Open academic community in Poland: New scholarly communication models during the transformation period

Maciej Ostaszewski
Information Processing Center, OPI, Poland

Abstract
Digital revolution has contributed to fundamental changes in the way research is conducted and its results are distributed. The emergence of Open Science paradigms has reflected the very essence of those processes. Ideas of Open Science gain popularity and become ever more accepted as a model for conducting research in the 21st century. Its acceptance at all stages of research process and scholarly communication gets ever stronger in international institutions, OECD and, European Commission, in particular. There are many new initiatives constituting awareness on those processes, good practices of implementing the ideas of Open Science are widely promoted, too. The recent recommendation position statement of the European Commission of July 17, 2012 sets a new landscape for future presentation of research results and their sharing. Member countries of EU have been requested there to define an agenda of the related implementation process.

Research scope
The main characteristics of Open Science include among others:

- *Treatment of knowledge as common good*
- *Maximally wide provision of the public free access to knowledge*
- *Building new knowledge communities around open content*
- *Openness and transparency at all stages of the research and communication*
- *Widening of the open access range to publications, research data, educational material, etc.*
- *Promotion of various open access models for knowledge dissemination*

Only during the last decade the understanding of research results as common goods gained wide acceptance. Thus, regardless all enthusiasm declared by a large part of the academic community, comprehensive studies on foundations, chances and risks accompanying the practical implementation of Open Science are needed [1,2].

Few remarks on the levels of Open Science implementation in Poland
Poland is only recently entering the path towards wide-range implementation of Open Science concepts. The launch of the virtual library of science e-infrastructure, open to the entire academic community (equally research staff and students), has been initiated around 1996, some years later accompanied by an expansion of the Digital Library community movement [1].
A groundbreaking change refers to the SyNaT project whose main objective has been to develop and launch a unified networked IT platform for hosting and communication within entire national resource of academic and research data, publications and knowledge documentation. The system has been designed and developed so as cover needs of the entire academic community on national level. As intended, the system will contribute to new quality in a wide range of research and academic education developments in Poland. As an underlying model, the schemes of Open Science are exploited. It is a clear sign on the represented attitude that the Polish Ministry for Science and Higher Education has announced the priority of introducing the open access to published research results based on public funding. The action will comply with the Communication and Recommendation released by the European Commission on July 17, 2012 [2].
Apart from ever more numerous activities contributing to the development of specific technical solutions, comprehensive studies are lacking in Poland that would show and provide diagnosis of the acceptance for those changes. The same refers to the knowledge on the general implementation of Open Science models in research work.

Open Science In Poland – A Quantitative Survey

Objectives Of The Study

A study undertaken at the OPI was focused on:

- Diagnosis of the attitude and awareness level among the Polish academic and research communities on broadly viewed open access to research publications, research data and, more generally, problems of Science 2.0.
- Analysis of the implementation range and level of the open science paradigms and tools to research, including among others open notebook science, open peer review, legal tools and scientific networking.

Motivations

Still an insufficient level of public discussion about Open Science and the lack of surveys dedicated to its main problems were our motives to do a survey on the readiness to accept and support open models in Polish academic community. Analytic data resulting from that research were considered as enabling a diagnosis of the potential barriers, bottlenecks and sources of fear within the community.

Survey Design

- **Survey questionnaire**
  A study based on distributing an e-questionnaire was launched using the LimeSurvey platformI[a]
- **Respondents**
  Polish scientists holding at least PhD degree
- **Methods of collecting data**
  The questionnaire was sent to over 24K scientists in Poland, holding at least PhD degree and registered in the "Polish Science" database, operated by the OPI for the Ministry of Science and Higher Education.
- **Respond rate**
  1300 respondents were logged in to the e-questionnaire. 849 completed it, 456 dropped it out (after starting).
- **Groups of questions**
  Four thematic groups of questions were asked (37 questions) on: Open Science – generalities, Open Access, Open Data, Science 2.0.
  In addition, a supplementary question was asked on: age, sex, discipline, participation in international projects, academic title/degree, experience as research team leader, type of the institution of affiliation

- **Survey specification**
  - Majority of the participants: male - 70%
  - Average age-46
  - PhD holders were the largest group 67%. Accordingly, 18% DSc, 12% professors
  - Field of science
    - Technical science – 24%
    - Arts and humanities – 22%
    - Natural sciences – 20%
    - Exact sciences – 19%
    - Medical sciences 10%
  - Type of institutions
    - Universities – 81%
    - Research institute – 10%
    - Polish Academy of Sciences – 6%
  - Respondents who are not leaders of research projects – 36%
  - Leaders of:
    - Projects – 46%
    - research programs – 6%
    - institutions – 5%
  - 53% respondents participated in international research projects, 48% did not

General Attitude Within the Community

1. OPEN SCIENCE – GENERALITIES

Most interviewed scientists met the concept of Open Science (70%). But the level of knowledge is different among them. Only 14% know exactly what does Open Science mean and the largest group includes those who, despite of facing the idea of Open Science know little or nothing about it (20%). There are still many people who have never met the Open Science idea before (26%).

In this connection, as the term Open Science is quite well known among Polish scientists, a natural question follows of whether Open Science is recognized a good tendency or rather bad?

It turns out that the development/ spreading of Open Science concept is viewed as a positive phenomenon and scientists almost unanimously indicated that it may bring a lot of good consequences for their respective research field (81%). And only few (9%) perceived it as a sort of threat or danger.

Positive attitude toward Open Science general concepts should be treated as a starting point and background for farther characteristics of the openness in science. Open Science certainly is not just abstract notion and its ideas are implemented in rather concrete way. Positive perception of openness values is not so surprising, as the latter validates real attitude of the scientists participating in the survey.

2. OPEN ACCESS

Strong promotion of free access to research literature also in Poland is still quite limited but year by year gains more popularity. It makes the open access aspects best recognizable and crucial dimension of the Open Science idea. It was clearly visible in the survey results. The vast majority of the respondents are very well acquainted with open access ideas while only few have never met the concept.

What is the level of support for the concept of open access? To what extent is the related attitude positive?

It is remarkable that the respondents very strongly sympathize with general principle that research outputs published as articles in scholarly journal should be accessible openly and without any restrictions – 85% declared such opinion. So strong positive attitude does not appear wherever we have asked about other open science dimensions (open data, Science 2.0).

Many respondents share rather positive than negative opinions about open access. Scientists indicate that free access to research literature contributes to raising international visibility of the Polish science and, what is crucial, open access provides wide information on research – such a recognition prevents from repeating research already done. On the other hand, scientists also recognize drawbacks such as higher fraction of poor quality publications due to lack of peer- review verification.

An important observation is shared that open access may be considered as a process more beneficial for the development of the entire system of knowledge rather than it is the case viewed from an individual scientist perspective and the associated research career. Also we have observed relatively large group of scientists who have not expressed opinion about basic issues of the open access (either positive or negative). Those results show that some open access concepts and notions are still not quite clear and known among Polish science community - the latter refers not only to open legal basis models of publishing (such as Creative Commons) but also to digital tools delivering/ providing access to knowledge – repositories. Many scientists (30%) still do not know if the university they are affiliated with provides any repository where they could store publications.

Apart from natural barriers restraining/preventing from fully adoption of Open Access models in Poland such as insufficiently developed economic and legal science system, or low awareness among the scientific community, we also noted, as already mentioned, very positive reactions. On the other hand when we asked the question: do you think scientists in Poland are ready for open access models? In most cases the answer was negative: 70% of respondents declared such statement. This can prove that scientist represents positive thinking toward open access adoption but in the same time think that others scientist are not ready.

However when we take a better look on the aspect of active participating in open access model (publishing) we could find that it is deeply above those positive declarations and not so directly linked. 46% of respondents published their scientific outputs in open access models (34% in golden line, and 12% in green line)

Such situation shows that individuals can be strongly convinced that sharing research results is in general interest of knowledge progress, still they refrain from undertaking any own supporting action. A sort of fear often prevents from any individual implementation of such a model.

The attitude towards open access models is in general positive. Scientists are convinced that it could enriched Polish science thus it should be developed and maintained. The positive opinion about

advantages of open access models mostly refers to science as whole, less to individual perspectives and scientific career. There are still many respondents who have never actively met any open access models. Among many indicated barriers limiting the development of open access models in Poland one of the most important refers to the lacking readiness on community level to adopt open access models. It is quite remarkable that there are many respondents among scientists whose knowledge about open access is rather limited.

3. OPEN DATA

Another central dimension of Open Science refers to the access to research data. Even though the problem of open data is less established in science system, during last years a large amount of new projects addressing this issue were launched. Those initiatives show an increasing need for changes in those matters [3]. In Poland, similarly to open access questions, discussions about open data problems have been just arising among some actors of the science system scene. It is only the context of the EU Horizont 2020 program that problem of access to research data has been articulated like never before.
A limited level of discussion and similar to this lack of any supporting action from the Polish government side may create impression that those issues would not be positively recognized among Polish scientific community. Natural fear seems to be more typical than a wide acceptance.
Those are quite common opinions, still the outputs of our research show that the situation is rather different.
The concept of open access to research data is known for more than half of respondents – 67% of them declare to be acquainted with this concept. 89 % of the respondents claim that sharing research data in research practice may positively contribute to a progress in their discipline. Such a high level of support complies with the main arguments addressed by advocates of open science, that giving and sharing research data would give extra boost to the process of scientific progress.
Problems of open access to research data are more and more recognized as of key importance by public agencies supporting research. Those institutions in many cases not only require open access to research publications funded but also impose obligations for prospective grant beneficiaries on granting access to produced research data (NSF). In Poland national funding agencies such as NCN and NCBiR do not require any specific data management plan yet.
Results of the survey show that those political changes in science even though have not yet directly affect Polish scientists, still could raise positive reactions among them. Most scientists claim that providing access to data after a research project is completed should become rule and obligation rather than just only an individual act of goodwill. 76% of the examined scientists gave positive answer to the question on an obligation of sharing data from publicly funded research.
Many authors consider the introduction of any model of open access to research data to represent a longer-term process. In this process, introduction/maintenance of effective and acceptable standards that define what data, what ranges of data, and specify the rules for data storage, plays a very important role [4]. This process is by no means easy and in many fields of research there are still a lot of controversies about that.
Our survey shows that many scientists have serious concern about this process. Scientists fear that data could be wrongly used and misinterpreted by others. Almost half of them share such an opinion.
It should be mentioned that a fear of providing open access to data is compensated by strong positive tendency to consider lack of access to data as the main source of problems in research practice. 71% of the examined scientists indicated that without access to data the inquiry is less effective while 63% declared that it causes serious slack in research process.
In both cases, equally for positive and negative opinions, we could observe that some 25% of the examined researchers do not have any specific opinion about those issues. The latter may suggest that there is still quite large group of scientists who possibly know little about it.
We also asked scientists about reasons for not sharing research data. The main indicated reason was the lack of appropriate locations to store data (21%) and missing established standards to do so (20%).
Another question refers to the type and range of research data scientists are ready to share. In this connection, 42% of the examined researchers indicated raw data and outputs of some analytics (scripts, code), 57% declared readiness to share detailed documentation of research experiments. 72% support sharing computer software data.

4. SCIENCE 2.0

It can be recently observed that Internet evolves more and more toward higher interactivity, collaboration between web services and stronger position of users who become active producers of the content. This trend contributes to setting up new tools, such as wiki, blogs and social services which allow users to share information and create new projects. Also changes in science practice go further in

this direction [5]. The way scientists could work with each other on global scale by communicating information and spreading ideas is possible like never before. New science of the 21st century, commonly referred to as Science 2.0 directly refers to the use of digital technology and infrastructures.
In our survey we addressed only few questions dedicated to those matters.
The results of the survey show that even though Science 2.0 is not yet popular - 40% of respondents have heard about this idea, with expectations raised. In particular, most of the respondents who have heard about social science portals think those portals can be useful as tools for Science development (76%). They also declared strong will to join such platforms once appeared and gather scientists from the appropriate research field - 77% indicated such expectation. When asked about any reason of not utiling such platforms, the most popular answer of the scientists was "I have no time to do it".
Like in previous cases (open access, open data), the survey showed that the level of participating is still very limited. Scientists hardly ever used blogs, wiki and social science portals dedicated to scientific community. Apart from the lack of time, probably the most obvious reason is that any appropriate incentives promoting such behavior are missing in the science system.
Many authors and advocates of Open Science claim that motivations for scientists to support open knowledge distribution are rather limited because the traditional academic evaluation rules do not promote such a behavior. In this connection, a main dilemma reduces to the question on how to change the culture and system of awards in academia, thus how to ensure that Openness becomes equally advantageous to the entire academic system as also to its individual members [6].

ANALYSIS OF THE ATTITUDE: LINEAR REGRESSION MODELS

The overall community attitude, presented above, gives only basic information of what scientists think about Open Science. In order to disclose more interesting relations between variables there was a need to carry out more advanced statistical tests. Our main goal was to find out how the attitude towards Open Science depends on independent variables: sex, discipline, participating in international projects etc.
The general presumptions we made before statistical analysis is that general attitude toward Open Science could be discussed in three aspects: cognitive (knowledge about Open Science), behavioral (publishing in Open Science), affective (emotions and feelings about Open Science models). This tripartite conception of attitude was taken from popular and classical distinction used in psychology and social psychology: the attitude toward some object/subject addresses these three main aspects [7].
In this survey we asked several different questions about Open Science which could be treated as referring to some dependent variables. To reduce the number of those variables we applied a standard factor analysis, aiming first to reduce the number of variables and secondly to detect an underlying structure of the relationships between them. Factor analysis was applied to those variables that cover cognitive, behavioral and affective aspects of the attitude toward Open Science. Finally we detected three dominating factors:

- Knowledge (about OS),
- Behavior (publishing in Open Access),
- Opinions (positive opinions towards Open Access).

Those three dominating factors were used as new dependent variables. We assumed that they would depend on some independent variables (predictors): age, sex, discipline, participation in international projects, academic title/degree, experience as research team leader, type of institutions, and would also mutually interfere with each other.
To test this presumption we used linear regression models – general statistical method to verify influence one variable on other.

CONCLUSIONS OF THE ANALYSIS OF ATTITUDE

A) OPINIONS ABOUT OPEN ACCESS
- Female scientists are more positive towards open access
- Representatives of arts and humanities are more positive towards Open Access than others.
- Researchers from institutes of the Polish Academy of Sciences are more conservative
- Age, participating in international projects, and experience as a team leader were not statistically significant

B) KNOWLEDGE ABOUT OPEN SCIENCE
- Female scientists show lower awareness of open models as well as staff members of research and R&D institutes (compared to university members)
- Stronger international research record leads to better knowledge on open science
- Age, discipline were not statistically significant

C) PUBLISHING IN OPEN MODELS
- Representatives of life and exact sciences tend more towards open access publishing
- Higher age leads to lower number of open access publications
- Team leaders are more positive towards open access publishing
- Academic title (Professors) holders publish more frequently in open access model
- International research record is strongly correlated with publishing in open access models

D) KNOWLEDGE AND OPINIONS ON PUBLISHING
- Knowledge about Open Science and opinions toward open access have positive impact on frequency of publishing in open models.
- This knowledge has greater positive impact than opinions

GENERAL CONCLUSIONS:
- Open Science adoption is so far limited in Poland, still Polish scientists consider open models as an important driving factor for the progress equally in science as whole and in its individual disciplines
- Strong open publishing promotion, the related systemic solutions and advantages for individuals contribute to the exposure of open access as the main aspect of open science. Other key aspects, incl. open data and Science 2.0, have still somewhat limited visibility.
- Polish research community is split in their attitude towards various dimensions of openness. Knowledge on open science positive attitude towards open publishing as well as publishing in open models are driven by factors whose majority is of international nature, however some local features referring to specific national research model can be observed, too.
- This report summarizes preliminary observations based on the conducted survey. More comprehensive analysis would still require some supplementary research, in particular of qualitative nature.

References:

[1] *Wdrożenie i promocja otwartego dostępu do treści naukowych i edukacyjnych,* (2012), http://www.nauka.gov.pl/fileadmin/user_upload/Nauka/Polityka_naukowa_panstwa/Analizy_raporty_statystyki/20120208_EKSPERTYZA__OA__ICM.pdf

[2] *Commission Recommendation of 17.7.2012 on access to and preservation of scientific information,* http://ec.europa.eu/research/science-society/document_library/pdf_06/recommendation-access-and-preservation-scientific-information_en.pdf

[3] Hofmokl Justyna, Tarkowski Alek, Bednarek-Michalska Bożena, Siewicz Krzysztof, Szprot Jakub (2009), *Przewodnik po otwartej nauce*, http://otwartanauka.pl/wpcontent/uploads/2010/01/przewodnik-po-otwartej-nauce.pdf.

[4] The Royal Society Science Policy Centre (2012) *Science as an open enterprise,* http://royalsociety.org/uploadedFiles/Royal_Society_Content/policy/projects/sape/2012-06-20-SAOE.pdf

[5] O'Reilly Tim (2005), *What is Web 2.0? Design Patterns and Business Models for the Next Generation of Software*, http://oreilly.com/pub/a/oreilly/tim/news/2005/09/30/what-is-Web-20.html.

[6] Nielsen Michael (2008), *The Future of Science*, http://michaelnielsen.org/blog/the-future-of-science-2.

[7] Mirosława Marody (1976) *Sens teoretyczny a sens empiryczny pojęcia postawy: analiza metodologiczna zasad doboru wskaźników w badaniach nad postawami,* http://otworzksiazke.pl/images/ksiazki/sens_teoretyczny_a_sens_empiryczny_pojecia_postawy/sens_teoretyczny_a_sens_empiryczny_pojecia_postawy.pdf

[a] https://www.limesurvey.org/

A Challenge of Research Outputs in GL Circuit: From Open Access to Open Use

Sergey Parinov, CEMI, Central Economics and Mathematics Institute, RAS Russian Academy of Sciences
Mikhail Kogalovsky, Market Economy Institute of Russian Academy of Sciences
Victor Lyapunov, Institute of Economics and Industrial Engineering,
Siberian Branch of Russian Academy of Sciences, Russia

Abstract
Open Access movement and currently formed GL circuit provide the scientific community with unique opportunity to modernize a fundamental part of research life-cycle: processes by which the scientists reuse research outputs when they produce new knowledge and then the community assesses their impact. When scientists mentally manipulate the research outputs, outcomes and other objects of scientific information space they discover relationships between the objects and thereby they reuse it to produce a new scientific knowledge. Some of these relationships become visible in scientists' articles (e.g. by citations). Most of them are directly not observable and may exist in a mental form only. In the paper we propose an "open use" approach for the research area and discuss a practical implementation of the model within a research information system "Socionet" supported a grey literature circuit.

Acknowledgments
We are grateful to Keith Jeffery and Cameron Nylon for useful comments on the "open use" concept. This research is partly funded by Russian Foundation for Basic Research, grant 12-07-00518-A.

Introduction
Cameron Neylon wrote (Neylon 2012): "open access must enable open use" that means the "innovators can manipulate the material" and from the technical side it requests a "standardizing the representations of data and knowledge in ways that make them easily transferable".
From 2000 we are building a step by step for Russian language scientific community a research information system Socionet[1] that implements a combination of the open access and open use approaches.
From the beginning Socionet supports the grey literature circuit. It provides tools for electronic depositing and distributing different types of research materials in a way not controlled by commercial publishers. Socionet services make a standardizing representation of deposited materials within the research data and information space in a way that makes them easily transferable. Socionet users have a personal information robot service, which notifies them about new materials relevant to their interests. At Socionet there is a statistical subsystem which collects various data about scientists' activities in this virtual research environment and produces publicly available and daily updated scientometric indicators.
In the first section of the paper we present our "open use" approach and in the second one – some technical details about the Socionet research information system and about the current state of this approach implementation within this system.

An "open use" approach for research area
The proposed "open use" approach is based on a following assumption: researchers use available research outputs to produce new scientific knowledge when they mentally manipulate the research artifacts, extracted from the reading materials, and discover scientific relationships between the artifacts and their own outputs. Some of these relationships become visible as citations in researchers' outputs. Another part of the relationships is directly not observable since the existed citation technique does not allow researchers to express them explicitly and correctly.
Information about some relationships remains in a mental form only. As a result it is not shared with the research community, it is not utilized in a global research process, and the community has no complete picture about scale and scope of research outputs using and impact.
To respond on this challenge we are developing a concept and building a digital technology of "open use" versus traditional paper-based technology which limitations are mentioned above.
We understand the "open use" for the research area[2] as a process specified by the assumption above, but with at least 5 additions.
1. An open access to research, which is the prerequisite of the open use: all research outputs and full corpus of the Science should be publicly available for using by the community to produce new scientific knowledge. Approaches and a technology to support open access are well known and we do not discuss it here.

2. An openness of results of researchers' manipulation of the materials. It should be clear specified what pieces of the materials were selected by the researcher as artifacts for its further using.

From technical point it needs an approach allowing scientists to share with the community research outputs in more reusable form than traditional journal articles, books, etc., allow to do it. Currently there are technologies supporting micro-publications (Clark et al. 2013), nano-publications (Groth et al. 2012) or research artifacts (Parinov 2010a, 2010b), which designed to be better reusable. Also there is an open annotation approach[3] which allows making research artifacts right over electronic version of publications in all traditional forms.

3. An openness of researchers' motivations to use selected artifacts in producing new scientific knowledge. But only part of researchers' manipulations with the material leads to real using of research artifacts. So it is important to share with the community also details about not using of artifacts, when there were tries and fails. In that case the result of researchers' tries and fails in using the artifacts also have to be publicly available.

Technically it can be resolved by implementing semantic linkage technique (Parinov 2012a, 2012b), which in combination with available ontologies allow scientists to express explicitly their knowledge, opinions and hypotheses about scientific relationships between research artifacts and so can visualize in computer-readable form facts and motivations of using or not using.

4. A guaranteed awareness of researchers on all facts of using their research outputs (tries/fails data and motivations) and about impacts of the outputs.

It can be achieved by creating electronic notification system which will trace facts of research objects using and will provide information about this for all interested parties.

