Apis-Hristovich EOOD v Lakorda AD. Case C-545/07.

[121]Judgment of the Court (Third Chamber) of 18 October 2012, Football Dataco v. Sportradar. Case C-173/11.

Judgment of the Court (Fifth Chamber) of 19 December 2013, Innoweb BV v Wegener ICT Media BV and Wegener Mediaventions BV. Case C-202/12.

[122]CONNELLY KOHUTOVÁ, Radka. Databáze ve věku informační společnosti a jejich právní ochrana. Vyd. 1. V Praze: C.H. Beck, 2013, s. 55. ISBN 978-80-7400-493-3.

Rights of third parties

Database rights

If we want to incorporate someone else's database (we will call it the External Database) or a part thereof into our database, we need to be careful not to infringe on the maker's SGDR or the author's copyright. Since this paper focuses on situations where we will probably not encounter any copyrightable databases, we will only consider the SGDR, but the considerations for copyright are similar to a certain degree.

External Database not published under a public license

We need to determine the following with regard to what has been discussed in the previous chapter:

1. Is the External Database protected by the SGDR? If so, then:
2. Will we include a substantial part of the External Database? If so:
3. Do we have a license to extract and re-utilize a substantial part of the External Database? If so:
4. Will the External Database remain relatively intact when included in our database? If so:
5. Does the license permit us to sub-license to others? Is the scope of the license broad enough?

1. Is the External Database protected by the SGDR?

Question no. 1 can be answered by applying the definition mentioned earlier – we must assess whether the creation of the External Database has required a substantial investment.

2. Will we include a substantial part of the External Database?

There is no exact delimitation of what constitutes a substantial part of a database, but when deciding on the answer to question no. 2, we should remember that a part can be substantial in terms of quantity or quality, where quantity refers simply to the volume of the contents and quality refers to the scale of the investment when obtaining, verifying and presenting the contents (sometimes a quantitatively small part may represent a large part of the investment).[123]

3. Do we have a license to extract and re-utilize a substantial part of the External Database?

Question no. 3 is rather straightforward – in order to include a substantial part of the External Database in our database, we need to extract that part, for which we need a license (or the maker of the External Database could transfer the SGDR to us, but that is not very likely).

4. Will the External Database remain relatively intact when included in our database?

As we mentioned before, the SGDR does not protect the contents of the database themselves, but rather the "databaseness" of the whole. The purpose of question no. 4 is basically to determine whether publishing our database would effectively allow the public to access the External Database (or a substantial part thereof) or just its contents.

[123]See Judgment of the Court (Grand Chamber) of 9 November 2004, The British Horseracing Board Ltd and Others v William Hill Organization Ltd. Case C-203/02

If we take a substantial part of the External Database and append it to our database without any significant modifications, the maker of the External Database will still hold the SGDR to that part of our database. This means that in order to be able to publish our database, we will need a license to re-utilize the External Database. If, on the other hand, we take the data from the External Database, but rearrange it, select only some of it, combine it with different data and then use it in our database, the maker of the External Database will hold no SGDR relating to our database. Given that these examples obviously merely constitute the ends of the spectrum of possible situations, it will often be difficult to determine whether there are third-party SGDR pertaining to our database.[124] It is therefore advisable to be careful when publishing databases that include third-party data and to obtain the appropriate licenses when in doubt.

5. Does the license permit us to sub-license to others? Is the scope of the license broad enough?

Let us say that we have included a substantial part of the External Database in a mostly intact form in our database. Even if the maker of the External Database does grant us a license to extract and re-utilize the External Database, this does not yet mean that we can publish our database under the terms of a public license. According to Article 2363 of the Civil Code (Act no. 89/2012 Coll., Civil Code): "*The licensee may grant an authorisation under the licence in full or in part to a third party only where the license agreement so stipulates.*" This means that the license has to expressly state that we can sublicense the right to extract and re-utilize the External Database to others.

Another important issue to consider is the fact that a public license usually offers a very broad set of permissions. Remembering the *nemo plus iuris* rule (i.e. that no one can transfer more rights than he himself has), we need to make sure that the license to the External Database grants us all the permissions that we intend to offer others through a public license. Standard public licenses apply worldwide, for the whole term of SGDR protection and, barring some express exemptions, without restrictions with regard to the ways of use or quantity.