5. An openness of usage statistics aggregated by a research output, a researcher (e.g. for all research outputs by this author) and an organization (e.g. for all research outputs produced by staff), including outgoing usage (e.g. how the object used research outputs) and ingoing as well (e.g. how the object was used by the community).

Technically it needs a monitoring service, which trace all changes in research objects and semantic linkages among them, collect and process this data to provide public scientometric indicators.

In the next section we present a current (November 2013) state of implementing this approach and building the "open use" technology within Socionet research information system.

Socionet overview

The Socionet system development was started in 1997 as a Russian Virtual Laboratory for Economists and Sociologists project. At the beginning it provided a mirror of RePEc.org data and functionality. It also included the first in Russia scientific open archive to submit scientific grey literature in Social Sciences for its online presenting, and some simple tools of virtual workspace (Krichel and Parinov 2002). In 2000 the designed information system got its current name "Socionet" (socionet.ru), since from that time it has own harvester, which federates more research collections and archives, than RePEc provided (Parinov et al. 2003). It allowed a building and, from that time, an everyday updating the Russian research data and information space (DIS) for Social Sciences.

In 2002 a Socionet Personal Zone service was created as add-in online workbench and a managing system for academic electronic assets including the grey literature (samizdat) materials. It allowed a depositing and managing of electronic scientific collections for 9 data types (e.g. "person", "institution", "paper", "article", "book", etc.). The Personal Zone service also included software of the "personal information robot" to trace new additions/changes within DIS according personal research interests of users and notify them about relevant findings (Parinov and Krichel 2004).

In 2004 Socionet users got some new tools to create and manage semantic linkages between information objects of DIS. From that time some information objects in Socionet, like personal and organizational profiles can represent professional social networks of appropriate research actors (Parinov and Krichel, 2004).

In 2007 monitoring of DIS changes and statistics automated services were started. The Socionet scientometric database has been accumulating from 2007.01.01. The Socionet statistics section provides a big set of time series indicators. It includes indicators of views/downloads aggregated according linkages between DIS information objects, e.g. a sum of views/downloads for all publications linked with a personal profile, or the next step of aggregation – a sum of personal indicators for all people linked with an organization's profile, and so on (Kogalovsky and Parinov 2008). The monitoring service of the Socionet can trace changes in linkages including its semantic. Appropriate scientometrics data is adding to the Socionet statistics subsystem (Kogalovsky and Parinov 2009).

In November 2013 the Socionet system federates more than 4000 collections with scientific materials organized on the base of RePEc.org and about 600 collections from Russian research organizations. In total it is about 2 M materials and with every day average surplus of 300 new materials and 1-2 new collections per week. It covers 15 scientific disciplines organized by 16 data types sections. From 2009 the Socionet works as a multidisciplinary RIS freely available for all types of academic actors and based on Open Science ideas (Parinov 2009, 2010b). Socionet tends to be a full-functional modern CRIS driven by the community of scientists speaking and communicating in Russian language (Parinov 2010a).

Socionet currently includes following main subsystems (see Figure 1):
1. Information hub (IH), which federates scientific metadata of RIS, RePEc archives and so on. The IH can harvest local metadata organized in different formats. At IH's output one gets accumulated and daily updated metadata in standardized form. Technically IH's output is designed to fit with software agents and give back metadata through OAI-PMH and other XML-based protocols (Parinov 2006).
2. Interdisciplinary research data and information space (DIS) as a visualization of full IH contents presents existed information objects and semantic linkages between them for navigation and searching by Socionet users.
3. Online workbench to create, manage and submit to DIS single materials, whole collections and archives, and also to create/manage networks of semantic linkages between DIS objects. Any authorized researcher or research organization can use it to provide to DIS a proper professional presentation. Profile of organization with linked collections can be represented as OAI-PMH archive[4].
4. Monitoring and scientometric services, which provides for research community useful scientometric database (updated daily) and notifications. All counted scientometrics indicators are public and can be used for research assessments and scientometric studies.
In November 2013 about 5M semantic linkages exist over research objects of Socionet DIS. At the moment only smaller part of them was created by scientists using Socionet services. About 700 thousands of semantic linkages the Socionet received with RePEc collections (linkages associated with personal and organization profiles). About 4M semantic linkages with the meaning “citation” were imported from the CitEc data base (Barrueco and Krichel 2005).

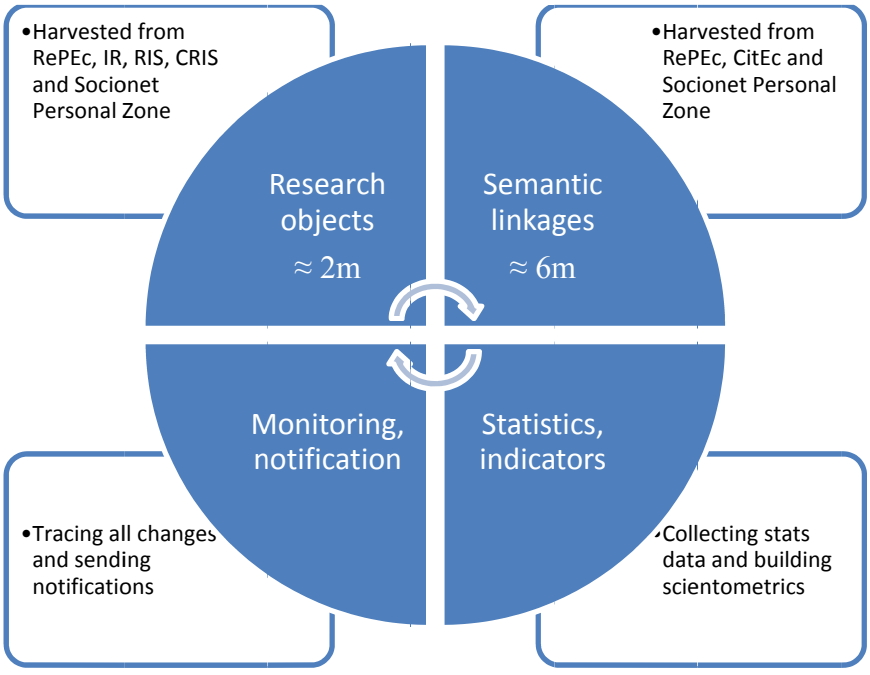


Figure 1. Socionet main subsystems

Automatically accumulated in Socionet semantic linkages data are used: a) to build a visualization of DIS structure in a form of a graph and to provide graphical navigation tool; b) to search linkages according specified parameters (e.g. by creation/revision date, or by usage characteristics, etc.); c) to create reports for notification system; and etc.
The Socionet system uses the CERIF model of standardizing for the representations of data and knowledge. CERIF Semantics and SPAR ontologies for using within Socionet were converted to a form taxonomy represented by a set of semantic vocabularies. The CERIF based semantic linkage technique after some upgrades allows scientists to link different pairs of information objects from RIS content. The semantic meanings assigned by the scientists to the created linkages carry information about classes of relationships between research information objects. The scientific relationship classes are defined by taxonomy based on controlled semantic vocabularies produced from available ontologies.

Implementing "open use" approach at Socionet

Since approaches and a technology to support open access were implemented at Socionet from the beginning in this section we make focus on new tools and services aimed to support the "open use" approach for the research area.

New forms of research outputs for better reuse

Socionet users have an ability to deposit information objects with types "artifact" (Parinov 2010a, 2010b) which has similar functionality with micro-publications (Clark et al. 2013) and nano-publications (Groth 2010). Benefits of depositing research outputs in such new form are (Clark, et. al 2013): "(a) their internal structure is semantically clear and computable; (b) citation networks can be easily constructed across large corpora; (c) statements can be formalized in multiple useful abstraction models; (d) statements in one work may cite statements in another, individually; (e) support, similarity and challenge of assertions can be modelled across corpora; (f) scientific assertions, particularly in review articles, may be transitively closed to supporting evidence and methods."
Motivations to use or not use research artefacts

Socionet users can create semantic linkages between any available research information objects (Parinov 2012a, 2012b). Using available scientific ontologies embedded into the semantic linkage technique users can express explicitly their knowledge, opinions and hypotheses about scientific relationships between research artifacts and so they can visualize in computer-readable form the facts and motivations of using research artifacts for their research process or not using them.
Available for Socionet users a list of motivations is specified by a set of semantic vocabularies presenting scientific relationship classes which can exist between pairs of research objects of different types.
Initial set of rendered scientific relationship classes has been built from different already existed ontologies (Parinov and Kogalovsky, 2011; Parinov 2012a) includes: (1) relationships between research outputs like inference, usage, impact, comparison, evaluation, etc.; (2) relationships between elements of the set {scientists, organizations}; (3) relationships between research outputs on the one hand and elements of the set {scientists, organizations} on the other.
Since a semantic linkage expresses a relationship between two objects, there should be determined which scientific relationship classes (semantic vocabularies) applicable for each combination of pairs from a list research objects' types: a source object type {"person", "organization", "research output", "project", etc.} -> a target object type {"person", "organization", "research output", "project", etc.}.
The initial classes of scientific relationships and a set of semantic vocabularies were proposed in (Parinov and Kogalovsky 2011). For the pair of object types "research output" -> "research output" following classes of scientific relationships and associated semantic vocabularies were specified (ontologies used as a source for semantic vocabularies are mentioned below in brackets):

- Type "Inference", initial semantic vocabulary (CiTO): "obtain background from", "updates", "used as evidence", "confirms", "qualifies", etc.;
- Type "Research usage", initial semantic vocabulary (CiTO): "contains assertion from", "uses data from", "uses method from", "corrects", "refutes", etc.;
- Type "Hierarchy and association relationships", initial semantic vocabulary (SKOS, SWAN): "broader", "narrower", "related", "alternative to", etc.;
- Type "Research material components", initial semantic vocabulary (DoCo): "duplicate", "revised", etc.

When researchers build a linkage between created own research artifact and some other research artifact and assign to the linkages semantics selected from one of four listed above semantic vocabularies, they express their motivations to use the artifact which the linkage is directed to.
Additionally there is a relationship class “Usage proposal” which is also valid for pair of data types "research output" -> "research output” and has initial semantic vocabulary: “can improve”, “can illustrate”, “can replace”, etc. Using it scientists can share with the community their ideas on what research outputs can be used to improve/develop some other research outputs.
For the pair of types "person" -> "research output" there is a class "Professional opinions" with initial semantic vocabulary (SWAN): "responds negatively to", "responds positively to", "responds neutrally to", etc. Using this class of semantic linkages scientists can express their results of “tries and fails” for attempts to use the research artifacts. They also can protest (the value "responds negatively to") against wrong opinions expressed by other scientists with their semantic linkages, etc.

Notifications about reuse

To provide a guaranteed awareness of Socionet users about facts of using their research outputs, including tries/fails data and motivations, and about impacts of the outputs, we are building an electronic notification system[5], which monitors all changes over a set of semantic linkages between Socionet information objects and send e-mail notifications to users who may be interested in this.
Different types of notifications produced by this service support a scientific circulation/communication by distributing signals about semantic linkages creation/revision. This service notifies:

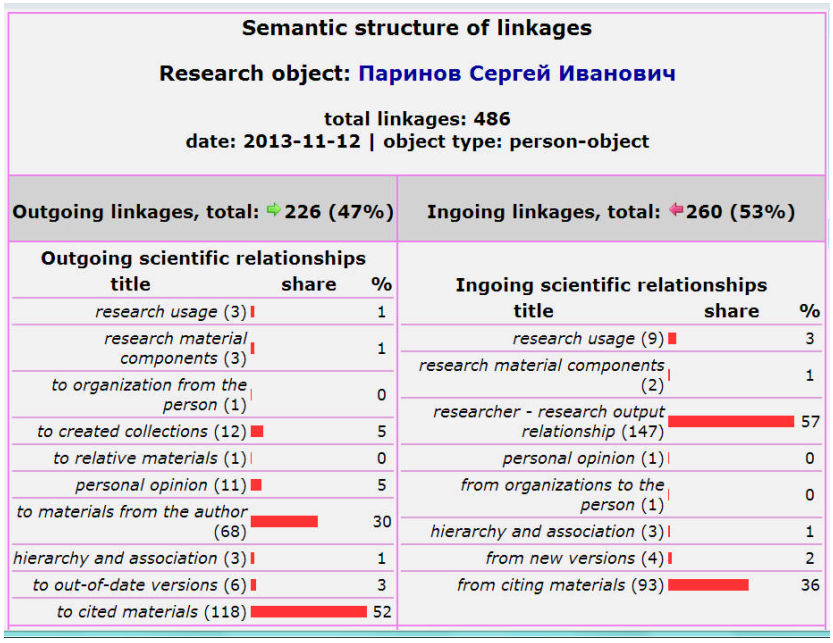
1. the authors of objects linked by created or revised semantic linkage, just to inform them about this event, let them know about specified semantics and give them an ability to react on this event (e.g. to protest against specified semantic);
2. the author who is changing his/her object (e.g. an article), if the object has linked (cited) in other objects (articles), that by this action she/he can violate have established linkages and/or its semantics;
3. the authors of semantic linkages, if there were changes in objects specified as a source and a target of the linkages, so they should reconsider their linkages and, if it necessary, correct it;
4. the users of research DIS while they are viewing some DIS object (e.g. the readers of electronic articles) that certain semantic linkages made for the displaying source object (e.g. citations in reading text) can be violated because of the target object (e.g. cited articles) was changed, and an author of the linkages has not updated suspicious linkages (e.g. citations).

If the first three types of notification in the list above can be made by e-mail only, the last one should work as warning, that displayed on the screen when it necessary.
Thus notification service creates additional “open use” approach benefits since it improves scientific circulation and communication because it immediately informs scientists about using their research outputs and authors of semantic linkages can receive a feedback on their actions from authors of linked research objects. It also improves global research cooperation because researchers can immediately react on how their research outputs were used by the community. A cooperation can have at least two ways: a support of the third party research where their outputs were used or a protest against of wrong using or their research outputs.

Statistics of reuse

The Socionet Statistics subsystem[6] recently was developed to provide needed openness of statistics about research artifacts reuse. An approach for designing research artifacts reuse indicators was presented in (Parinov and Kogalovsky 2013). At the moment (November 2013) this additional functionality is under testing and evaluation. In Socionet still there is no statistically significant amount of semantic data. Examples of reuse indicators below are provided for illustrative purposes only.
The Figure 2 gives an example of overall statistical distribution of scientific relationship classes (described above) where only three of them classify reuse of research artifacts: "Research usage", "Hierarchy and association relationships" and "Research material components". The Socionet semantic vocabularies also cover "Inference" class of relationships which does not presented at Figure 2 but the same classifies a diversity of research reuse.
The Figure 2 illustrates a use case of aggregation of all existed in Socionet scientific relationships between a researcher’s profile (one of authors of this paper[7]) and other research information objects. Total relationships divided on two sets: a) expressed by outgoing semantic linkages, i.e. created by the researcher, and b) ingoing, i.e. created by other scientists with intention to express relationships with research objects belonged to the researcher.

This division on two sets particularly for relationships expressed diversity of research reuse illustrates how the researcher used research artifacts (left column at the Fig. 2) and how the community used research artifacts created by the researcher (right column at the Fig. 2).



Semantic structure of linkages

Research object: Паринов Сергей Иванович

total linkages: 486
date: 2013-11-12 | object type: person-object

Outgoing linkages, total: 226 (47%)

Outgoing scientific relationships

title	share	%
research usage (3)		1
research material components (3)		1
to organization from the person (1)		0
to created collections (12)		5
to relative materials (1)		0
personal opinion (11)		5
to materials from the author (68)		30
hierarchy and association (3)		1
to out-of-date versions (6)		3
to cited materials (118)		52

Ingoing linkages, total: 260 (53%)

Ingoing scientific relationships

title	share	%
research usage (9)		3
research material components (2)		1
researcher - research output relationship (147)		57
personal opinion (1)		0
from organizations to the person (1)		0
hierarchy and association (3)		1
from new versions (4)		2
from citing materials (93)		36

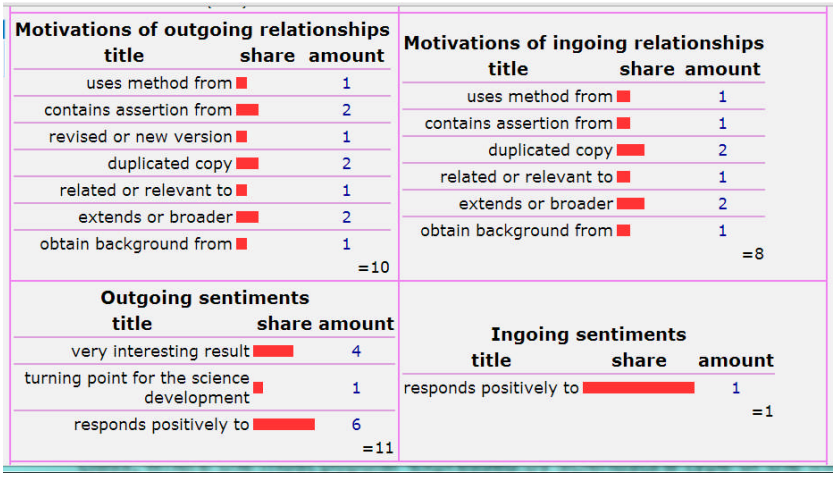
Figure 2. An example of scientific relationship classes distribution for a person

A table on the Figure 3 characterizes the same personal profile as on the Fig. 2 and it illustrates: 1) distributions of outgoing/ingoing motivations in using research artifacts (top row), and 2) distributions of outgoing/ingoing researchers' sentiments about research artifacts and results to use them (bottom row).

Data in the Fig. 3 top row is built as a subset of scientific relationships that classify research using only (4 classes mentioned above), and it is presented on the Fig. 3 by subclasses, i.e. by titles of motivations.

The left column on the Fig. 3 (outgoing motivations) presents a structure of the researcher's motivations in his using research artifacts. The right column (ingoing motivations) presents motivations of the community to use researcher's artifacts.

Data in the bottom row on the Fig. 3 represents a distribution of motivations specified by the relationship class the "Professional opinions" and the associated semantic vocabulary. In the case of "outgoing sentiments" (left column) it characterizes a structure of the researcher sentiments resulted from his mental manipulation with research artifacts. The right column with the "ingoing sentiments" demonstrate a structure of sentiments of the community about the researcher's artifacts.



Motivations of outgoing relationships

title	share	amount
uses method from		1
contains assertion from		2
revised or new version		1
duplicated copy		2
related or relevant to		1
extends or broader		2
obtain background from		1
		=10

Motivations of ingoing relationships

title	share	amount
uses method from		1
contains assertion from		1
duplicated copy		2
related or relevant to		1
extends or broader		2
obtain background from		1
		=8

Outgoing sentiments

title	share	amount
very interesting result		4
turning point for the science development		1
responds positively to		6
		=11

Ingoing sentiments

title	share	amount
responds positively to		1
		=1

Figure 3. An example of motivation distributions of a person

Socionet statistics subsystems provides similar usage "portraits" not only for authors, but also for research organizations and, of course, for research outputs.

This 5th element of our "open use" approach also gives benefits for the community.

Such "semantic cloud" in combination with data about groups of research objects linked by certain scientific relationships makes possible a multilayer stratification of Socionet DIS. By this way one can build a "usage map" for scientific areas, disciplines, specific objects, groups of authors, etc.
Since the semantic data can be aggregated by using information about linked objects (Parinov and Kogalovsky 2013), the aggregators can characterize different objects with variation in selected relationship classes or subclasses. It can present e.g. accumulated usage information about research outputs of one author, or a distribution of motivations and sentiments expressed by one scientist, or the same aggregators for a research organization, a scientific journal, an academic publisher, and so on.
Handling some specific classes of relationships or motivations we can make studies for selected groups of research outputs, authors, or scientific disciplines: which research outputs is used in some specific way, e.g. as a background for scientific inference of another research result, what results are claimed to be a theoretical generalisation of another, and many other according our taxonomy of relationships.

Conclusion

A grey literature circuit supported by a research information system gets a lot of improvements and researchers – users of the system - have essential benefits when the open access and the open use approaches and tools are implemented in the system.

References

Barrueco, J. M., Krichel, T. (2005). Building an autonomous citation index for GL: RePEc, the Economics working papers case. The Grey Journal, 1(2), 91-97.

CERIF 1.3 Semantics: Research Vocabulary. (2012). CERIF Task Group, euroCRIS, http://www.eurocris.org/Uploads/Web%20pages/CERIF-1.3/Specifications/CERIF1.3_Semantics.pdf

CERIF 1.3 Vocabulary. (2012). CERIF Task Group, euroCRIS, http://www.eurocris.org/Uploads/Web%20pages/CERIF-1.3/Semantics/CERIF1.3_Vocabulary.xls

Clark, T., Ciccarese, P., Goble, C. (2013). Micropublications: a Semantic Model for Claims, Evidence, Arguments and Annotations in Biomedical Communications. http://arxiv.org/abs/1305.3506

Groth, P., Gibson, A., Velterop, J.: The Anatomy of a Nano-publication. Information Services and Use 30(1/2) (2010), http://iospress.metapress.com/content/ftkh21q50t521wm2/

Jörg, B., Jeffery, K.G., Dvorak, J., Houssos, N., Asserson, A., van Grootel, G., Gartner, R., Cox, M., Rasmussen, H., Vestdam, T., Strijbosch, L., Clements, A., Brasse, V., Zendulkova, D., Höllrigl, T., Valkovic, L., Engfer, A., Jägerhorn, M., Mahey, M., Brennan, N., Sicilia, M.-A., Ruiz-Rube, I., Baker, D., Evans, K., Price, A., Zielinski, M. (2012a). CERIF 1.3 Full Data Model (FDM): Introduction and Specification. euroCRIS, http://www.eurocris.org/Uploads/Web%20pages/CERIF-1.3/Specifications/CERIF1.3_FDM.pdf

Joerg, B., Ruiz-Rube I., Sicilia M., DVOŘÁK J., Jeffery K., Hoellrigl T., Rasmussen H. S., Engfer A., Vestdam T., and Barriocanal E.G. (2012b). Connecting Closed World Research Information Systems through the Linked Open Data Web. International Journal of Software Engineering and Knowledge Engineering 22, no. 03 (2012): 345-364.

Kogalovsky, M.; Parinov, S. (2008). Metrics of online information spaces. In Economics and Mathematical Methods, v.44, no. 2, 2008 (in Russian), authors' version - http://socionet.ru/publication.xml?h=repec:rus:mqijxk:17

Kogalovsky, M.; Parinov, S. (2009). Scientometrics by using a citation type of linkages in Socionet system. Deposited by authors at Socionet (in Russian), http://socionet.ru/publication.xml?h=repec:rus:rssalc:web-32

Krichel, T.; Parinov, S. (2002). The RePEc database and its Russian partner Socionet. In Russian Digital Libraries Journal vol. 5, no. 2, 2002, http://www.elbib.ru/index.phtml?page=elbib/eng/journal/2002/part2/KP

Neylon, C. (2012) Science publishing: Open access must enable open use. Nature, 2012/12/20/print, http://dx.doi.org/10.1038/492348a

Parinov, S.; Lyapunov V.; Puzyrev R. (2003). Socionet system as a platform for developing information resources and online services for researchers. In Russian Digital Libraries Journal vol. 1, no. 6, 2003 (in Russian) http://www.elbib.ru/index.phtml?page=elbib/rus/journal/2003/part1/PLP

Parinov, S.; Krichel, T. (2004). RePEc and Socionet as partners in a changing digital library environment, 1997 to 2004 and beyond. In Russian Conference on Digital Libraries, Pushchino, Russia, http://eprints.rclis.org/archive/00001830/01/bonn.pdf

Parinov, S. (2006). Information Hubs. Deposited by author at Socionet (in Russian), http://socionet.ru/publication.xml?h=repec:rus:mqijxk:9

Parinov S. (2007). e-Science: Online Future of the Science. In Information Technology, No. 9, 2007 (in Russian), author's version - http://socionet.ru/publication.xml?h=repec:rus:mqijxk:19

Parinov, S. (2009). Electronic libraries development is a way to Open Science. In Proceedings of the XI All-Russian Research Conference "RCDL2009", Petrozavodsk, Russia (in Russian), author's version - http://socionet.ru/publication.xml?h=repec:rus:mqijxk:21

Parinov, S. (2010a). CRIS driven by research community: benefits and perspectives. In proceedings of the 10th International Conference on Current Research Information Systems. Aalborg University, Denmark, June 2-5, 2010, pp. 119-130. http://socionet.ru/publication.xml?h=repec:rus:mqijxk:23

Parinov, S. (2010b). The electronic library: using technology to measure and support Open Science. In: Proceedings of the World Library and Information Congress: 76th IFLA General Conference and Assembly, Gothenburg, Sweden, August 10-15 (2010), http://www.ifla.org/files/hq/papers/ifla76/155-parinov-en.pdf

Parinov, S., Kogalovsky, M. (2011). A technology for semantic structuring of scientific digital library content. In: Proc. of the XIIIth All-Russian Scientific Conference RCDL 2011. Digital Libraries: Advanced Methods and Technologies, Digital Collections, October 19-22, pp. 94–103. Voronezh State University (2011) (in Russian), http://socionet.ru/publication.xml?h=repec:rus:mqijxk:28

Parinov, S. (2012a). Open Repository of Semantic Linkages. In: Proceedings of 11th International Conference on Current Research Information Systems e-Infrastructure for Research and Innovations (CRIS 2012), Prague (2012), http://socionet.ru/publication.xml?h=repec:rus:mqijxk:29

Parinov, S. (2012b). Towards a Semantic Segment of a Research e-Infrastructure: necessary information objects, tools and services. Metadata and Semantics Research, Communications in Computer and Information Science. J. M. Dodero, M. Palomo-Duarte, P. Karampiperis, Eds. Springer, vol. 343, pp. 133-145. http://socionet.ru/pub.xml?h=RePEc:rus:mqijxk:30

Parinov, S., Kogalovsky M. (2013). Semantic Linkages in Research Information Systems as a New Data Source for Scientometric Studies. Published at Scientometrics, Springer Link, August 2013. http://socionet.ru/pub.xml?h=RePEc:rus:mqijxk:31

Shotton, D. (2010b). Introduction the Semantic Publishing and Referencing (SPAR) Ontologies. October 14, 2010. http://opencitations.wordpress.com/2010/10/14/introducing-the-semantic-publishing-and-referencing-spar-ontologies/

Shotton, D. (2010c). CiTO, the Citation Typing Ontology. Journal of Biomedical Semantics 2010, 1(Suppl 1): S6. http://www.jbiomedsem.com/content/1/S1/S6

Shotton, D., Peroni, S. (2011). DoCO, the Document Components Ontology. 17/02/2011. http://purl.org/spar/doco/

Shum, S.B., Clark T., de Waard A., Groza T., Handschuh S., Sandor A. (2010). Scientific Discourse on the Semantic Web: A Survey of Models and Enabling Technologies. Semantic Web Journal: Interoperability, Usability, Applicability. Special Issue on Survey Articles, Ed. Pascal Hitzler, http://www.semantic-web-journal.net/content/scientific-discourse-semantic-web-survey-models-and-enabling-technologies

SWAN (Semantic Web Applications in Neuromedicine) - Scientific Discourse Relationships Ontology Specification. http://swan.mindinformatics.org/spec/1.2/discourserelationships.html

[1] http://socionet.ru/
[2] We consider the open use for research outputs only within the research area and do not discuss in this paper other types of using research, e.g. using of research outcomes, etc.
[3] http://www.w3.org/community/openannotation/
[4] see 21 Socionet based OAI-PMH archives at http://roar.eprints.org/view/geoname/geoname=5F2=5FRU.html
[5] Currently, in November 2013, it is on testing stage
[6] http://socionet.ru/stats.xml
[7] See the source at http://socionet.ru/stat-lnk.xml?h=repec:rus:ecoper:parinov_sergey.56054-1&l=en

Information support of research information interactions of PhD. students in Slovakia

Jela Steinerová and Andrea Hrčková
Comenius University Bratislava, Slovakia

Abstract
The support of research process in information practices has long attracted attention of information science, especially in terms of digital scholarship and science 2.0. However, there is a need to manage large volumes of digital data in science and new information strategies of young researchers should be designed. The goal of this paper is to report on research of information needs of doctoral students in contexts of information support off digital research data and processes. Based on the concept of information interactions we concentrate on information practices of doctoral students. The study was designed as the follow up of previous research projects on relevance judgments of doctoral students and information ecology of the academic environment. The concept of the study includes the following information interactions: research behavior, information use, information seeking, organization of information, information production, social media. We applied qualitative methodology of semi-structured interviews with doctoral students in different disciplines and visualization of information horizons. Results of data analyses confirm differences in information needs and information strategies in research behavior of doctoral students. Research of dicusssion groups is briefly summarized from the viewpoint of interactive interfaces. Main information problems of doctoral students were identified, namely finding focus, expert support, networking and collaborative information behavior in discussion groups. The final model of information interactions in different disciplines is described and recommendations for information portal design are proposed. Grey information objects in information interactions in digital scholarship are identified and value-added services for the community are also articulated.
Keywords: information practices, digital scholarship, doctoral students, "grey" digital information products

Introduction
Information science and scholarly communication become closely integrated within interactions in the electronic environment. A number of new information products (which can be categorized into „grey" literature) appear within these information interactions, e.g. scientific blogs, collaboratories, discussions, large volumes of empirical data, valude-added annotations, comments, categorization, personalization, etc. This paper addresses the issues of emerging „grey" information objects spaces in scholarly information interactions. We present results of a survey of information behavior of doctoral students and a survey of discussion groups and derive a final model and recommendations for value-added ecological services based on (grey) information products embedded in information interactions in digital scholarship.

Digital scholarship and information science
Digital scholarship builds on cyberinfrastructure and ubiquitous digital environment where we can note the convergence of digital libraries and intellectual scholarly information activities (Borgman 2007, Sonnenwald 2013). The data-intensive science concentrates on support of scholars in every phase of their scholarly work. Information science builds new models of social scholarly practices, e.g. collaboration, management of social networks, trust, and security. Some authors introduced the new concept of science 2.0 (Shneiderman 2008). New types of information seeking appear, e.g. knowledge discovery, exploratory searching, faceted searching, intelligent searching. Management of large volumes of data (big data) requires new approaches to processing of scientific records, but also special tools for assistance in interpretations and collaboration in digital humanities, e.g. presentation of research stories, integration of digital cultural objects, sharing data and management of access and copyright.

Information interactions
Human information interaction is a multidisciplinary area that focuses on relationships between people and information (Fidel 2012) and elaborates on information behaviour research. Information interactions in the electronic environment include information practices in using sources, seeking, reviewing, interpretation and production of information. A number of interactive models determine information interactions in information science (e.g. Belkin, Saracevic, Spink and Cole) (Steinerová et al. 2010). In digital scholarship we can apply these interactions to information behavior of different actors, including shared cognitive attributes and information resources and building special information spaces in digital libraries and collaboratories.

Challenges of information science are connected with building new sociotechnical systems and new tools for structuring of knowledge. The concept of information ecology as part of information science can shed light on information interactions from the perspective of relationships between information technologies and people, including procedures, goals, community values and tools. (Steinerová et al. 2013). Critical ecological issues of information interactions are new tools for elimination of information overload and risks of information use, including regulation of information environment, integration and re-use of resources, personalization and optimization of interfaces. The documentation of the research process has long been modeled in CRIS (Current Research Information systems) and CERIF (Common European Research Information Format) for building repositories of recorded information in the research process. It is important that CERIF can help derive new relationships between recorded documents and build rich contexts for ecological information use. Another challenge for information science is the concept of open science (Science 2020) as making scientific data and resource available through intelligent openness mediated by digital data collection, data mining, knowledge discovery. Several examples include bioinformatics and genetics ontology websites, management of scientific text data lifecycle in university repositories, simulations of special cases in humanities.

Related issues of open data and open access help in free access to electronic sources of scholarly production in traditional subject or institutional (university) repositories (e.g. arXiv, CogPrints, RePec). For example in digital social sciences the data of social records (e.g. housing, education, shopping, etc.) are subject to special analyses, re-used in different contexts with special tools regarding verification, privacy protection and copyright. In digital social sciences and humanities new user-driven innovations and creative processes appear in the course of information interactions. Research information interactions can be characterized as information processing in design of research, problem formulation, analyses and syntheses of sources, data gathering, interpretations, experiments, simulations and deriving conclusions. Although it differs in different scientific disciplines, the common intellectual information processes include seeking, reading, writing, creating, citing, disseminating. Our assumption is that with electronic environment and open access the range of grey literature extends to these and other information interactions.

In these contexts we designed a qualitative study of doctoral students in Slovakia with respect to their shared cognitive information needs, communicative and collaborative information practices. We build on previous studies of doctoral students in information science (e.g. Drachen et al. 2011, Steinerová, Grešková, Šušol 2007).

Survey of doctoral students

The context of our research project is focused on cognitive traveling through the web. A part of the project concentrates on information practices of doctoral students as young researchers and teachers. The framework of the research is determined by different cognitive, affective contexts, information tasks, social and organizational contexts.

The main research questions of the study were articulated as follows: Which information needs and behaviors can be identified with doctoral students? Which information interactions are typical for doctoral students in digital environments? Can we develop a model of information interactions for digital scholarship?

The goal of this qualitative research is to model information skills and interactions in research behavior, information use, information production and social media. The research instrument for gathering data was designed, using methodological guidelines for semi-structured interviews including 28 questions. Altogether, 18 doctoral students from different disciplines participated in semi-structured interviews, including 10 women, 8 men, the average age was 26,8 years. (Steinerová 2013). Results confirm differences in information handling in disciplines (types of research) and the need to pay more attention to methodological training of doctoral students. The concept of the study is depicted in table 1.

Tab. 1 The concept of the research

Research behavior	selection of topic, planning of the research process
Information behavior in information use	information strategies, practices serendipitous information gathering
Information gathering and seeking	types of sources information horizon
Organization of information	sorting of sources sorting tools
Social media	use, benefits
Information behavior in production	publishing types of sources; selection of journals, publisher, forms

Information support of doctoral students

Results of data analyses confirm differences in information needs and information strategies in research behavior of doctoral students. Main information problems of doctoral students were identified, namely finding focus, expert support, networking and collaborative information behavior in real and virtual groups. As examples we also used drawing of information horizons of doctoral students.

As for information strategies, the most frequent interactions include browsing, keyword searching, filtering, citation chaining and monitoring of selected authors. The information resources consulted represent Google Scholar, digital libraries and scientific journals. Natural curiosity as the incentive of information need connects with verification of information, problem solving, argumentation. In information sharing new types of products and communications are used, especially discussions, blogs, wiki systems, forums and informal social events in social media. Interactions with social media are mainly passive, especially reading, sharing, distribution of questionnaires and the prevailing usage is mainly for private, personal purposes. It is important to notice that especially in the mixture of formal and informal information itneractions and communication the productions of „grey" documents emerges. The framework of information ecology can help explain other interactions producing grey information products – re-use of data, information and succcessful information strategies.

The section on the influence of the academic community confirmed that academic community can support information practices of doctoral students, especially writing theses, managing citations, sharing of sources and strategies. Navigation and guidance in concept mapping and international expert networking were also considered. Most frequent barriers in information interactions were represented by lack of time, access to sources, disintegration of systems and services and information overload. The problems were identified in terminology development and outdated publications in libraries. Help would be welcome in building methodological knowledge, collaborations with colleagues, e-learning, use of electronic sources. Methodological knowledge is represented by terminology, problem statement, methods and tools which help create the personal information space. Principles of content, context and convenience are typical for information interactions of the community of doctoral students. Doctoral students are in the process of creating their expert networks and the role of supervisors is the most important (Steinerová 2013).

Implications for digital scholarship indicate the tendency of making implicit knowledge (e.g. in social media) explicit in new grey documents and new genres (scientific and research blogs, wikis). Transition from lower level of context to high levels of context help discover knowledge in manipulations with digital objects. Information interactions in digital environment can add value to traditional representations of information objects by means of visualization, interpretation and re-use.

Discussion forums (groups)

In the research of discussion groups (forums) (Hrčková 2013) several contexts of emergent grey documents were revealed. Researchers discover new perspectives (viewpoints) on their topics or issues, best practices (implicit practical experience from colleagues), news, feedback, and problem solving strategies.

Discussion forums are virtual places that gather individuals from the same field of knowledge, interest or expertize with the aim of exchanging ideas and answers (Stuhlman 2010). The role of forums in the scientific environment is slightly underestimated. Nevertheless, currently we can find many discussion forums used by scientists on the Internet. Naming a few of them, e.g. TheScienceforum, Physforum, Sciforums etc., altogether they connect thousands of scientists all around the world without any physical barriers.

According to the typology of Burnett (2000), three main types of content can be found in virtual communities as discussion forums: news, questions and answers and group projects. Members can therefore access the complex solutions of a concrete problem and current news in the field that are not published officially. As such, we can consider discussion forums as "grey" source of information. PhD. students utilize discussion forums for writing their theses and for the direct feedback. Discussion forums are, however, mostly used in computer science in which PhD. students exchange knowledge and solutions to technical problems. Doctoral students utilize the forums to communicate the unclear topics related to their work and a quality "grey" content is to be found in this kind of social media.

In our research (Hrčková 2013) we build on the framework on Preece´s concept (Preece 2003) using the term of sociability to enhance the usability, social interaction and social interface. The framework includes categories of purpose, people, rules, social interaction. We asked directly 161 discussion forum users to evaluate 53 factors that are potentially influencing their activity in discussion forums. The values were set ranging 1 to 5, meaning 5 as necessary factor. We included the technological factors as usability, credibility, security and privacy, sociability factors and content factors. The aim was to create complex recommendations for these special virtual places based on the opinions of discussion groups'

users. We found out that the active as well as passive users need the external motivation for participating in the group in a small extent. The users, according their own words, don´t need any help or special treatment as newcomers. On the other hand, users strongly agreed with the need of creating usable and safe technologies to enable their communication. The basic features of forums as registration, signing in/ out, browsing and contributing should therefore remain simple, learnable and understandable. In the process of contributing, it should be clear, which topic the user is contributing to. The authors should also be able to manipulate with their own contribution (edit, erase and move to another subject if necessary). We concluded this, as placing contributions to the right category gained surprisingly big importance.
If value-added services should be provided for doctoral students, we have to include the information behavior of this target groups into the design of interfaces. In the case of discussion groups, it can be concluded that the design should be clear and simple to enable the fast and professional interaction between the individual PhD. students.

The final model of information support of information interactions
Based on common characteristics of information practices of doctoral students as curiosity, critical thinking, knowledge discovery and on knowledge of discussion forums we designed the final model of information support of information interactions. It is also based on the analysis of digital environments and new emerging representations of information objects (unpublished grey documents). The model identifies information interactions in which possible new information objects in scholarly communication emerge (fig. 1).

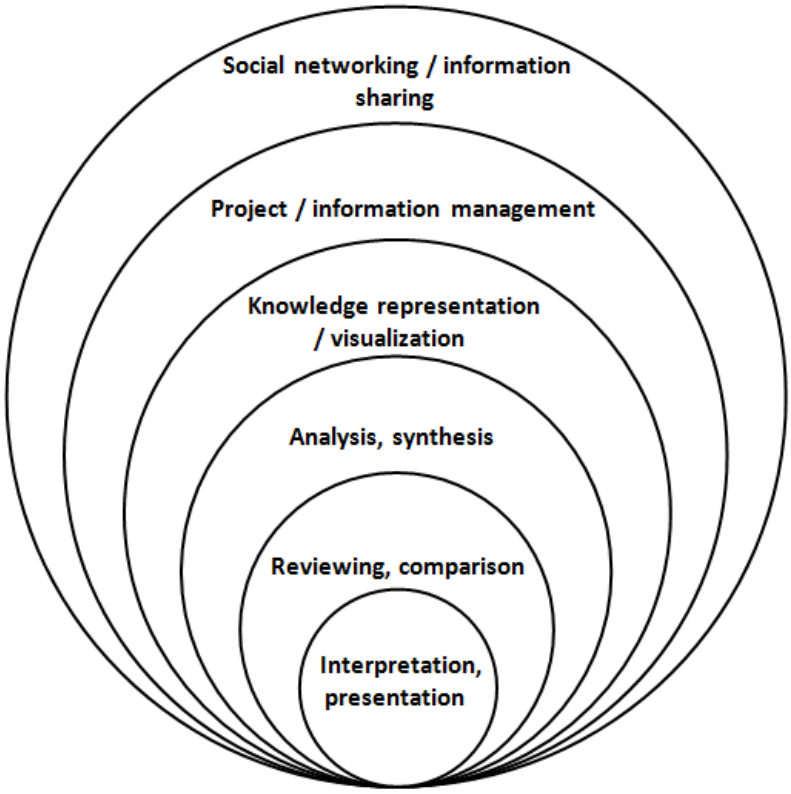


Fig. 1 The model of information support of information interactions

The nested model depicts the main information interactions from social networking (in social media, discussion groups) through information management porcesses (project management) and representation of knowledge and visualization of concepts to basic intellectual information processing in analysis, synthesis, reviewing, comparison, and final interpretation and presentation. The grey information objects embedded in these interactions include contributions and posts in social media, explanations in information sharing, research project proposals and research project reports (temporal, final content reports), visualized representations of content (information horizons, concept maps, other images and pictures representing knowledge), analytical studies and reports (state-of-the-art, literature reviews), peer reviews of projects, papers for journals, theses, expert reviews, comparative studies, interpretations of information objects and artifacts, presentations (e.g. ppt etc.), representations and interpretations of research and learning objects.

Recommendations

These findings lead to recommendations for building special value-added information services based on the common characteristics of doctoral students, knowledge of their problems and information needs for support in information interactions. It has been confirmed that value-added ecological information interactions for digital scholarship should be based on principles of availability, visibility and convenience. Special services should cover support of networking, collaboration, and creativity.

Main features of value-added services of the community portal for doctoral students should cover tools for project management, tools for methodology and methods of research (finding focus, support of creativity). Other features include orientation in professional (electronic) sources and development of terminology (concept mapping and representation of knowledge). Special features should support interactions with supervisors and other experts, social networking with colleagues, and access to methodological knowledge (best practices, methods, tools). As for information literacy, especially methodological literacy for doctoral students is recommended (including concept mapping and terminology, finding focus and synthesis and interpretations of data, and social interactions with experts) (Steinerová 2013).

Emergence of grey documents and information objects in digital environments

Network information environment and changes in information behavior of doctoral students and researchers form cultural challenges for design of information products. New configurations of information products in academic knowledge networks relate to e-products (e-journals, e-book) and to processes of personalization, collaboration, remixing of content. The chain content – service – product is different in production and manipulations of information objects and learning objects in e-learning and e-science. New media formats are closely connected with communities (e.g. doctoral students) with imporrtant characteristics of social awareness, privacy and trust.

Information objects in contexts of digital scholarship provide users with contexts and interests of user communities. The most important feature of new value-added information products is making knowledge visible in representations of values and concepts in contexts. These products can be categorized as those which imitate traditional information products (journals, books, reports, encyclopedias, dictionaries) and those typical for online digital environments (e.g. signal RSS feeds, e-books, websites, community portals). User-generated products transfer social experience into blogs, folksonomies, portals, web archives or special digital libraries, especially by common characteristics and information needs of research communities.

Many of the “grey“ scale information products integrate varied forms of media (texts, videos, photographs and music). Diverse range of new information products is connected by interactivity including such web 2.0 features as adding value by creative use, classification, tagging, comments, annotations, discussions.

In digital scholarship we can determine special interactive digital spaces and places (e.g. research collaboratories) which result from collaborative and social information behavior of scientists. These spaces are „inhabited“ by e.g. large numbers of empirical data (from surveys), statistical data, medical images, digital cultural objects, annotated human genomes, etc.