According to Czech law, if a license does not state any limitations pertaining to the ways of use, the quantity, the territory or the time, these limitations will be determined according to the purpose of the license agreement and, if the scope of these limitations is not otherwise implied by the purpose of the agreement, it is deemed to be as follows:

"*a) The territorial scope of the licence is limited to the territory of the Czech Republic;*

b) The time scope of the licence is limited to the normal time for the given type of work and way of use, however not longer than one year from the date on which the licence was granted; where the work ought to be delivered only after the granting of the licence, the term shall be one year from the date of delivery;

c) The quantity scope of the licence is limited to the normal quantity for the given type of work and way of use."[125]

As we can see, if the license to extract and re-utilize the External Database does not expressly specify its scope, we will have to deduce it from the purpose of the agreement. If it isn't clear that the purpose of the agreement was to allow us to publish our database under a public license, it may be difficult to argue that the purpose of the agreement implies a broad enough scope for

[124]Caution! Remember that regardless of the answer to this question, we still need a license to extract a substantial part of the External Database in the first place

[125]Art. 2376 of the Civil Code, in connection with Art. 2389

the license. Therefore, if we are able to set or negotiate the terms of the license, we should aim to have the license either expressly stipulate that it allows us to publish our database under the terms of a specific public license, or expressly set out its scope at least at the extent of the public license we intend to use.

External Database published under a public license

Using a substantial part of an External Database licensed under a public license is relatively simple (the purpose of public licensing is, after all, to make this simple), but there are still some things that we should bear in mind. Firstly, we need to determine whether there is a limitation to the public license that would prevent us from utilizing the External Database (such as a No Derivatives limitation, clashing Share-Alike clauses, a Non-Commercial condition if we want to use our database commercially, or any other restrictions like technical requirements on derivative databases etc.). As public licenses generally do not restrict multiple licensing, we could try to obtain an individual license from the maker, if the terms of the public license would prevent us from making use of the External Database. We also need to be sure to attribute properly, while remembering that the External Database might contain parts of other databases, so there might be more than one database maker to give attribution to – it is important to find and carefully read the attribution notices.

Personal Data

According to the Personal Data Protection Act (Act no. 101/2000 Coll., on the Protection of Personal Data and on Amendment to Some Acts) and the Data Protection Directive (Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data), personal data is "*any information relating to an identified or identifiable natural person ('data subject').*"[126] This means, for example, that a person's name in connection with any other piece of information about them is deemed to be personal data and therefore some metadata, like bibliographic data, will contain personal data.

Storing personal data, or performing any other operations upon it (i.e. processing personal data[127]), is usually only allowed with the data subject's consent. However, the Personal Data Protection Act provides some exceptions where consent is not needed. Given that we will mostly be processing personal data which is publicly available on the Internet, we will be able to make use of the exception[128] for processing lawfully published data.[129] If we were to process data that was not published, we would almost certainly still be processing the data for the purposes of legitimate interests which also does not require the data subjects' consent. In that case, however, we would need to inform the data subjects of the fact that their personal data was being processed.[130]

There are also other requirements set by the Personal Data Protection Act which aim to ensure that all processing is fair, lawful and not excessive, but considering the minimal scope of

[126]Art. 4(a) of the Personal Data Protection Act and Art. 2(a) of the Data Protection Directive

[127]Art. 4(e) of the Personal Data Protection Act

[128]Art. 5(2)(d) in connection with Art. 4(l) of the Personal Data Protection Act

[129]See NONNEMANN, František. Další zpracování oprávněně zveřejněných osobních údajů. Právní rozhledy: časopis pro všechna právní odvětví. Praha: C. H. Beck, 2011, č. 18.

[130]Art. 11(5) in connection with Art. 5(2)(e) of the Personal Data Protection Act

personal data processing that is required for maintaining and publishing a database of metadata, they will not be an issue for us.

Copyright in contents

If the metadata in our database contains copyright-protected works, such as abstracts in bibliographical databases, we need to consider whether we have the right to use the works ourselves (especially the right to communicate them to the public[131]) and to license this right to others in order to allow for the good re-usability of our database.

These works will often be subject to the Employee Work regime set out in Article 58 of the Copyright Act. In such cases, if we are the employer, we will usually exercise the author's economic rights to the work and should therefore have no problems with it. If we do not exercise these rights, we should try to obtain a very broad license to these works (similarly to the issue described in 3.1., question 5). If we do not succeed in obtaining an appropriate license to some works, we should consider removing such works from our database altogether.

Contracts and policies

Finally, we should consider whether there are any contractual or policy limitations to how we can license our database. For example, we may have received some external funding, which requires the results to be published in a certain way, or there might be guidelines on publishing set by institutional/company policy.

Selecting the right public license

We need to select the appropriate public license based on the limitations that we have discussed in previous chapters and on the purpose we aim to achieve by licensing our database under a public license. An analysis of the available options would be beyond the scope of this paper, but we would probably choose from the range of Creative Commons licenses version 4.0[132], the Open Data Commons licenses[133] or one of the public domain dedication tools such as CC0[134] or ODC PDDL[135]. If none of these licenses suit our needs and limitations, we can consider creating a bespoke license, but it is usually better to stick with the standardized licenses, if at all possible, especially for the sake of interoperability and ease of re-use.