In several cases the technological interactive features are integrated with content, especially scientific blogs, discussions, reviews, commentaries, annotations, user profiles. Blogging and weblogs develop a special blogosphere and can help discover implicit knowledge in professional communication. Wiki pages and other social software tools support people in collaborative development of documents and new knowledge. Folksonomies are open classificatory systems generated by users through addding value (e.g. keywords) to content. Concept maps and other knowledge maps help visualize knowledge and contexts for better learning and retrieval, including different multimedia forms.

The space for „grey“ information products is opened to other new products, e.g. big data can be represented through linked data into research stories, simulations, genealogies, service prototypes. Further examples of products can include knowedge bases and knowledge maps, ontologies and other products of the use of special knowledge organization tools. Scholars and doctoral students are both producers and users, sometimes publishers. The spectrum of information objects include student products, working papers, images, preprints and postprints, electronic journals and books, digitized objects, data, tools, archives and other outpus of the intellectual life of universities. It is important to integrate the documentation of these grey information objects in digital repositories and special research information systems.

Conclusions

Information support of digital scholarship wil be important in transformation of the digital environment into value-added information spaces with information objects represented by special interactive technological tools. Main information interactions in our model produce special information objects, e.g. from social networking sites to special analytical reports based on data and text mining.

Multiple scholarly interactions require new models for community digital services and products. High level of interactivity can help develop new products in new digital spaces especially in social networking, dialogues and discussions, and reviewing. „Grey“ information products emerge from integration of new tools and technologies with knowledge of information behavior and design of digital systems. Information support is important for building conceptual infrastructures and guidance in research work based on ecological principles of information cleaning, information re-use and interpretations. The identified interactions and examples of „grey“ products can be further modeled in features of digital research information systems.

Acknowledgements

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-0208-10 and the project of the“University Scientific Park” of Comenius University in Bratislava (ITMS 26240220086).

References

BORGMAN, Christine L. 2007. *Scholarship in the Digital Age. Information, Infrastructure and the Internet*. Cambridge: MIT Press, 2007, 336 p. ISBN 978-0-262-02619-2.

BURNETT, Gary. 2000. Information Exchange in Virtual Communities. In: *Information Research* [online]. Sheffield: University of Sheffield. Department of Information Studies, 2000, vol. 5, no. 4 [ret. 2013-09-20]. ISSN 1368-1613. Available from: http://informationr.net/ir/5-4/paper82.html

CONNAWAY, Lynn Silipigni. 2009. Towards a profile of the resarcher of today: what can we learn from JISC projects? In: *JISC Virtual Research Environment and Digital Repository Project* [online]. London: JISC – Joint Information Systems Committee, 2009 [ret. 2013-09-20]. Available from: http://www.jisc.ac.uk/whatwedo/programmes/inf11/researchbehaviourstudy.aspx

DRACHEN, Thea Marie et al. 2011. Information behaviour and practices of PhD. students. In: *hprints.org* [online]. Copenhagen: The Royal Library (Det Kongelige Bibliotek), The National Library (Nationalbibliotek) and Copenhagen University Library (Københavns Universitetsbibliotek), 2011 [ret. 2013-09-20]. Available from: http://hprints.org/docs/00/59/90/34/PDF/Information_behaviour_and_practices_of_PhD_students_appendices.pdf

EMMOTT, Stephen et al. Science 2020. In: *SCIENCE 2020* [online]. Cambridge: Microsoft Research. Microsoft Corporation, 2006 [ret. 2013-09-20]. Available from: http://research.microsoft.com/en-us/um/cambridge/projects/towards2020science/background_2020.htm

FIDEL, Raya. 2012. *Human Information Interaction: An Ecological Approach to Information Behavior*. Cambridge: MIT Press, 2012, 348 p. ISBN 978-0-262-01700-8.

HRČKOVÁ, Andrea. 2013. *Discussion Groups as an Information Resource* [Dissertation Thesis]. Bratislava: UK, 2013.

KROLL, Susan and Rick FORSMAN. 2010. *A Slice for Research Life: information Support for Research in the United States* [online]. Dublin, Ohio: OCLC Research, 2010 [ret. 2013-09-20]. Available from: http://www.oclc.org/research/publications/library/2010/2010-15.pdf

LYNCH, Clifford. 2008. The Institutional Challenges of Cyberinfrastructure and e-Research. In: *EDUCAUSE Review* [online]. Boulder: Educause, 2008, vol. 43, no. 6 [ret. 2013-09-20]. ISSN 1945-709X. Available from: http://connect.educause.edu/Library/EDUCAUSE+Review/TheInstitutionalChallenge/47446

PREECE, Jenny and Diane MALONEY-KRICHMAR. 2003. Online Communities: Focusing on Sociability and Usability. In: JACKO, Julie A. and Andrew SEARS. *The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications.* London: Lawrence Erlbaum Associates, 2003, pp. 596-620. ISBN 0-8058-3838-4.

SHNEIDERMAN, Ben. 2008. *Science 2.0*. In: *Science* [online]. Washington: American Association for the Advancement of Science, 2008, vol. 319, no. 5868, pp. 1349-1350 [ret. 2013-09-20]. Available from: http://www.sciencemag.org/content/319/5868/1349.short

SONNENWALD, Diane H. 2013. Visioning the Future of Information and Library Science: Challenges and Opportunities. In: HOBOHM, Hans-Christoph. *Informationswissenschaft zwischen virtueller Infrastrukur und materiellen Lebenswelten = Information science between virtual infrastructure and material lifeworlds : Proceedings des 13. Internationalen Symposiums für Informationswissenschaft (ISI 2013), Potsdam, 19. bis 22. März 2013* [online]. Glückstadt: Hülsbusch, 2013, pp. 22-34 [ret. 2013-09-20]. Available from: http://opus4.kobv.de/opus4-fhpotsdam/frontdoor/index/index/docId/392

STEINEROVÁ, Jela, Grešková, Mirka and Jaroslav Šušol. 2007. *Prieskum relevancie informácií*: *Výsledky rozhovorov s doktorandmi FiFUK*. Bratislava: CVTI SR, 2007, 150 p. ISBN 978-80-85165-93-7.

STEINEROVÁ, Jela. 2008. Seeking relevance in the academic information use. In: *Information Research* [online]. Sheffield: University of Sheffield. Department of Information Studies, 2008, vol. 13, no. 4 [ret. 2013-09-20]. ISSN 1368-1613. Available from: http://informationr.net/ir/13-4/paper380.html

STEINEROVÁ, Jela, GREŠKOVÁ, Mirka and Jana ILAVSKÁ. 2010. *Informačné stratégie v elektronickom prostredí.* Bratislava: Vydavateľstvo UK, 2010, 192 p. ISBN 978-80-223-2848-7.

STEINEROVÁ, Jela et al. 2012. *Informačná ekológia akademického informačného prostredia. Final report / Záverečná správa z výskumu VEGA 1/0429/10*. Bratislava: Vydavateľstvo UK, 2012, 96 p. ISBN 978-80-223-3178-4.

STEINEROVÁ, Jela. 2012. Information ecology - emerging framework for digital scholarship. In: *Libraries in the Digital Age (LIDA) Proceedings* [online]. Zadar: University of Zadar. Department of Information Sciences, 2012,vol. 12 [ret. 2013-09-20]. Available from: http://ozk.unizd.hr/proceedings/index.php/lida2012/article/view/66/37

STEINEROVÁ, Jela. 2013. Information Interactions as part of digital scholarship. In: Huvila, I. (edt.). *AEW 2013. ASIS&T European Workshop 2013: Proceedings of the Second ASIS&T European Workshop 2013 June 5-6* [online]. Åbo: Åbo Akademi University, 2013, pp. 33-49 [ret. 2013-09-20]. Available from: http://www.abo.fi/institution/en/infovetskrifter

STEINEROVÁ, Jela. 2013. Methodological Literacy of Doctoral Students – an Emerging Model [in press]. In: KURBANOGLU, Serap et al. *ECIL 2013. Conference Proceedings*. Cham: Springer Internationl Publishing Switzerland, 2013, pp. 148-154.

STUHLMANN, Daniel D. 2010. Discussion Forum. In: *Knowledge Management Terms* [online]. Chicago: Stuhlman Managment Consultants, 2010 [ret. 2013-09-20]. Available from: http://home.earthlink.net/~ddstuhlman/defin1.htm

The SK CRIS system as a source of unique information about scientific activities and their outcomes

Danica Zendulkova and Juraj Noge
Slovak Centre of Scientific and Technical Information, CVTI SR, Slovakia

Abstract
Majority of R&D outcomes consists of data, information and documents integrated into a category of Grey Literature. From portfolio of these outcomes especially publications, patents, products and innovations are interesting for scientific community and wide range of R&D stakeholders indoors and abroad.
Current Research Information Systems (CRIS) are one of key software tools for data collection and access, dissemination of information about scientific activities and their outcomes. The EU standards for these systems, primarily the CERIF data format, are covered by the international association euroCRIS.
The aim of providing the Information System about Science, Research and Innovation SK CRIS is to integrate all available research information at the national level. The system was designed on a data structure compatible with the CERIF data format and it became a member of the CRIS systems family.
The SK CRIS contains data about research projects funded by public resources, a registry of researchers, a registry of research organisations and research and development results. Data acquisition and refilling is possible by using integration interface importing data from external systems. It concerns mainly project data imported from systems of grant agencies. Direct data entry made by research subjects by online forms is the second way how to receive data.
The SK CRIS integrates also other available external data sources: registries of researchers from universities and Slovak Academy of Sciences and also publications from universities registered in the Central Registry of Publication Activity.
The SK CRIS makes contextual data about science and research accessible during their life cycle. It means that a user can see whether some document was created as a result of the project. The information about its authors and originators, persons and organisations is also available, as well as usage of a document and its citations. The SK CRIS is not limited to work only with metadata. It contains also the functionality allowing the integration of collected data with fulltext documents.
The benefit of the SK CRIS is mostly its ability to offer aggregated information from entered data and to present their relationship. Main characteristic of the SK CRIS is the CERIF based concept of objects (entities) with attributes, with a data model guaranteeing interoperability and full language variability. The relations and semantics allow to record the objects with relationship and characterise properties and time attributes by roles and time details. However, the system will be used mainly by scientific community but also by research management, decision makers and public.
Enlargement of the SK CRIS functionality and integration of the other external data sources is our vision for near future. In the same time we consider the interconnection with other European CRIS systems. The SK CRIS information system was developed within the Activity No 4 of the NISPEZ national project implemented by the Slovak Centre of Scientific and Technical Information: Enlargement of Central Information Portal for Research, Development and Innovation (CIP RDI) with new functionalities complying with EU standards.

Introduction

The Slovak Current Research Information System (SK CRIS) as a part of the Central Information Portal for Research, Development and Innovation (CIP RDI) has the status of Information system of public administration, defined by the Act 275/2006 on information systems of public administration. The Slovak Centre of Scientific and Technical Information (SCSTI) is responsible for operating, maintaining and providing technical support to this system on behalf of the Ministry of Education, Science, Research and Sport of the Slovak Republic (ministry).
The SK CRIS has been developed within the Activity No 4 of the national project NISPEZ*: Enlargement of Central Information Portal for Research, Development and Innovation (CIP RDI) with new functionalities complying with EU standards.* The system is focused on effective work with research projects funded from public resources on national level including a connection with other, science related information systems.
In the process of system building, the experience of SCSTI with the R&D information system building and operation has been utilised and generally accepted principles of best practice for current research information systems are followed.

The system SKCRIS offers wide-range and detailed map of Slovak science, research and development (R&D) and provides the information in mutual relations to all target groups. Implementation of EU standards for CRIS systems, especially using of data format CERIF ensures a possibility of interoperability with similar systems across EU countries. The system was put into operation in January 2013.

SK CRIS content

The content structure of SK CRIS allows the collection and maintaining of all common types of R&D data (Turňa, 2011). The details of core content entities of our new R&D system SK CRIS are presented below (figure 1).

Figure 1 SK CRIS Main webpage

Registry of Projects

The Research and Development Project is considered as main object (entity) of current research information system: CRIS system is project centric. The information about projects seems to be most relevant for users from different target groups.

Other entities (mainly organisations, researchers, publications, but also events, facilities, equipment, services etc.) are entered into the system mainly via the relationship to any R&D project.

The Registry of projects has been built up since 2000 as part of different versions of R&D information systems. Currently, it contains the data about approximately 10000 projects.

The data about different categories of projects are collecting mainly by integration interface interconnecting SK CRIS with other systems collecting similar data. The systems that will be connected to SK CRIS are mainly the following ones: the system of the Slovak Research and Development Agency (SRDA); Scientific Grant Agency VEGA; Cultural and Educational Grant Agency KEGA and the system of Incentives for Research and Development.

The integration interface allows an automated data import from external systems to SK CRIS. The validation of all imported data is included. The aim is firstly to collect a comprehensive data about research projects included also the responsible organisations and persons, and secondly there is an effort to avoid duplicity of data entries from different systems.

Registry of Organisations

Registry of organisations has been built from several information sources. The first one is annual additional statistical survey of R&D Potential of the Slovak Republic. The information about which organisations carry out research and development is also provided by the Slovak Statistical Office, which receives information about activities of organisations from statistical reporting. Whereas the additional statistical survey of R&D Potential has been carried out since 2000, a significant number of forms were filled in the past. This data is stored in the system and it has been updated annually.

The part of entries into the Registry of organisations is being acquired also during the process of the Certification to perform R&D, which is administrated within the information system. As certification is compulsory for each organisation applying for the support for research from public sources, the registry includes not only data about universities, Slovak Academy of Sciences and state research organisations, but also a significant number of research organisations from non-for-profit and private sector. Currently, the Registry of organisations contains about 1300 subject records. The user interface is shown on figure 2.

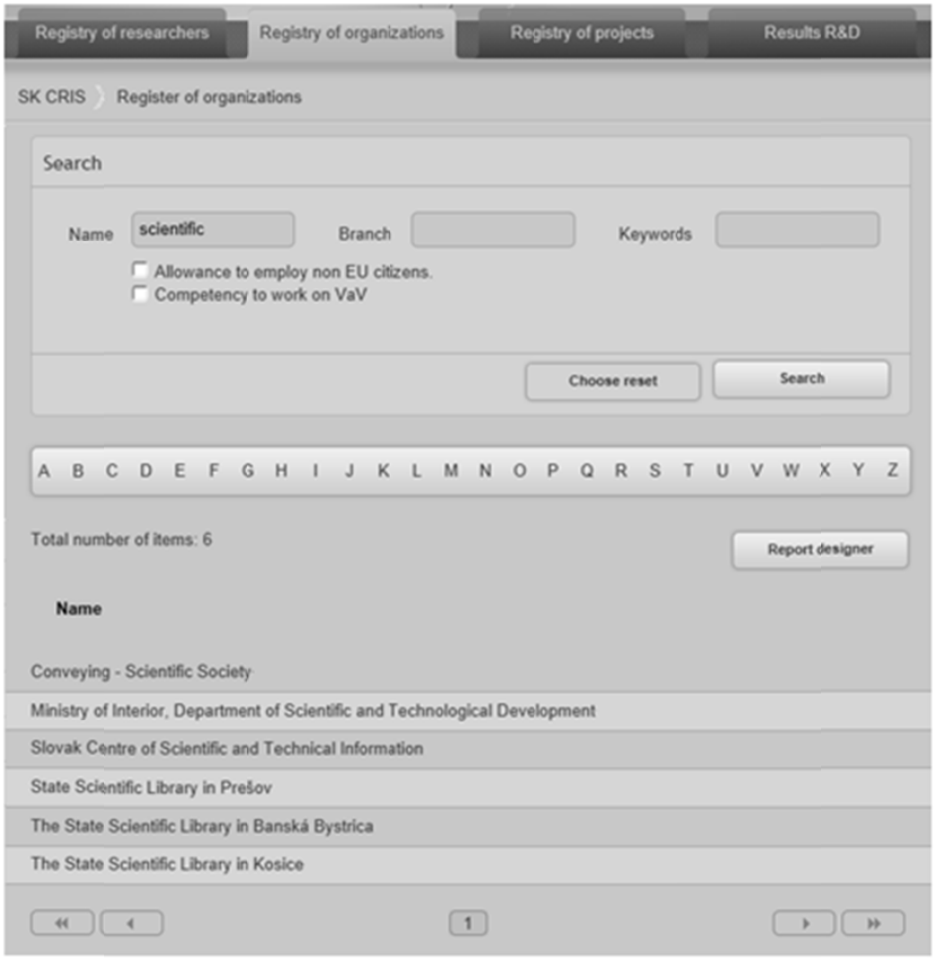


Figure 2 Registry of Organizations, user interface

Registry of Researchers

Registry of researchers has been established on the basis of data from database of experts – evaluators of R&D projects funded by the ministry (1100 personal records). However, the overall goal is to provide the data about all scientists (more than 23000 researchers) in the registry. In first phase of the SKCRIS operation we tried to cover significant and representative part of scientific community - researchers from universities and Slovak Academy of Sciences (SAS) and from state operated research institutes. We realised automated import of entries from registry of employees in higher education sector, and from registry of SAS employees.

Between the other data sources on researchers the registration of research projects have been the most important. It is possible to extract relevant researchers' data (name, affiliation, role in the project) from successful grant applications administrated by different ministries and grant agencies.

Except for the data sources described above, the basic data (name, surname, institution) about researchers from universities can be extracted also from the Central Registry of Publication Activity operated by the SCSTI.

We assume that the base records about researchers created on the basis of publicly accessible sources or imported from other systems will be continually updated by respective researchers, or by repetitive imports of updated data. It is necessary to mention the continuous need for check of data quality. Desirable is manual data control, software validation and elimination of possible duplication.

The result is the bilingual up-to-date database of researchers containing professional profiles that will serve for searching experts according to scientific area, but also for presentation of scientific capacity in Slovakia and abroad. On the end of October 2013 the Register of Researchers contains more than 18000 personal records.

Registry of R&D Results

The registration of R&D results as an independent entity is novelty in the Slovak research information system. The results of projects have been published on web. The similar situation applies for registration of researcher's publication activity. The publications records inserted by researchers into the database would accessible to public.

In SK CRIS system all results will be published. It will not be only bibliographic data about publications, but also data about intellectual property objects (referred to as patents, trademarks, designs, utility models etc.) and data about other results of R&D activities (e.g. products and innovations).

For obtaining relevant data, the interconnection of SK CRIS with Central Registry of Publication

Activity was put into operation. The data collected as R&D results (project results, researchers' publications and results) inserted by organisation in applications for Certification to perform R&D has been also published. The results will be continuously updated.

Additional second level entities substantially supplement the core and result entities mentioned above. These are the calls for projects, finances, laboratory infrastructure (facilities and equipment; offered services), but also events, awards and different full text attachments, i.e. CV of researchers.

Grey SK CRIS

Typically Grey literature consists of technical research documentation. Although many different kinds of publication has been classified as grey. The material is usually not peer reviewed as is white literature. Some of them might be commercial documents in confidence and they could contain intellectual property of value to the organisation.

We can see two possible way how system CRIS could be important for collecting and providing of grey literature. The first way is based on the fact that CRIS system is considered as storage of grey objects (Jeffery,2007). Important part of CRIS system (and CERIF data format) entity research results could be considered as grey literature. The results of R&D are listed below (Asserson,2012) :.

(a) publications: journal / conference papers, technical papers, theses, dissertations, reports
(b) patents
(c) products: prototypes, fully engineered products
(d) results: data and its presentation / visualisation
(e) know-how and IPR: reports, procedure instructions
(f) education and training: documentation, courses
(g) publicity: press releases, product or organisational posters

The second way means that CRIS system is considered as software tool for administrate, provide and access contextual information about grey literature.

CRIS systems based on data format CERIF should support each type of result. As mentioned above, the SKCRIS system contains portal website where users can find contextual research information (projects, calls for proposals, researchers and their expertise, research institutes and their activity). The data are stored in SQL database using data format CERIF (CERIF,2012), containing metadata on research results including „grey" documents .

As yet a fulltext scientific repository solution is not operated, the SKCRIS substitutes some repository functionality beyond the common CRIS properties. The SKCRIS collects not only metadata but also full-texts of publications (figure 3), but we could not consider our system as CRIS-repository application .

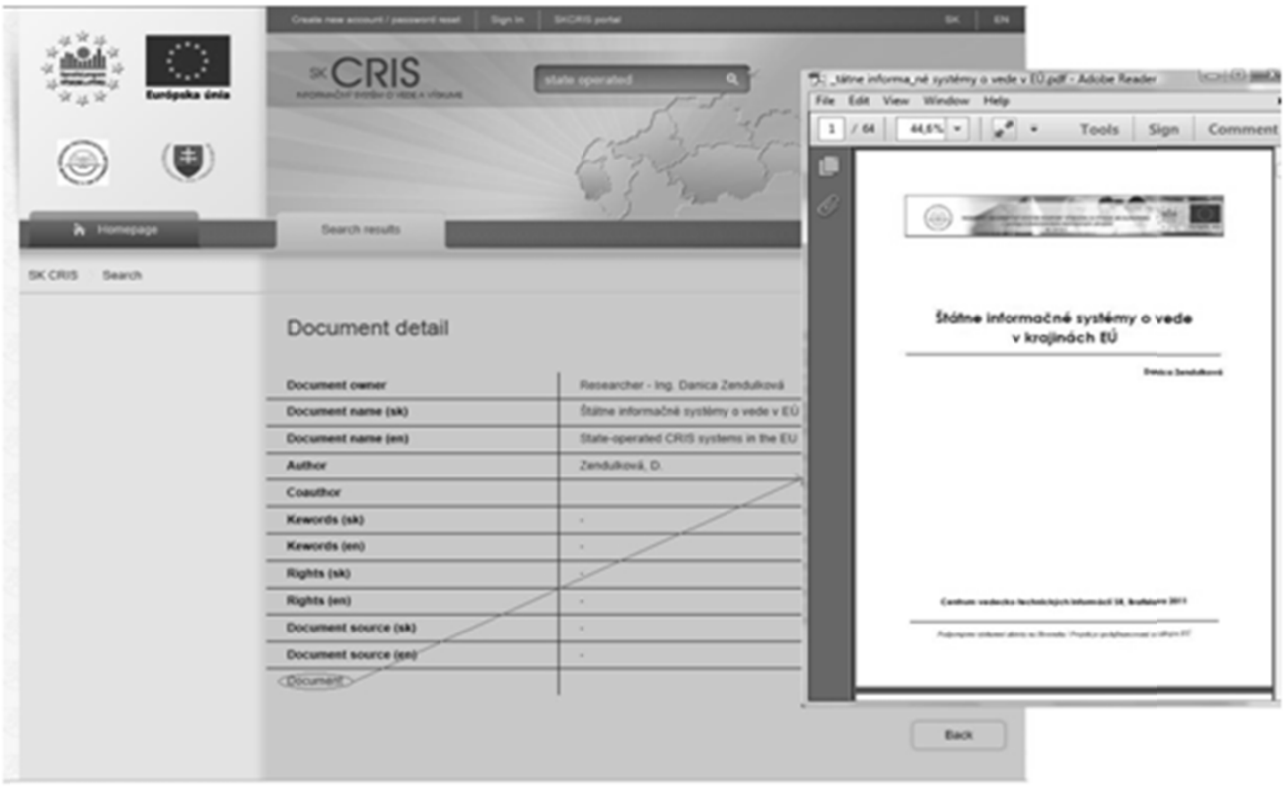
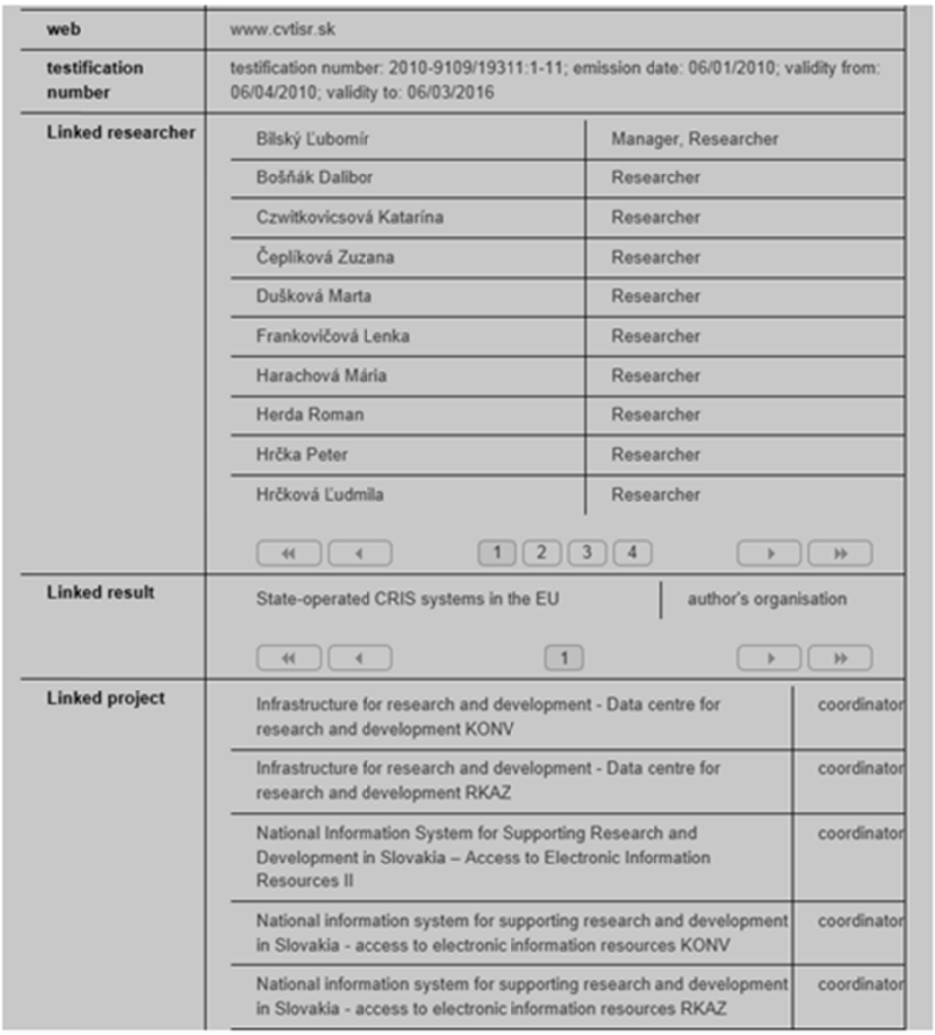


Figure 3 R&D Results: Document metadata and fulltext

In this consequence we need to mention also news items from research institutes' websites, published on Central Information Portal of Research, Development and Innovation.
The aim of providing the portal SK CRIS is to enable digital access to all informa tion concerning science, research and innovation funded from public resources.

Properties: how SKCRIS works
The functionality of the SK CRIS is user oriented. The most important functionality cornerstones are following:

1. *Search interface*. The search interface contains the possibility of use of different search techniques: simple, advanced, but also full text searching and faceted navigation.
2. *Bilingualism*. Bilingual Slovak and English user interface with selected data by CERIF requirements will access the information about Slovak science to users from different countries worldwide.
3. *Semantics.* The part of SK CRIS data model is also CERIF semantics, including different code lists and classification schemes. Their use results from Slovak legislation. The example of semantics and link entities between core and result entities is shown in Figure 4.
4. *Relationship*. The search results will present mutually related information. The main linking entities are Person – Organisation, Publication – Organisation and Project Organisation. The linked objects are connected by hypertextual link. It means it is possible from detail of organization on Figure 4 by click to open detail of linked researcher, project and result.



web	www.cvtisr.sk
testification number	testification number: 2010-9109/19311:1-11; emission date: 06/01/2010; validity from: 06/04/2010; validity to: 06/03/2016
Linked researcher	Bilský Ľubomír — Manager, Researcher Bošňák Dalibor — Researcher Czwitkovicsová Katarína — Researcher Čeplíková Zuzana — Researcher Dušková Marta — Researcher Frankovičová Lenka — Researcher Harachová Mária — Researcher Herda Roman — Researcher Hrčka Peter — Researcher Hrčková Ľudmila — Researcher « ‹ 1 2 3 4 › »
Linked result	State-operated CRIS systems in the EU — author's organisation « ‹ 1 › »
Linked project	Infrastructure for research and development - Data centre for research and development KONV — coordinator Infrastructure for research and development - Data centre for research and development RKAZ — coordinator National Information System for Supporting Research and Development in Slovakia – Access to Electronic Information Resources II — coordinator National information system for supporting research and development in Slovakia - access to electronic information resources KONV — coordinator National information system for supporting research and development in Slovakia - access to electronic information resources RKAZ — coordinator

Figure 4 Semantics and linking entities in SK CRIS

The specification of user interface functionality was inspired by the results of European CRIS systems desk research conducted as part of the SK CRIS analysis in 2009-2010 (Zendulková, 2011). The data already present in the system will not be required to be entered again. When online form opens, the relevant archived data will be loaded automatically. The system allows to track the history of data that can be changed (surnames, organisation names, etc.)

SK CRIS data model

Implementation of the data format CERIF will simplify and clarify the data entry, but also their presentation as mutually related objects. CERIF (Common European Research Information Format) is XML data format to support the management of Research Information. It is recommended by the European Commission as standard for automation in area of R&D information (CERIF, 2012).

The CERIF data format is based on data model, which allows metadata representation of research entities, their activities, interconnections and their output (results). The CERIF elements have defined the core structure, semantics and link entities and they are divided into hierarchical categories (Jörg, 2009 & 2011).

SK CRIS database is based on the standard CERIF 2008 v1.2. Databases Certification (applications for Certification to perform R&D) and Statistics (additional statistical survey of R&D Potential) contain data collected about each entity outside the scope of CERIF data model SK CRIS.

Besides core entities (Organisation, Project, Person) and result entities, SK CRIS contains the data about laboratory infrastructure. The next 2nd level CRIF entities will be added in future.

Instead of CERIF data model the SK CRIS registers statistical data (additional survey of R&D) and data of applicants for Certification to perform R&D. The online application form contains several indicators outside the data format CERIF scope. The workflow of administration process is part of SKCRIS functionality.

Data model SK CRIS (Valkovič, 2011) is schematically shown in Figure 5.

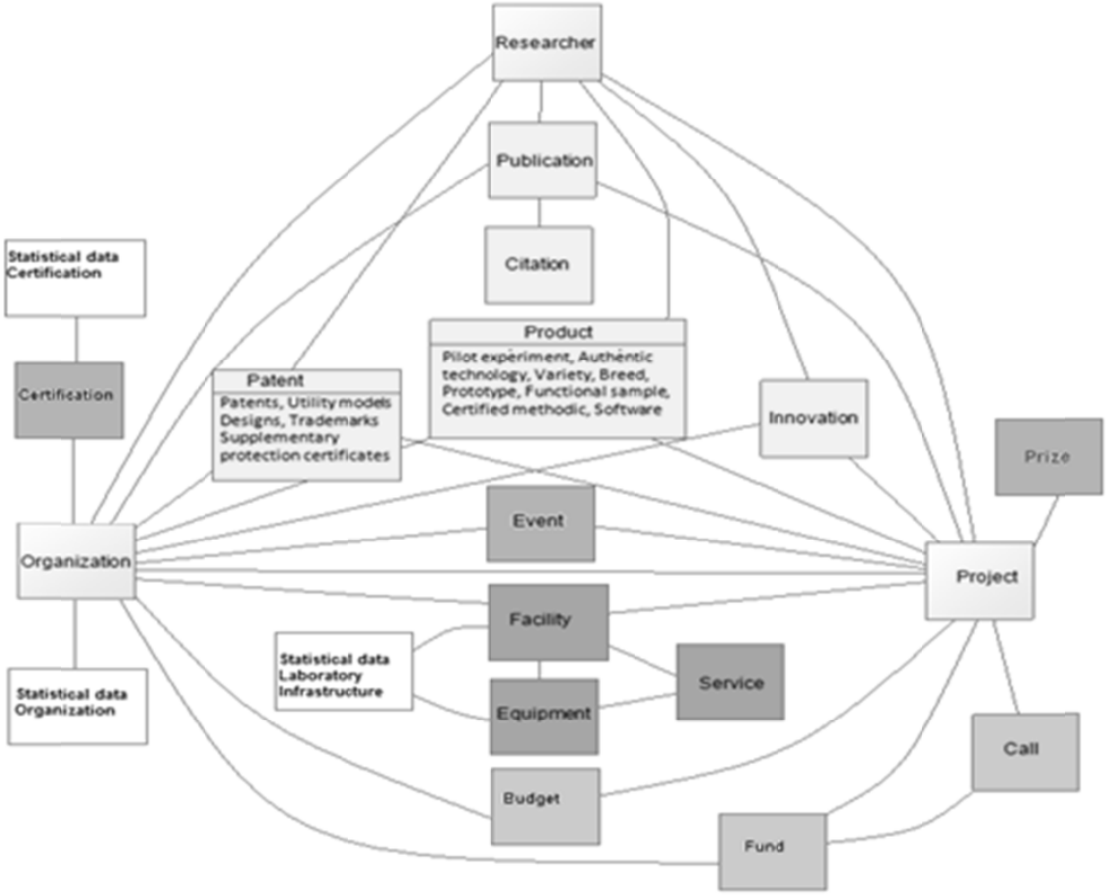


Figure 5 SKCRIS data model

Integration with other systems

The integration interface interconnecting SK CRIS with external systems containing relevant R&D information is one of important SKCRIS functionality. The following data have been imported by this interface:

- data about different categories of projects stored in grant agencies systems
- data about researchers from Registry of Slovak universities employees and Registry of employees of Slovak academy of Science
- Data about publications from Central registry of Publication Activity (CRPA) containing publication activity of Slovak universities.

The interconnection of SK CRIS with CRPA will allow to get data about the significant volume of R& D results and also the core data about researchers – publications' authors. The simplifying functionality for research results entry will be a part of SK CRIS. SKCRIS user – researcher receives during the entry of research results the verified list of publications from CRPA. He/she identifies relevant publications, which eliminates duplicity of the same data entries. For SK CRIS users, the data integrated from CRPA will complete the field of research results that have been continuously inserted by researchers to the SK CRIS.

In Figure 6 we can see the process how publication registered to CRPA by the university in UNIMARC data format (on top) is integrated into SK CRIS. The publication was imported into SK CRIS by researcher identifying CRPA publications where she is author. After inserting, the publication was marked as the result of registered project by project responsible person. The record about publication obtained the linkage to the project, responsible organisation and to all authors – researchers having personal records in SKCRIS. The figure also shows that four of five authors are employees of Comenius University - Jessenius Faculty of Medicine and one is not (by SK CRIS data) employee of university.

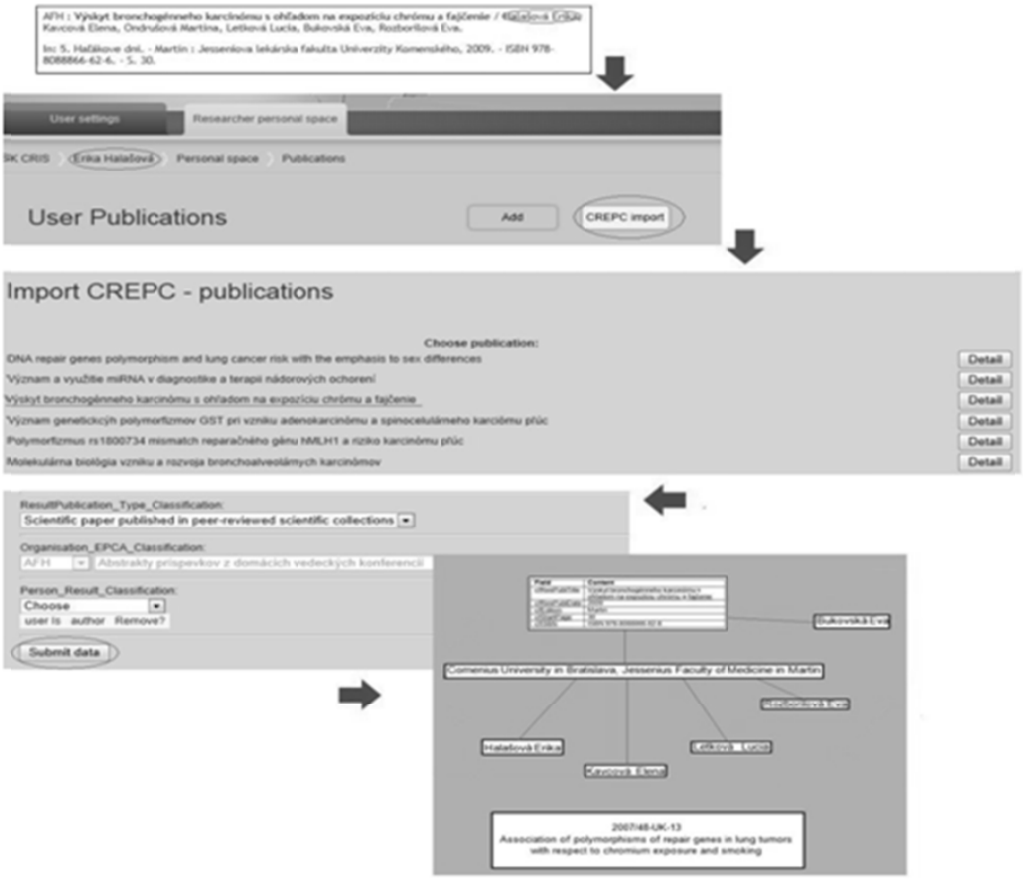


Figure 6 Connection between CRPA and SK CRIS

Benefits

First months of SKCRIS providing show several benefits for different target group of users.

1. The presentation of information and new knowledge and its interchange between scientists and researchers in electronic environment.
2. Accelerate the application of knowledge into the practice (technology transfer), information about research results for enterprises and business and their involvement in active utilisation of this knowledge in practice.
3. To obtain general support for science and research on national level. To provide information for decision makers about scientific and research activities and results and for state administration.
4. Popularisation of science and technology, including RTD results in a comprehensible way attractive for public. These activities should positively influence the perception of science by society.

Challenge for future

Despite the SKCRIS was put into operation some months ago, we are confronted with several challenges for future. Between possibilities of SK CRIS improvement, we consider following aims as crucial:

- Following CERIF development and implementation of actual CERIF data format version
- Integration of data from more external systems
- Use CERIF XML format for data exchange
- Interconnection with the fulltext repository when it will be prepared
- The process of data quality improvement is separate task. Responsibility of data correctness, completeness and timeliness is on the side of data creators, not on the side of SKCRIS administrator.
  But we identified some steps for reach better data quality:
  - To complete and to validate incoming data
  - Avoid duplicity of data coming from different systems
  - To use universal unique identifier for similar objects in integrated systems

The national CRIS system SKCRIS as result of the NISPEZ project funded from structural EU funds is only first step, groundwork in organization of research information workflow in Slovakia. Our goal is to build complete “house for research information” consisting of institutional CERIF based network of institutional CRIS system connected with fulltext scientific repository.

References

1. CERIF: the Common European Research Information Format. (2012) [Online]. [cit. 05.10.2013] Retrieved from: <http://cordis.europa.eu/cerif/ >
2. Jörg, B. (2009): CERIF: Spoločný formát informácií o výskume. Formálne kontextuálne vzťahy na prechod cez labyrint vedeckých poznatkov. Informačné systémy o vede v EÚ. Štandardizácia a kompatibilita. Zborník z medzinárodnej konferencie. CVTI SR, Bratislava 23. 4. 2009. [Online]. [cit. 05.10.2013] Retrieved from: <

   http://nispez.cvtisr.sk/userfiles/file/Aktivita%204.1/Konf2009%20prezentacie/Zbornik_konferencia_SK.pdf>
3. Valkovič, L. (2011): Detailná funkčná špecifikácia systému SK CRIS. Bratislava, 2011. [cit. 05.10.2013]
4. Zendulková, D. (2011): Štátne informačné systémy o vede v EÚ. CVTI SR: Bratislava, 2011.- 66 s. ISBN:978-80-89354-04-7 [Online]. [cit. 05.10.2013] Retrieved from: <http://nispez.cvtisr.sk/index.php?&menu=158 >
5. Turňa, Ján; Noge, Juraj; Zendulková, Danica: (2012)The system SK CRIS, scientific publications and theses – mirror of Slovak science. In: Jeffery, Keith G; Dvořák, Jan (eds.): E-Infrastructures for Research and Innovation: Linking Information Systems to Improve Scientific Knowledge Production: Proceedings of the 11th International Conference on Current Research Information Systems (June 6-9, 2012, Prague, Czech Republic). Pp. 211-220. ISBN 978-80-86742-33-5. [Online]. [cit. 05.10.2013] Retrieved from: <http://www.eurocris.org/Uploads/Web%20pages/CRIS%202012%20-%20Prague/CRIS2012_23_full_paper.pdf >
6. Asserson A. (2012): Grey Literature & CRIStin. In: Seminar on Providing Access to Grey Literature 2012*:* The 5th year of the seminar focused on storage and providing access to the grey literature, 23th October 2012 [online]. Praha: National Technical Library, 2012. [Online]. [cit. 05.10.2013] Retrieved from: http://nusl.techlib.cz/Sborniky. ISSN 1803-6015.
7. Jeffery, K. – Andersson, A. (2007): Hyperactive Grey Objects. In: The Grey Journal, 3/2007 (2): 159-163.
8. Jörg, B. (2011): CERIF tutorial CERIF 1.3 release. In: euroCRIS Membership Meeting November 2nd, 2011, Université de Lille 3 – Charles de Gaulle. [Online]. [cit. 05.10.2013] Retrieved from: <http://www.eurocris.org/Uploads/Web%20pages/members_meetings/201111%20-%20Lille,%20France/CERIF-Tutorial_Brigitte-Joerg_CERIF1.3.pdf >

Auditing Grey in a CRIS Environment

Keith G. Jeffery, Consultant, United Kingdom
Anne Asserson, University of Bergen, Norway

Abstract
We define grey as information that is not peer reviewed scholarly publications. In the CRIS (Current Research Information System) domain Grey includes not only non-peer reviewed publications (typically technical papers) but also performance art, art artifacts, design documents, models (e.g. for engineering or architecture) and – increasingly importantly – research datasets.
Recently national governments - inspired by the work of W3C on LOD (Linked Open Data in the context of the Semantic Web) – have made government information available for citizen rights (the information should be available to those who paid for its collection) and business leverage – the information is used by businesses for strategic planning and a growing ICT business sector provides applications using LOD for business advantage.
Government portals to publicly-funded data collections have been implemented (usually as data.gov.nn sites where nn = country) using 'flat' metadata standards such as DC (Dublin Core) and CKAN (Comprehensive Knowledge Archive Network). These metadata standards provide limited information and – worse – it is very difficult to assure integrity of the associated information object because there is no integrity checking mechanism. Much of the presented material is of a summary nature and based on more detailed research activities; rich metadata is available, both contextual (project, funding, persons, organisations, related white publications etc.) and detailed (schema level for software to interact with the information object). This rich metadata improves information object integrity in discovery and utilisation.
The ENGAGE project (http://www.engage-project.eu. portal at www.engagedata.eu) aims to enhance the metadata associated with the usual data.gov portals with the rich metadata available around the supporting research information objects in order to improve discovery and to ensure appropriate use of the information objects in context. A 3-layer model has been proposed with discovery metadata (DC, CKAN and several others) generated from the contextual layer (CERIF: Common European Research Information Format – an EU recommendation to Member States) which in turn points to the detailed metadata associated with each domain or even individual information object. The mappings from CERIF to/from each metadata format are done once and manually, thereafter conversion is automatic. This architecture combines the easy browsing / discovery in the semantic web/LOD world with the formal rigour of underlying rich metadata stored in information system with full integrity constraints.
This architecture allows confidence that any audit of the quality of the semantic web/LOD environment will be favourable because it is generated from the underlying integrity-rich environment and not just generated by manual input and linkage.

1. Introduction

In a world of ever increasing information, auditing (quality checking) is important. We argue that the vast majority of information that exists may be classified as grey (peer reviewed scholarly publications form a very small part of the corpus). Furthermore, in this paper we use information object to mean 'digital grey object' where the object may be information encoded in any form and any medium. In particular we are concerned with the metadata related to the information object and how it is used for auditing.