Conclusion

There are a lot of legal obstacles to navigate before one can publish a database of metadata under a public license. It is important to remember that, while sometimes the publishing of the database is simple and without problems, there are often issues that need to be resolved first and in some cases it simply isn't possible at all. However, this should not be discouraging,

[131]Art. 18 of the Copyright Act

[132]About The Licenses. [online]. [cit. 2014-11-20]. Available from: http://creativecommons.org/licenses/

[133]Licenses | Open Data Commons. [online]. [cit. 2014-11-20]. Available from: http://opendatacommons.org/licenses/

[134]Creative Commons – CC0 1.0 Universal. [online]. [cit. 2014-11-20]. Dostupné z: http://creativecommons.org/publicdomain/zero/1.0/

[135]Open Data Commons Public Domain Dedication and License (PDDL) | Open Data Commons. [online]. [cit. 2014-11-20]. Available from: http://opendatacommons.org/licenses/pddl/

because we should be able to identify the threats and choose the right approach when armed with some knowledge and a bit of foresight.

References

LEGISLATION

Act no. 121/2000 Coll., on Copyright and Rights Related to Copyright.

Act no. 89/2012 Coll., Civil Code.

Act no. 101/2000 Coll., on the Protection of Personal Data and on Amendment to Some Acts.

Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases.

Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data.

Berne Convention for the Protection of Literary and Artistic Works and Art.

JUDICIAL DECISIONS

Judgment of the Court of European Union of 9 November 2004, Fixtures Marketing Ltd v Oy Veikkaus Ab. Case C-46/02.

Judgment of the Court (Grand Chamber) of 9 November 2004, Fixtures Marketing Ltd v Organismos prognostikon agonon podosfairou AE (OPAP). Case C-444/02.

Judgment of the Court (Grand Chamber) of 9 November 2004, Fixtures Marketing Ltd v Svenska Spel AB. Case C-338/02.

Judgment of the Court (Grand Chamber) of 9 November 2004, The British Horseracing Board Ltd and Others v William Hill Organization Ltd. Case C-203/02.

Judgment of the Court (Fourth Chamber) of 9 October 2008, Directmedia Publishing GmbH v Albert-Ludwigs-Universität Freiburg. Case C-304/07.

Judgment of the Court (Fourth Chamber) of 5 March 2009, Apis-Hristovich EOOD v Lakorda AD. Case C-545/07.

Judgment of the Court (Third Chamber) of 1 March 2012, Football Dataco Ltd and Others v Yahoo! UK Ltd and Others. Case C-604/10.

Judgment of the Court (Third Chamber) of 18 October 2012, Football Dataco v. Sportradar. Case C-173/11.

Judgment of the Court (Fifth Chamber) of 19 December 2013, Innoweb BV v Wegener ICT Media BV and Wegener Mediaventions BV. Case C-202/12.

Feist Publications, Inc., v. Rural Telephone Service Co., 499 U.S. 340 (1991).

OTHER SOURCES

CONNELLY KOHUTOVÁ, Radka. Databáze ve věku informační společnosti a jejich právní ochrana. Vyd. 1. V Praze: C.H. Beck, 2013, xiv, 207 s. ISBN 978-80-7400-493-3.

LEŠČINSKÝ, Jan. Právní ochrana databází. Vyd. 1. Praha: C.H. Beck, 2003, xiv, 173 s. ISBN 80-717-9833-9.

NONNEMANN, František. Další zpracování oprávněně zveřejněných osobních údajů. Právní rozhledy: časopis pro všechna právní odvětví. Praha: C. H. Beck, 2011, č. 18.

CJEU case on 'screen-scraping' has potential to affect business models, says expert. In: *Out-Law.com* [online]. 2014 [cit. 2014-11-20]. Available from: <http://www.out-law.com/en/articles/2014/november/cjeu-case-on-screen-scraping-has-potential-to-affect-business-models-says-expert/>.

About The Licenses. [online]. [cit. 2014-11-20]. Available from: <http://creativecommons.org/licenses/>.

Licenses | Open Data Commons. [online]. [cit. 2014-11-20]. Available from: <http://opendatacommons.org/licenses/>.

Creative Commons – CC0 1.0 Universal. [online]. [cit. 2014-11-20]. Available from: <http://creativecommons.org/publicdomain/zero/1.0/>.

Open Data Commons Public Domain Dedication and License (PDDL) | Open Data Commons. [online]. [cit. 2014-11-20]. Available from: <http://opendatacommons.org/licenses/pddl/>.