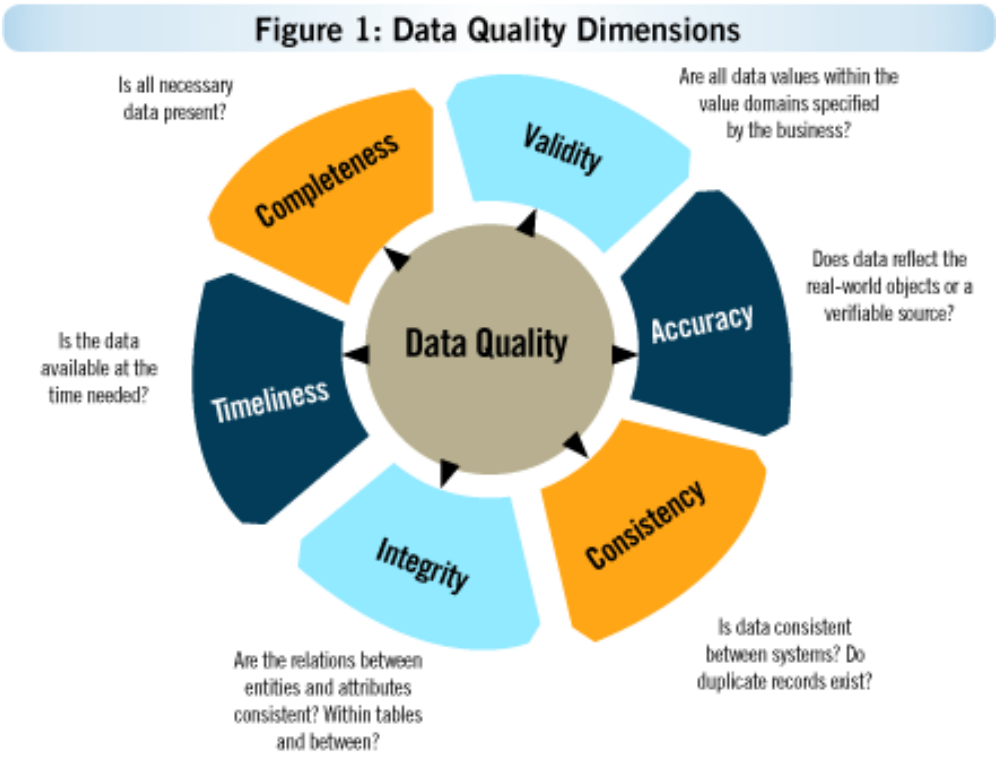
2. Reliable information

2.1 Introduction

Reliable information is essential for quality decision-making. We suggest that an audit of an information object (more specifically the metadata describing the object) to assure its reliability has three elements: quality, context and availability.

2.2 Quality

The dimensions of quality may be summarized as in Figure 7.



(With acknowledgements to FINETIK)

Figure 7: Dimensions of Data Quality

First and foremost data quality (using data in the wide sense of any kind of information presented using any medium[1]) depends on integrity: that is the verisimilitude with which the real world of interest is represented. This can only be assured by using a schema (which defines the representation of the data and the relationships of one data item to another) with associated constraints – expressions in logic that ensure the data values are within range or are from a predetermined list of values) and that dependencies (referential and functional) are respected.

The accuracy of the data (e.g. the accuracy of measurement using scientific instruments or observation) is important and – together with the precision with which the measurement was taken - determines whether the data may be used for the purpose intended. The real world has inconsistencies and observations are incomplete. It is essential that incompleteness is represented (by so-called null values) and that inconsistency is recorded with as much contextual information as possible (Section 0). Furthermore data values may change with time; it is important to know the temporal validity (for which period of time it is representing the real world) of the data item. Much of the above can be assured by independent validation – commonly through peer review – and / or by quality rating by the interested community (as done for example concerning products available through amazon.com or hotel booking systems).

2.3 Context

Context describes the environment in which the information was collected and may be used. It consists of related entities that give confidence that the information object of interest is understood in that environment and that its quality and appropriateness may be assessed. We assert that a solution – (CERIF) – exists already. CERIF has already been in use widely in 42 countries and is an EU Recommendation to Member States. CERIF is maintained, developed and promoted by (euroCRIS) at the request of the European Commission. The CERIF datamodel is already quite well-known in the Grey Literature Community (Jeffery and Asserson 2005) but the overall model is reproduced here (Figure 2) to illustrate the entities that are recorded together with their relationships thus giving the context of the research.

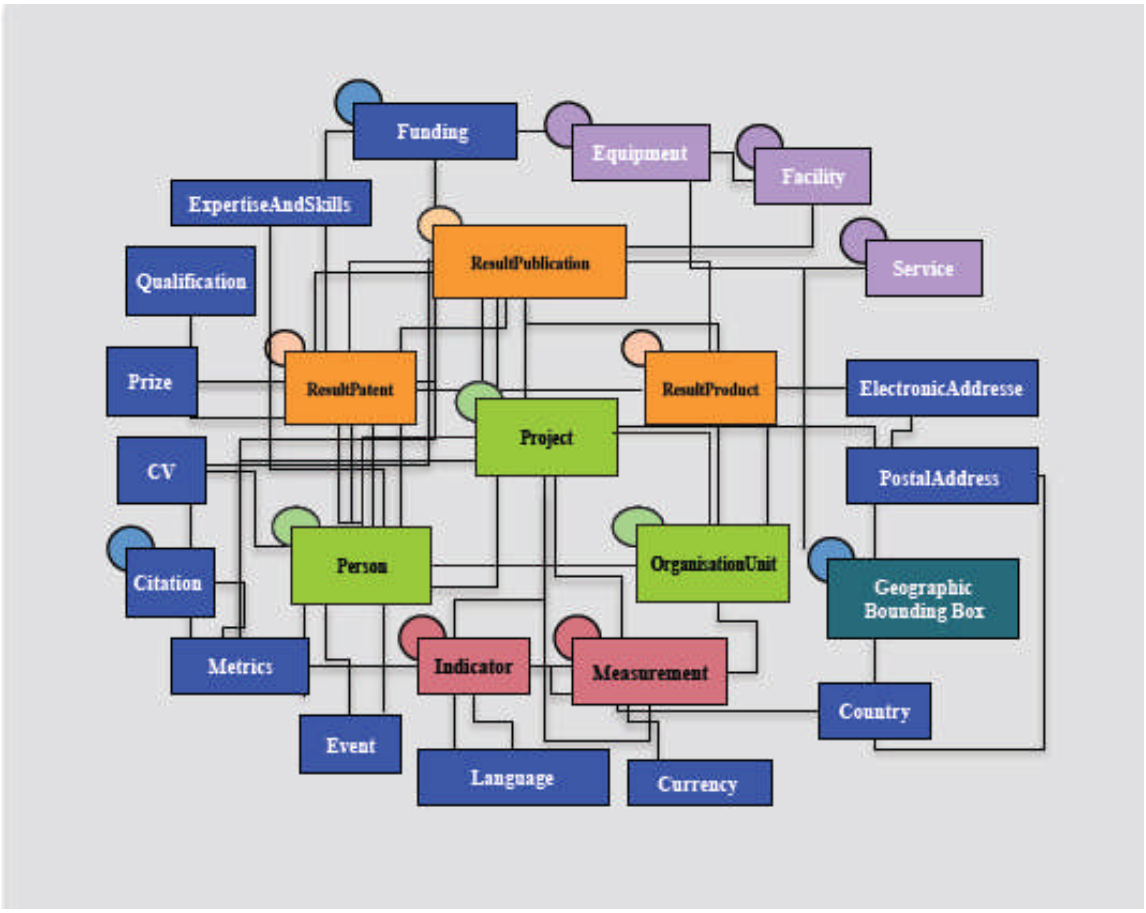


Figure 8: The CERIF Datamodel

2.4 Availability

Data (in the wider sense of information) is not reliable – and its reliability cannot be audited - if it is unavailable through lack of persistence or lack of accessibility.

2.4.1 Persistence

The key requirements for persistence are (1) media migration to ensure the information is readable; (2) a declared syntax and semantics to ensure the information is understandable; (3) preservation of related software to process the information: this is required because the software may well encapsulate information about the information object and may be the only method of accessing the object. Unfortunately software systems are ephemeral so the specification of the software and the processing environment (operating system, compilers) may be the best mechanism.

2.4.2 Accessibility

Hidden information may be valuable for some purposes (particularly for commercial exploitation of the IP (Intellectual property) encapsulated in grey material) but for scholarly discourse access should be open and toll-free at the point of use. In this way the grey material – output of research or related activity - can be evaluated for quality by the peer community or for applicability to a problem or opportunity by others (innovators, entrepreneurs, educators). Accessibility is controlled (explicitly or not) by rights and licenses and these must be recorded.

2.5 Conclusion

The reliability of information depends critically on one technology: metadata. The metadata itself must be of as high or higher quality than the information being described and thus must conform to the aspects of reliability outlined above. The authors have long urged the grey community to adopt richer metadata and specifically CERIF (Jeffery 1999), (Jeffery and Asserson 2007), (Jeffery and Asserson 2008), (Jeffery and Asserson 2010).

3 Open data

3.1 Introduction

Recently there has been much interest in open data. This has been caused by a conjunction of the open movement (for scholarly publications, software) and technology developed under the aegis of W3C (World Wide Web Consortium). A major motivation has been government policy resting on two premises: (1) that open transparent government means that data collected using public funds should be available to the public; (2) that making this government data openly available will stimulate new business opportunities – both for providers of software services and users of those services for business benefit. Such information (the majority is documents rather than structured datasets) is usually classified as grey since objects are not subject to formal (scientific) peer-review although one could argue that parliamentary scrutiny of e.g. national financial data is a kind of peer review in assuring quality.

The technologies involved are LOD (Linked Open Data) and SW (Semantic Web). The former provides a mechanism using RDF (Resource Description Framework) of making a syntactic relationship between two things (e.g. X is related to Y) and the later allows the naming of that relationship to be expressed in a defined vocabulary with the usual ontological relationships (X is father of Y). These triples (subject-relationship-object) are usually encoded in XML.

Sometimes raw data is encoded in RDF. However, more usually RDF is used for the metadata describing such a dataset. DC (Dublin Core) has evolved from its early textual representation through qualification of elements and XML representation to a RDF encoding. However, the vast majority of examples of DC metadata are in textual or qualified form. CKAN (the modestly named Comprehensive Knowledge Archive Network) has a metadata format that is based on DC but extended. It is encoded in RDF.

3.2 The Problems

Most portals provide only a clickable list of available information objects; clicking on the name provides a screen of metadata including the URL of the object itself. In fact most available objects are not structured data at all, they are unprocessable documents in pdf or limitedly processable spreadsheets in xls format although a recent move by W3C intends to encourage the non-proprietary csv (comma separated value) format[2]. Very little metadata exists for most objects and that which does exist is 'flat' and poor; that is it has the properties of a library catalogue card rather than exposing the richness of the structures of the metadata (Jeffery 1999). Many objects are of summary or aggregated form and do not make reference to the underlying detailed – often publicly-funded research – objects.

3.3 A problem is an opportunity

The problems provide an opportunity: to support the unreliable open data represented by metadata in DC or CKAN with reliable open data from the domain of research represented by rich metadata namely CERIF.

4 Engage

The ENGAGE Project (ENGAGE) assists users in accessing PSI (Public Sector Information) available through open government data sites but also underpins the objects there with access to underlying information generated during publicly-funded research (datasets and related publications – which may be classified as white). Naturally the information at present is dominantly in the domain of SSH (Social Sciences and Humanities) but the concept is extensible across all domains. ENGAGE provides not only user access but also facilities for loading datasets and a social networking approach to commenting on datasets and annotating them. Furthermore it offers a community forum for requests or discussion.

A key aspect of the architecture is the 3-layer metadata model (Figure 9): this provides separation of (1) metadata used for discovery (simple, flat) from (2) metadata used to understand how the object(s) might be used in context (CERIF) from (3) the detailed and specific metadata of the object itself – usually a metadata format common to only a few objects in a limited domain. The discovery metadata is generated from the contextual metadata to ensure consistency – and convertors are available from and to CERIF for (DC), (CKAN), (eGMS), (INSPIRE), (ADMS), (DDI), (SDMX) and other mappings are being done all the time. The contextual metadata points to the detailed metadata associated with the object, used dominantly to connect software to the object for processing.

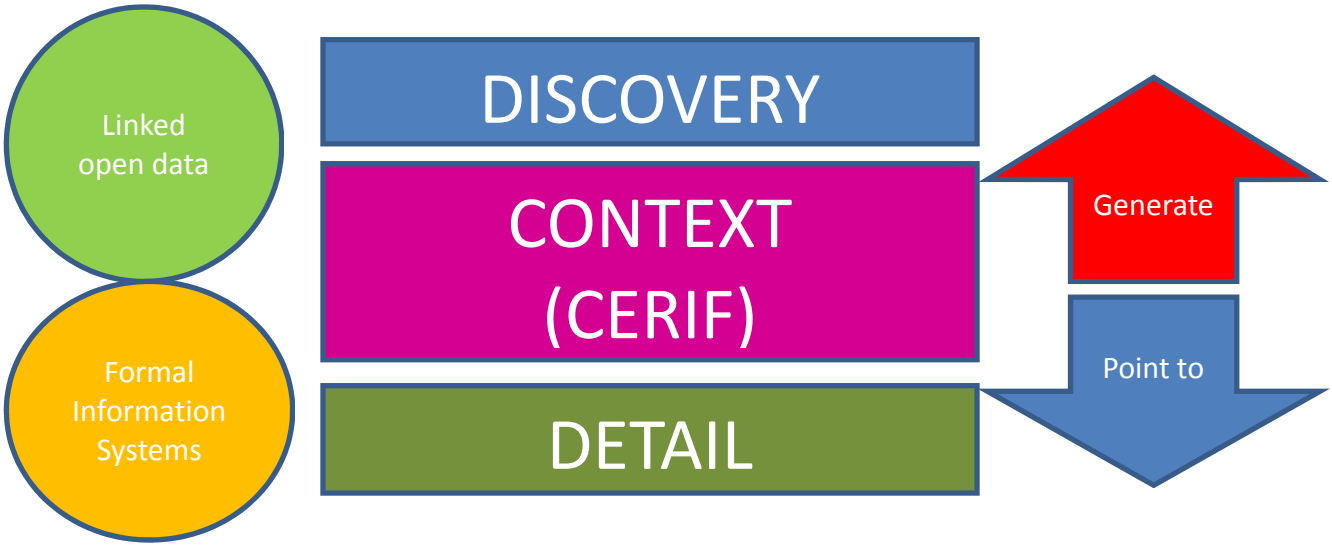


Figure 9: ENGAGE Metadata Architecture

The ENGAGE metadata architecture describes the data model within a developing rich e-research architectural environment providing models (described by metadata) for not only data (in the sense of information) but also users, processes and resources.

5 Conclusion

euroCRIS has defined the metadata architecture of an ambitious project. The purpose is to underpin summary open government information objects (described by limited metadata) with research objects (described by rich metadata) from which the former were derived. The integrating characteristic is CERIF. This provides formality and assurance. euroCRIS is also providing the metadata interconvertors so CERIF acts as the superset 'master' format generating the others and acting as a 'translator' between them. The enhanced metadata provides the required quality for audit to assure the quality of the information object.

Acknowledgements

The authors acknowledge the contributions of colleagues in the ENGAGE project part-supported by EC Contract 283700.

References

(ADMS) https://dvcs.w3.org/hg/gld/raw-file/default/adms/index.html (Retrieved 20131028)

(CERIF) www.eurocris.org/cerif (Retrieved 20131028)

(CKAN) http://ckan.org/ (Retrieved 20131028)

(DC) http://dublincore.org/ (Retrieved 20131028)

(DDI) http://www.ddialliance.org/ (Retrieved 20131028)

(eGMS) http://www.esd.org.uk/standards/egms/ (Retrieved 20131028)

(ENGAGE) http://www.engagedata.eu/ (Retrieved 20131028)

(euroCRIS) www.eurocris.org (Retrieved 20131028)

(INSPIRE) http://inspire.jrc.ec.europa.eu/index.cfm/pageid/101 (Retrieved 20131028)

(Jeffery 1999) Jeffery, K G: 'An Architecture for Grey Literature in a R and D Context' Proceedings GL'99 (Grey Literature) Conference Washington DC October 1999 http://www.konbib.nl/greynet/frame4.htm

(Jeffery and Asserson 2005) K G Jeffery, A G S Asserson 'Grey in the R&D Process'; Proceedings Grey Literature 7 Conference, Nancy, December 2005; TextRelease; ISBN 90-77484-06-X ISSN 1386-2316

(Jeffery and Asserson 2007) Keith G Jeffery, Anne Asserson: 'Greyscape' Opening Paper in Proceedings Grey Literature 9 Conference Antwerp (GL9) 10-11 December 2007 pp9-14; Textrelease, Amsterdam; ISSN 1386-2316

(Jeffery and Asserson 2008) Keith G Jeffery, Anne Asserson: 'INTEREST' Proceedings Grey Literature Conference Amsterdam 8-9 December 2008 in Tenth International Conference on Grey Literature : Designing the Grey Grid for Information Society, 8-9 Dec. 2008, Science Park Amsterdam, The Netherlands ed. by Dominic J. Farace and Jerry Frantzen ; GreyNet, Grey Literature Network Service. - Amsterdam : TextRelease, February 2009. GL-Conference series, ISSN 1386-2316; No. 10. - ISBN 978-90-77484-11-1.

(Jeffery and Asserson 2010) Keith G Jeffery, Anne Asserson 'GL Transparency: Through a Glass, Clearly' Proceedings Twelfth International Conference on Grey Literature, Prague December 2010 Ed Dominic Farace and Jerry Frantzen 95-100, TextRelease Amsterdam Series ISSN 1386-2316; no 12; ISBN 978-90-77484-16-6

(SDMX) http://sdmx.org/ (Retrieved 20131028)

[1] Formally information is structured data in context

[2] http://www.w3.org/2013/05/lcsv-charter.html

Grey Literature in European Commission Projects

Sara Goggi and Gabriella Pardelli, Istituto di Linguistica Computazionale;
Silvia Giannini and Stefania Biagioni, Istituto di Scienza e Tecnologie dell'Informazione, CNR, Italy

Abstract
The survey is focused on the documentation produced by the European Commission (EC) projects involved in the Framework Programme for Research and Technological Development (hereafter FP7) and managed by the Italian National Research Council (hereafter CNR). In particular, the Grey Literature (GL) available on CORDIS[1] and European Projects websites was analysed. In order to verify how it is managed and whether it is compliant with EC recommendations, some categories were introduced to identify, measure and evaluate the usability and availability of projects production. Data was obtained from a sample of European projects websites.

1. Introduction

The latest recommendations issued by the European Commission go towards the revision of their policy on dissemination and preservation of scientific information in order to promote the access to the results of the community-funded research by especially implementing the open access policy within 'Horizon 2020', the EC Framework Programme for Research and Innovation, 2014-2020 (*European Commission, 2012*). In the proposal for a Regulation of the European Parliament and of the Council laying down the rules for the participation and dissemination in 'Horizon 2020 we can read:

> *"The rules regarding intellectual property, exploitation and dissemination have been modelled on the widely acknowledged Seventh Framework Programme provisions with further improvements and clarifications. Specific new emphasis has been put on open access to research publications and an opening was made for experiments with open access to other results. The enlarged scope and new forms of funding as well as the need for flexibility in this area of the rules has been taken into account by the possibility to lay down additional or specific provisions where appropriate. Access rights for the European Union, and in the field of security research also for Member States, have been foreseen"* (*European Commission, COM 2011*).

The websites of the EC projects represent an essential vehicle for both the acquisition and the diffusion of documentation and could also become an important resource within a European infrastructure able to overcome the disconnected and scattered nature of their content in order to optimize their reuse. The EC considers the structure and the contents of the websites generally *[suffering] from a contextual and structural neglect (European Commission, Best practices)*.

In order to improve the documentation management, EC published some guidelines that proved to be a very useful tool for optimizing and handling information on the dedicated portals of the community-funded projects.
The guidelines consist of general and specific recommendations for the websites and for each element of the websites structure. The general recommendations focus on the importance of using social media as well as webmaster tools and virtual meeting facilities (as web streaming) and of adopting a "eu" domain. Among the general recommendations are also identified: the use of sustainable systems aimed at the retrieval of sites by the search engines, the update of the sites contents and the use of keywords in metadata and contents page. We concentrated on the specific recommendations providing the "Best Practice" for the structure of websites, identifying the components and providing the explanation of each concept and content.
The CNR has an important role in the development of the European Union programme: it considers the programmes of the European Union as a priority and strongly encourages the participation of researchers. The CNR has participated in European projects since the beginning and since 2007 joined the FP7 (2007-2013). The CNR also "*welcomes the architecture designed by the European Commission for Horizon 2020 (H2020) and considers its structure innovative and suitable to promote European science, technology and innovation"* (CNR, Position on Horizon2020).
Starting from this context, the aim of this survey is to obtain information about the current practices carried out by the partners of the European projects as a means to improve the projects visibility, the usability and the access to their production.

2. Materials and Methods

We measured and estimated the projects documentation and its level of compliance with EC recommendations. CORDIS was the starting point of our explorationand its contents have allowed us to create a sort of "identity card" for each project.

In order to get a specific range of data, we selected a subset of European projects from the CORDIS website using the following query criteria:

i) Search for >*Consiglio Nazionale delle Ricerche*

ii) Refinements >Programme: *FP7* **>**Coordinator Country>*Italy* **>**Participant Country>*Italy*.

We got 226 projects matching the search criteria: this group of projects was the core of the analysis.

Data was collected between June and August 2013.

The information was extracted from

- ✓ CORDIS directory contents
- ✓ Project site structure and contents

We gathered the

- ✓ Projects production

and analysed the documents by

- ✓ a)Type of products
- ✓ b) Format
- ✓ c) Availability/access

The objective of detecting all types of documents has been achieved through different steps: first we extracted the information provided by the CORDIS directory: Acronym, Title, Grant, Start/End, Status, CORDIS URL, CORDIS keywords, CORDIS objective, CORDIS related links, Total cost, Area, Contract type and verified the existence of the project websites. This checking allowed us to understand immediately how many projects have their own website even if CORDIS does not always provide the link nor precise information on project sites. In this case we made specific searches via Google to verify the site existence.

Secondly, we checked the projects websites in order to gather information about the websites structure and the projects production.

The third step was the categorization of the documents and the defining of the dataset. Finally, there has been the analysis of the results.

3. Explorative phase: CORDIS website analysis

The following tables and figures show some results: the CNR is the Coordinator in 62% of the cases and in 38% is participant.

Projects subset		
Status	N.	%
accepted	14	6,2
completed	99	43,8
execution	113	50
Total	**226**	

Projects subset		
Projects websites	N.	%
yes	140	62
no	86	38
Projects websites by status		
accepted	1	0,8
completed	60	42,8
execution	79	56,4

Tab. 1. – Projects websites

Compared to the total, half of the projects appear to be in execution while 43.8% has been completed. A small group of projects is only accepted. About 62% of the projects have their own website: the category of accepted projects shows the presence of only one website compared to 13 which have not developed it yet. Some projects have not created a proper website and are using only a few spaces within the sites of the partners institutions. We included in the group of projects with a website only those which make available some information and/or documentation. This type of website affects the total number of projects for the 8,5%. With respect to this sample, the presence of the projects website does not seem closely related to the project in execution.

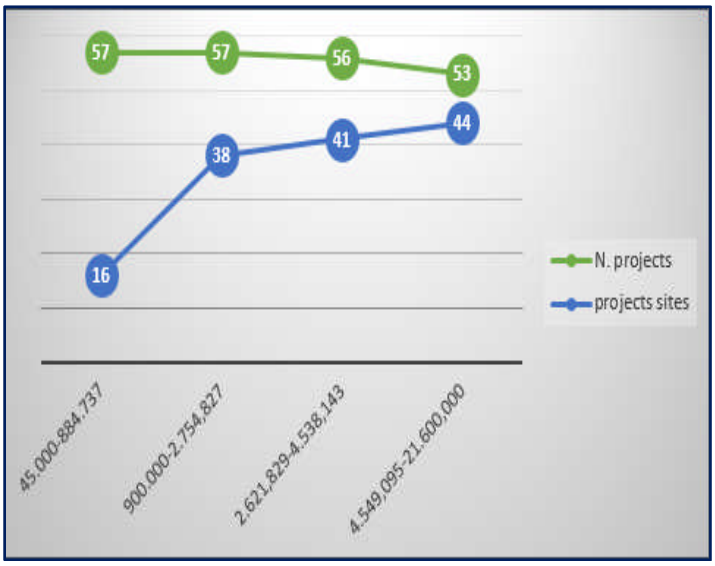


Fig. 1 – Projects websites/projects total cost ranges

It is quite understandable that among the projects recently activated only one has already its own website, but it is not so clear why the other projects with websites are only 56% of the total.
The presence of the project website seems related to the increasing financing; we checked the total cost of projects: when the cost increases, the presence of projects websites increases too (Fig. 1). Probably, the growing financial availability lets to devote more professional management of the sites (e.g.: webmasters, librarians and so on) and a greater number of resources moves towards the Web. Even a higher number of partners requires large and more organized structures and more funding that can help the implementation and the improvement of the project in the network.
The analysis of products managed in CORDIS indicates only two types: publications and reports. CORDIS contains documentation even in the absence of the project website remaining, therefore, the only source of documentation for 24 projects.
There is a large group of projects that does not have any kind of documentation. Then we have the highest percentage of research articles published in journals (Fig. 2), which CORDIS draws directly from OpenAIRE (Open Access Infrastructure for Research in Europe).

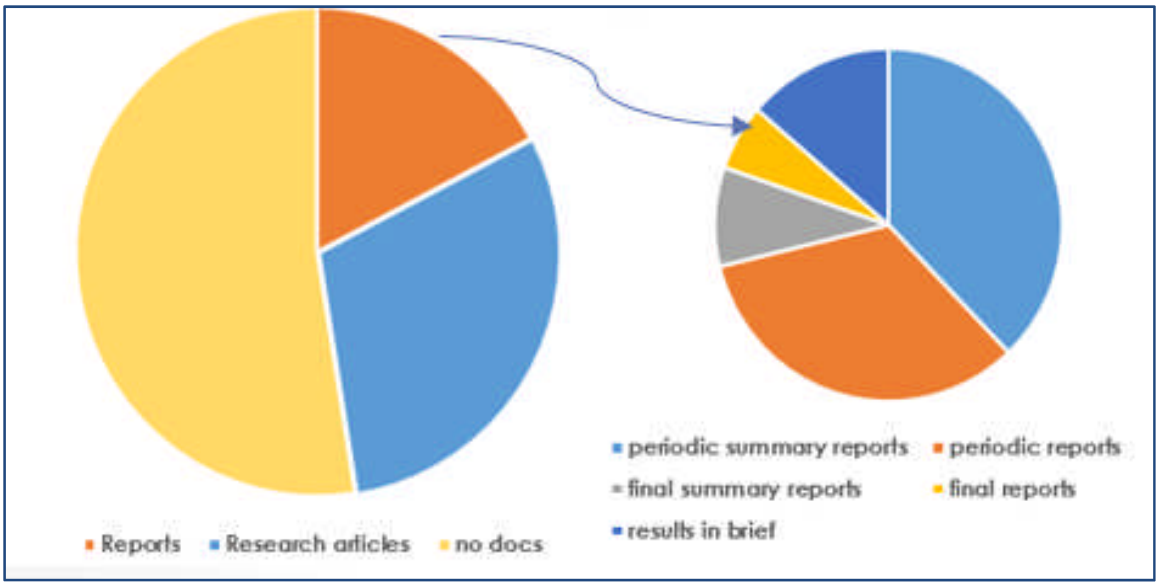


Fig. 2 – CORDIS documentation and reports types

4. Results: projects websites analysis

4.1 Categories

Our next step was the analysis of the projects websites. The collection of contents allowed us to identify the types of documents produced and to gather them into categories, types and subtypes. The nine sections proposed by the guidelines with respect to the structure of the website (Overview - Consortium - Management structure - Case Study, Deliverables, Publications, Events, Media center, Glossary) and the logical association of related materials were the two main criteria followed in the material categorization.
We identified 13 categories, composed by different types and subtypes (Tab. 2).
We treated 7 of them as individual categories, because they can be associated into a single type (Tab. 3).

Categories	**n. Projects**	**%**
Courses material	12	8,5
Deliverables & Reports	74	52,8
eNewsletters	31	22,1
Image material	48	34,2
Informative material	138	98,5
Media press	17	12,1
Projects meetings material	50	35,7
Normative documents	2	1,4
Patents	1	0,7
Research/scientific articles	109	77,9
Technical documents	96	68,5
Tertiary documents	20	14,2
Theses	4	2,8

Tab. 2 – Categories

Some categories represent types of materials not specified in the guidelines. We split or merged some categories in order to give more emphasis to the informative content (for example we split *Media centre* in *eNewsletters*, *Image material*, *Media press*, *Normative documents* but merged *Case studies* in *Technical documents* and *Events* in *Informative material*). For the same reason, we preferred to provide a generic level of information about the type *Events*, because their presentation is varied and confusing. Events are almost never divided between past, present or future events and only few of them shows a calendar, as suggested by the guidelines. Very often we found together generic news, not necessarily related to events. Our choice was the creation of the category *Projects Meetings material* whereby we isolated the information concerning the meetings of the projects (workshops, seminars or conferences) thus dividing them from the disorder of the type Event.

Categories	**n. Projects**	**%**
Courses material	12	8,5
eNewsletter	31	22,1
Image material	48	34,2
Projects meetings material	50	35,7
Media press	17	12,1
Normative documents	2	1,4
Patent	1	0,7
Theses	4	2,8

Tab. 3 – Individual categories

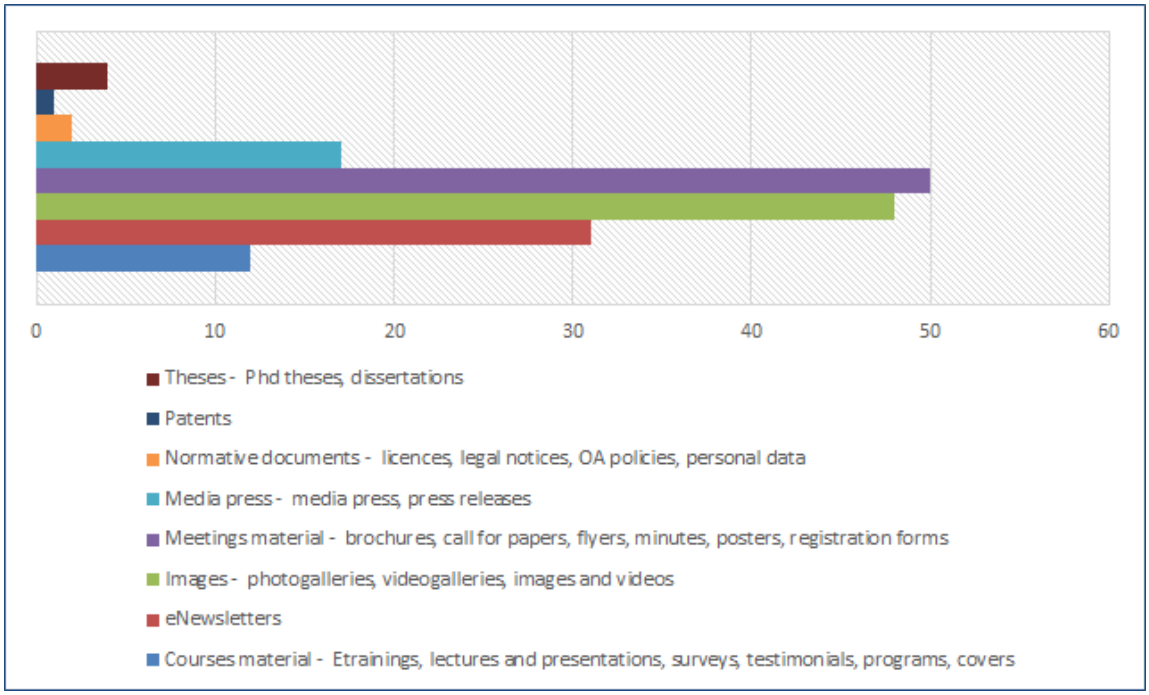


Fig. 3 – Individual categories (subtypes)

Figure 3 shows the types and the subtypes composing the individual categories. *Media press* could flow in the category Information material but it was isolated as contains both newspaper articles devoted to the project as well as articles published by the project itself on national or international press. We considered more interesting to provide this category as separate information. Twelve projects have established courses, training and / or summer school and have chosen to disseminate and make available the related documentation. We checked also several projects that present *Theses* conceived and realized within the project itself. The category *Image material* collects all those projects that have chosen to publish on their own sites various kinds of videos or images not closely related with a specific document type. In these cases, the object itself has been treated as a type. Conversely, when the object is placed as a medium used in support of a given type, we considered it as *pdf*, *html* and *other* formats. The category e*Newsletter* lists the projects that publish their own newsletter in electronic format and make available the full text. Two projects present their normative documents such as licenses, policies and policies for data processing.

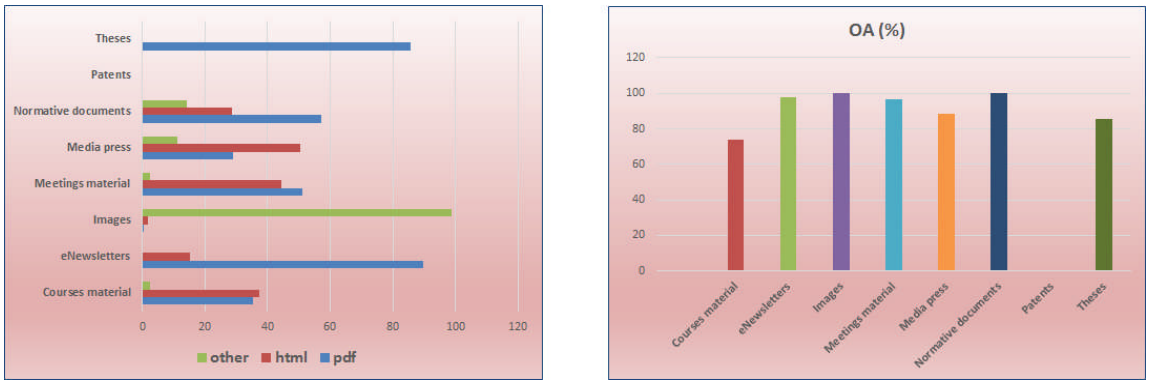


Fig. 4 – Individual categories (formats/accessibility)

The prevalent formats are *html* and *pdf*. The exception is represented the category *Images* which, by its nature, is almost 100% composed of multimedia material fully available. In this case, the formats are jpeg, tiff, png and other video formats. All formats differing from *html* and *pdf* are gathered into the item *other* (Fig. 4).

Talking about access to the full text, we can say that several types make it 100% available (or anyway at very high percentage), so almost all of these types can be open accessed. We could not find any news about the only project that shows a patent because it is only announced but not described, nor viewable.

4.2 Informative material

After the Homepage, the first sections of the guidelines are Project overview and Consortium. Both areas have been included in the category *Informative material*. We can see that almost all projects with a website provide at least a minimum of descriptive information. The contents of these sections are the biggest of the whole category. In Fig. 5 we can see, in fact, the different types and their incidence in the general category. Here we have also materials not mentioned in the guidelines such as Announcements, News, Related links etc.

The objectives of the project are indicated by the guidelines as the basic information for the website design because the treatment of this type of information is highly recommended. They are present almost always although not always introduced by this term. Frequently the objectives are introduced by terms such as: about, overview, abstract, description, mission, but the contents express the objectives of the project.

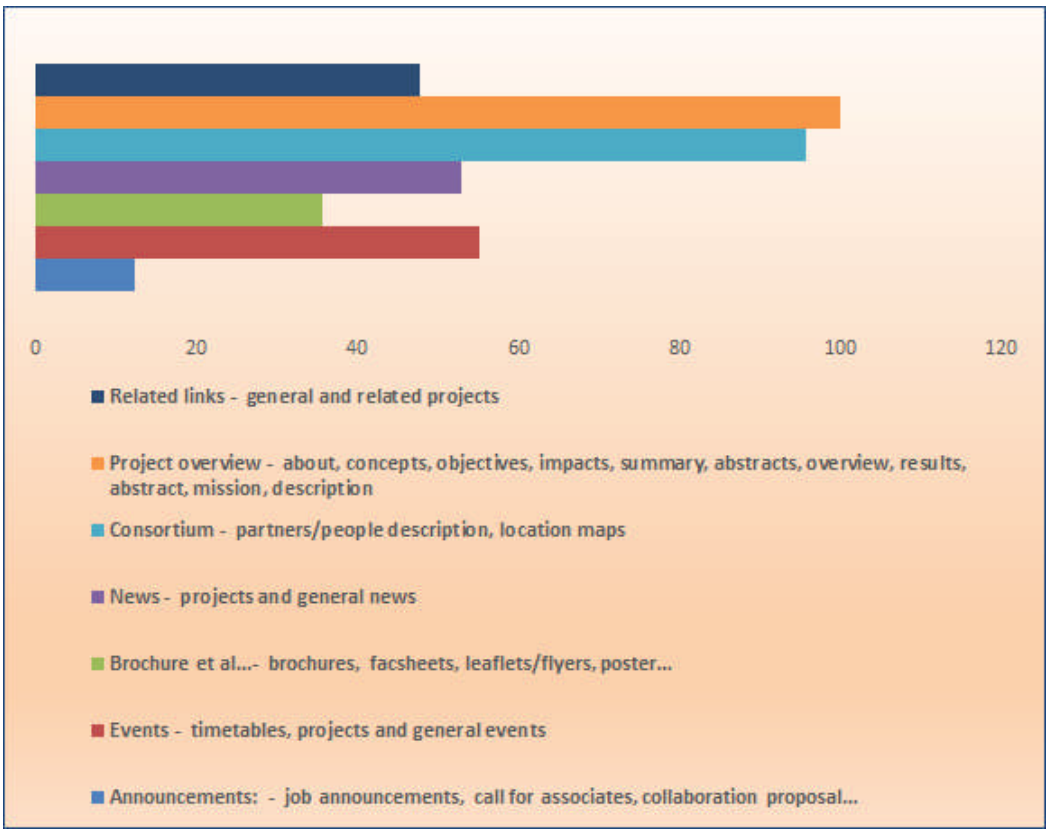


Fig. 5 – Informative material (types/subtypes**)**

The analysis shows that all projects provide their own overview and that only two out of 140 projects do not give any kind of information on the Consortium.

However, the extent data varies from project to project: most of them propose very little descriptive information while others produce a lot of documentation on scope, objectives, and information on the operational context. The analysis of the completeness of descriptive information shows that about 68% of the projects does not seem to provide detailed data about their operating context. This documentation is almost never provided in a downloadable pdf format, or conforms to a specific pattern. It is given freely and usable as a *html* page (Fig. 6).

Also the presentation of the Consortium varies from project to project. The 95% of the projects indicates the number of participants and their nationality while a slightly lower percentage (87%) is for the contact information that is omitted in some cases.

Significantly less is the presence of location maps, that is the geographic distribution of participants in graphical form. Three projects give a specific presentation about women involved in the project, accompanied by their profile and the description of their activities. In general this is introduced by the term *Gender aspects*.

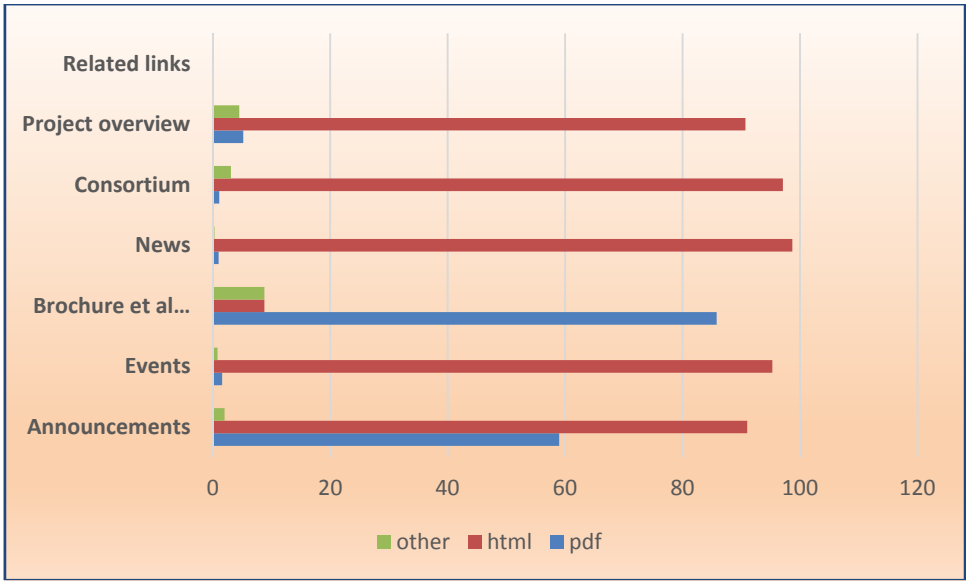


Fig. 6 – Informative material (formats)

Most of the descriptive pages about the partners present the name of the institution, *the* location, the logo and the URL, leaving to the website of the single participating institution the task of presenting the single partners. Several projects, however, perform the description of partners on their own. In this case, the projects propose a detailed description of the activities carried out by the partners and their institution, the subject area covered, the mission and their expertise.
The type Related links indicates the percentage of projects with links to related projects. The format has not been quantified because we have only links to other websites, which were not subject of detection. Data is interesting in order to see how many projects offer the link to similar projects from their websites: compared to the total of the projects with a website, they are about 44%.

4.3 Technical documents

The category *Technical documents* includes some types stated in the sections 3.4 and 3.5 of the guidelines and other types that, in our opinion, are conceptually similar to them, such as roadmaps, prototypes or strategy, just to name a few. The types Research plan, Scientific methodology and Management structures are the largest (Fig. 7).
The projects provide mostly descriptive documentation about their management bodies and about the handling of the project. Approximately, 47% of projects also provide a description of the work packages and/or their management structure. Twentythree projects provide the graphical representation of their research plans, requested from EC guidelines. Also in this case the detail of information is quite varied and the lacking descriptions are numerous. The management structure is presented in different forms and even the terminology used by the projects is not completely uniform and not always explanatory as well as the relationships between government bodies, activities and methodology.

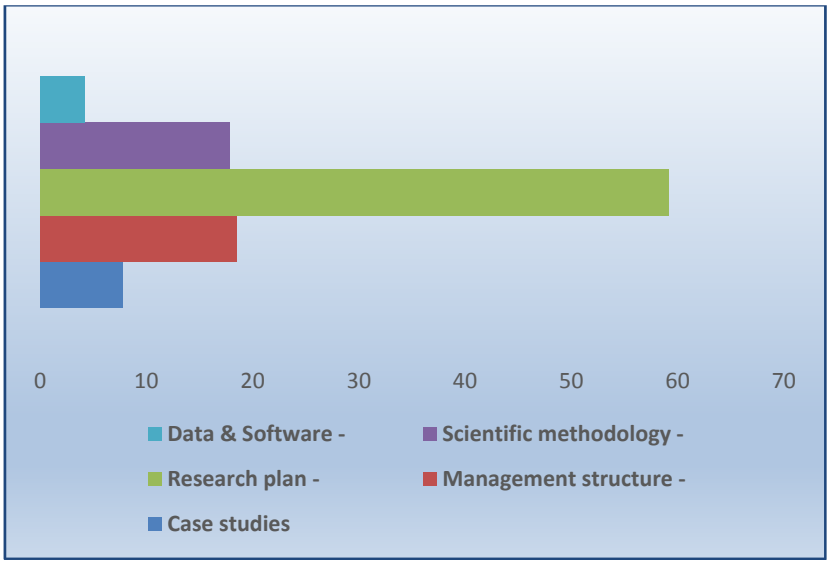


Fig. 7 – Technical documentation (types/subtypes)

A few projects present *Case studies*. It is true that the guidelines recommend the description of *Case Studies* only whether they are particularly relevant. Moreover, in the guidelines, the concept of *Case studies* appears closely related to geographic studies. It is also expressly recommended that the description of *Case studies* should be accompanied by a geographical representation of the same, by the creation of global or European maps. If we give a quick look at the research themes of the projects presenting *Case studies*, we see that mostly belong to the Space and Environment areas. Comparing the total number of projects with a website and the projects that describe *Case studies*, the percentage is very low and their impact within the category *Technical documents* is around 13%. By restricting the calculation to the total percentage of the projects contained in these areas, it appears that 65% of the projects have *Case studies* and 10% of the latter accompanies the description with geographic maps. As for the other types, most of the technical documents are made available through descriptive *html* pages. Some projects use simple links to the studied sites, or to images and photographs of specific subjects. Among *Case studies*, only one project makes available the description of the site.

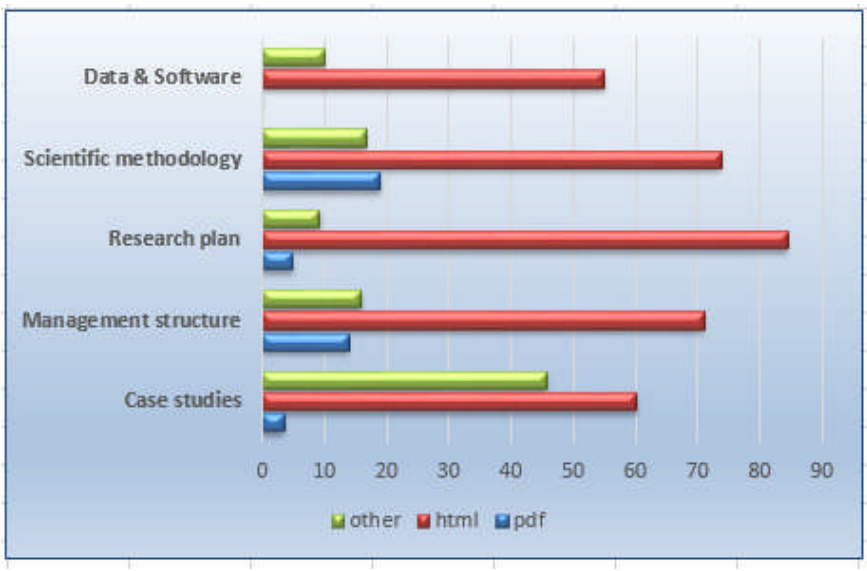


Fig. 8 – Technical documents (formats*)*

All documents converged in this category are usable in open access.

4.4 Deliverables & Reports

The guidelines give recommendations only about public *Deliverables*. These should be listed in tables and provided with specific descriptive metadata that can be used for the search and retrieval. It is also recommended the use of *pdf* format for downloading. The Research articles should be visible in specific tables (separate from those of the *Deliverables*) identified and described through the common metadata and, if open access, should be downloaded directly from the project website.

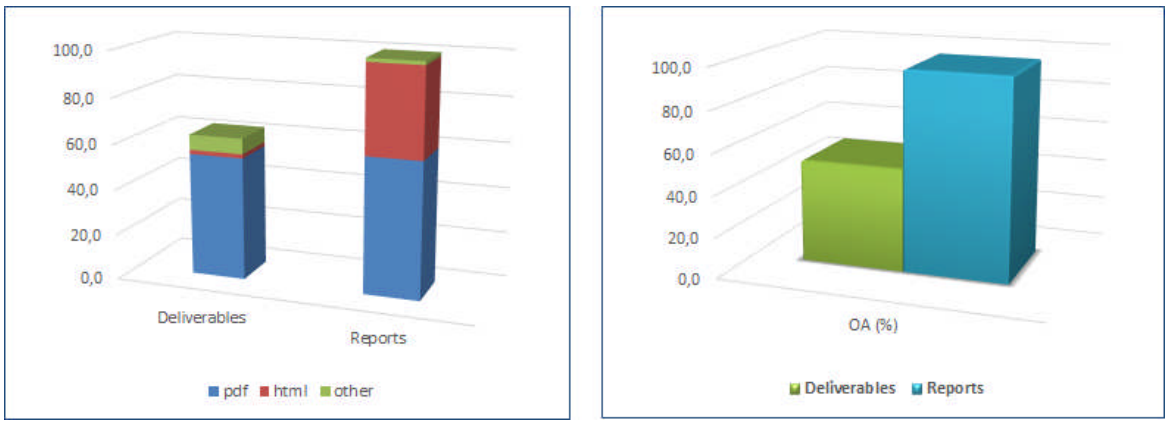


Fig. 9 – Deliverables & Reports (formats/accessibility)

The guidelines do not give indications about *Reports* but almost 60 % of the projects make visible some *Reports* on their sites (Fig. 9).

The subtypes are the same already met in CORDIS: periodic or final reports of activity and results in brief. However, there are also a few working papers and/or technical reports. The *Reports* are freely available in 97% of the cases.
The *Deliverables* are considered the main product of the projects activities. They allow checking the progress of work.
The 54% of the projects present their *Deliverables*. It seems that 46% of projects does not show them but many projects maintain a restricted area so it is possible that all *Deliverables* are deposited inside, without distinction between public and private. It is difficult to understand, because the projects that explicitly state to the partners that *Deliverables* are stored in the reserved area are rare, most of them do not provide any information about the contents of this area. The presentation of the *Deliverables* varies: some projects give news in the form of a list, calling them directly from the page menu; others put them in different voices such as documents or publications; and others give the information together with the work packages.

4.5 Research/scientific articles

The presentation of the *Research/scientific articles* category varies from project to project and is presented using different terms. It is rarely available in tabular form, as suggested by the guidelines, and even more rarely is structured in form of searchable databases. The form of presentation is often a list with links to the publisher's site for the access to the full text. Compared to the *Deliverables*, we noted a bigger compliance to the guidelines about the use of descriptive metadata: generally the bibliography contains the basic elements useful for the resource identification. We verified the there is no full correspondence between research articles appearing on the projects sites and on OpenAIRE: the most frequent case is that in which the project shows a greater number of references than OpenAIRE.

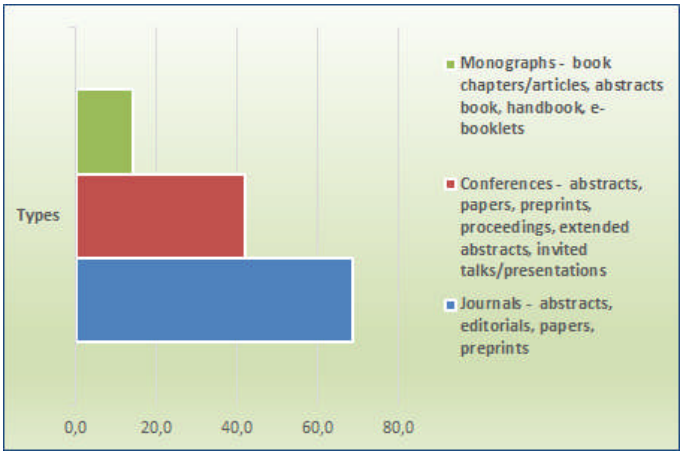


Fig. 10 – Research/scientific articles (types/subtypes)

The highest percentage of the group is related to research articles published in journals, followed by publications in conferences and then in monographs. The subtype shows that the category includes papers, abstracts, editorials and proceedings. OpenAIRE detects articles published in journals.
It is rare that OpenAIRE shows articles published in Proceedings.

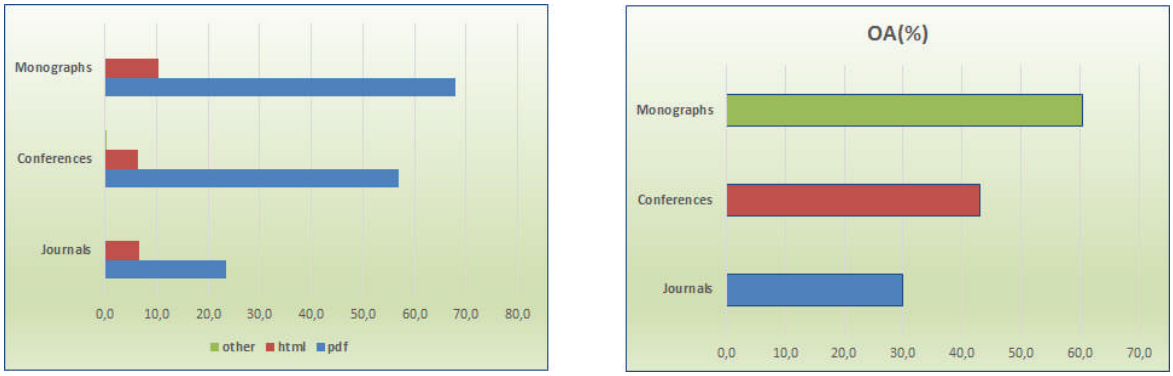


Fig. 11 – Research/scientific articles (formats/accessibility)

The observation of formats and accessibility on figure eleven shows that the more intensive use of the *pdf* and the availability of full-text are attributable to monographs, followed by Conferences and Journals. This is due to the presence of the subtypes e-booklets and handbooks in Monographs and the subtypes abstracts, extended-abstracts and invited talks / presented in Conferences. Most of the books and papers published in journals, book chapters and proceedings are available, but not accessible in full text. The availability of papers in open access is attributable to the presence of the document in Arxiv, in institutional repositories and/or in open access journals.

4.6 Tertiary documents

The category includes all types considered as secondary sources in this context. The projects that show this kind of products are numerically very low. Both the catalogues and the bibliographies are structured in the form of *html* pages (Fig. 12).

The category includes the *Glossaries*, treated in the last section of the guidelines that recommend to projects the production of a glossary of terms and abbreviations and of a Wiki. We checked four projects responding to this recommendation. They make available on their websites the explanation of terms and expressions of their own field of study. The nature of the glossary depends on the scope of the research project. Only one project has structured the glossary as a searchable database.

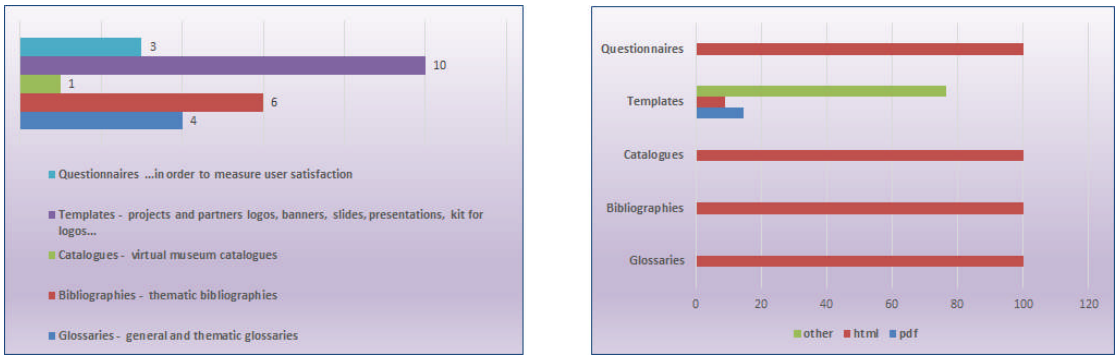


Fig. 12 – Tertiary documents (types/subtypes/formats)

Five projects propose a Wiki, more or less structured and articulated. In general they are organized as all wikies or shared websites and give more advices about the project. Sometimes they create special pages and recommendations about technical details and analysis of the possible approaches to the topics of study.

The questionnaires are offered by three projects with the aim of mapping the intended audience, to check the skills and to acquire information. They also aspire to measure the degree of user satisfaction about the contents of the project and their presentation on the websites.

5. Concluding Remarks

We encountered several difficulties in the identification of the contents offered by the projects. Sometimes the websites are very articulated and the organization of the material is not immediately understandable. In order to provide clear and explanatory results, we collected the documents in some categories. The *Grey Literature Vocabulary* (Pejšová P. et al., 2012), was our starting point and our guide in categorizing: more or less all types of materials encountered in the data collection found their collocation within the categories proposed by the vocabulary.

The analysis revealed that the documents produced by projects and available from their websites are numerous and more varied than the items included in the EC guidelines and that there are important differences in the way the contents are disseminated. Resuming, we can say that there is a basic compliance to EC guidelines in terms of content but a minor compliance towards the use of *pdf* format, compared to the increased use of *html*, and to the observation of specific templates and models in the preparation of contents. Almost all categories are available in open access apart from the deliverables and the research articles published in journals.

Surely, the European Commission is very interested in the divulgation of projects results and shows a remarkable attention to the communication plan and to the involvement of user groups. Our society is increasingly technological and needy of documentary sources visible and shareable. The attention given by the EC to open access formats and to the retrieval of information resources of different domains, seems to be a confirmation of the will that knowledge becomes a common heritage, not only accessible by an elite of specialists. A contribution in this sense could come from the more extensive use of tools such as Forums, Blogs and Social Networks. Our analysis reveals that 14% of the projects has already

created a Forum and/or a Blog and that 17% is included in the circuit of the most common Social Networks such as Facebook, LinkedIn and Twitter.
In our opinion this is a good result. The use of these tools can provide alternative routes and give more visibility to the realities of the projects and the supported activities. The sharing of collaborative platforms may facilitate the activities, the insights and the comparisons and act as a major vehicle of information at all levels.
Another important topic of reflection is related to the mode of content presentation. We detected a wide heterogeneity on the choice of terms which label the documentation: documentation itself, publications, library, results, dissemination are only some of the terms used by the projects to collect and make available their documents. This heterogeneity primarily involves the Grey Literature that is the majority of the documentation produced. Grey Literature remains too often in the shadow. OpenAIRE checks only *Research/scientific articles* produced by the projects, but any type of grey documents is gathered.
The scholars involved in European projects need to be updated on the state of the art of the research conducted by other projects: current, completed and/or belonging to different and past frameworks. They need of all types of documents. So, it becomes very important the creation of infrastructures that merge, describe and make easily available all documentation, grey literature included.
Equally helpful may be the adoption of taxonomic criteria for extracting from the sites informative redundancy. So, the resources could be labelled with more strictness.
Harmonize the terminology of the categories, especially of GL categories within each project, through the use of guidelines careful to the taxonomic criteria could be of great help to the visibility of the documentation.
We detected that also CORDIS, in most of the cases, is the only source for establishing links between projects belonging to the same thematic areas but these areas are very general and do not specify the topics studied by the projects. A significant example in this sense is given by thematic areas such as Scientific Research, Information Systems or Coordination and Cooperation, which do not appear to reach a significant level of detail to guide the search for users. The projects websites do not offer links to related projects by research areas or specific subjects. Only one project publishes on its website the keywords of the studied topics and only three projects present Tag Clouds, that could be used to visually represent the keywords used in a website and the related fields of research projects.
We do not believe it is our task to make further recommendations and/or suggest their extension to improve the visibility and accessibility of the vast documentation produced by the projects. We only provide a map of the results and some thoughts on what emerged from the analysis but we hope that European Commission can work at the creation of infrastructures careful to the relationship between the content itself and accompanied by adequate tools to search and retrieve information. We retain that the use of controlled vocabularies, a greater emphasis on conceptual content and the use of descriptive and semantic metadata that can label and describe the documents in a timely manner could be very important. Structures designed this way could collect and make more accessible the whole documentation.

6. Future plans

In the future, we would like to extend this study to other realities. For example, it could be very interesting to analyse projects coordinated by partners from other countries and/or by different institutions, such as the universities or the industries. The study of the terminology adopted by the EC inside the different frameworks and in the projects websites it could be equally interesting. Moreover, we found that among the information provided by CORDIS, particularly stimulating are those related to the involved research/thematic areas of the projects. The areas cover the majority (approximately 83%) of the total of areas involved in the seventh Framework Programme. Most of the projects pertains to the following areas: *People* - devoted to the "human potential" in Europe; *ICT* - devoted to Information and Communication Technologies; *NMP* - devoted to Nanosciences and Nanotechnologies; *IDEAS -ERC* - devoted to " frontier researches". Among these four areas, *ICT* and *NMP* have an equally significant number of websites. Conversely, in *People* - that is the area of greatest concentration - only 21% of the projects shows to have their own website. For some areas, which count a relatively small number of projects, the number of websites is coincident with one of these projects, for which the hedge is total. Topics and thematic areas in relation *1:1* are the majority, but some topics are shared between different thematic areas.
It was not possible, here, to study them in depth, therefore we just gave a few indications. For this reason, a deeper analysis of the relationships between projects, websites and research/thematic areas could be another possible directions for further development and future studies.

References and Webliography[2]

CORDIS - Community Research and Development Information Service
<http://www.cordis.europa.eu/projects/home_it.*html*>

EUROPEAN COMMISSION. *Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions. Towards better access to scientific information: Boosting the benefits of public investments in research*. Brussels, 17.7.2012 COM (2012) 401 final.
<http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2012:0401:FIN:EN:*PDF*>

EUROPEAN COMMISSION. *Proposal for a Regulation of the European Parliament and of the Council laying down the rules for the participation and dissemination in 'Horizon 2020 – the Framework Programme for Research and Innovation (2014-2020)*'. Brussels, 30.11.2011 COM(2011) 810 final 2011/0399 (COD).
<http://www.europarl.europa.eu/document/activities/cont/201203/20120316ATT41112/20120316ATT41112EN.*pdf*>

EUROPEAN COMMISSION, Directorate General, Research & Innovation, Directorate I , Environment. *EU Project Websites – Best Practice Guidelines*.
<http://ec.europa.eu/research/environment/*pdf*/project_website_guidelines_en.*pdf*>

CONSIGLIO NAZIONALE DELLE RICERCHE, *Position of the National Research Council of Italy (CNR) on HORIZON2020*.
<http://www.attivitaeuropee.cnr.it/sites/default/files/antonella/POSITION%20PAPER%20on%20H%202020%2016052012_1.*pdf*n>

EUROPEAN COMMISSION - Research & Innovation. *Horizon 2020*
<http://ec.europa.eu/research/horizon2020/index_en.cfm?pg=h2020-documents>

EUROPEAN RESEARCH COUNCIL. *Supporting top researchers from anywhere in the world*.
<http://erc.europa.eu/media-and-events>

PEJŠOVÁ P., MYNARS J., SIMANDLOVÁ T.. *A linked-data vocabulary of grey literature document types: Version 1.0*. Thirteen International Conference on Grey Literature (GL13), Washington D.C. (US), 5-6 December 2011, TextRelease, Amsterdam, 2012, (GL Conference Series, ISSN 1386-2316; no 13).
<http://code.google.com/p/grey-literature-typology/>
<http://invenio.nusl.cz/record/81435/files/idr-146_2.*pdf*>

POCH M. *PANACEA Platform documentation index*. WP3. Internal report 3.0, 2012.
<http://www.panacea-lr.eu/system/tutorials/PANACEA-Platform_documentation_index_v3.0.*pdf*>

SCHÖPFEL J., 2010. *Access to the European Grey Literature*, in PetraPejšová ed., Grey Literature Repository, RadimBačuvčík, VeRBuM,2010.
<http://nrgl.techlib.cz/images/Book.*pdf*#page=20>
<http://www.panacea-lr.eu/en/info-for-researchers/data-sets/>
<http://libguides.gwumc.edu/content.php?pid=44946&sid=4110551>
<http://cordis.europa.eu/search/index.cfm?fuseaction=search.simple>
<ftp://ftp.cordis.europa.eu/pub/fp7/ssh/docs/synopsis_ssh_projects_fp7_low_definition_with_web_links.*pdf*>
<ftp://ftp.cordis.europa.eu/pub/fp7/ssh/docs/ssh-project-synopsis-fp6_en.*pdf*>
<http://cordis.europa.eu/fp6/>
<http://cordis.europa.eu/fp7/ict/programme/home_en.*html*>
<http://cordis.europa.eu/search/index.cfm?fuseaction=prog.document&PG_RCN=13628900>
<http://bibliontology.com/examples>

[1] Community Research and Development Information Service, is an information space devoted to European research and development (R&D) activities and technology transfer. http://cordis.europa.eu/guidance/helpdesk/faq_en.html

[2] URL last access: November 21 2013

Back to Grey:
Disclosure and Concealment of Electronic Theses and Dissertations

Joachim Schöpfel, Charles de Gaulle University Lille 3 / ANRT – GERiiCO
Hélène Prost, INIST-CNRS, France

Abstract
The open access principle requires that scientific information be made widely and readily available to society. Defined in 2003 as a "comprehensive source of human knowledge and cultural heritage that has been approved by the scientific community", open access implies that content be openly accessible and this needs the active commitment of each and every individual producer of scientific knowledge. Yet, in spite of the growing success of the open access initiative, a significant part of scientific and technical information remains unavailable on the web or circulates with restrictions. Even in institutional repositories (IRs) created to provide access to the scientific output of an academic institution, more or less important sectors of the scientific production are missing. This is because of lack of awareness, embargo, deposit of metadata without full text, confidential content etc. This problem concerns in particular electronic theses and dissertations (ETDs) that are disseminated with different status – some are freely available, others are under embargo, confidential, restricted to campus access (encrypted or not) or not available at all. While other papers may be available through alternative channels (journals, monographs etc.), ETDs most often are not. Our paper describes a new and unexpected effect of the development of digital libraries and open access, as a paradoxical practice of hiding information from the scientific community and society, while partly sharing it with a restricted population (campus). The study builds on a review of recent papers on ETDs in IRs and evaluates the availability of ETDs in a small panel of European and American academic IRs and networks. It provides empirical evidence on the reality of restricted access and proposes a model of independent variables affecting decisions on embargo and on-campus access, together with a table of different degrees of (non) open access to ETDs in IRs.

The context

Scientific grey literature stands for intellectual works not controlled by commercial publishers, of sufficient quality to be collected and preserved, but often difficult to obtain. The difficulty of acquisition and collection building was one of the main characteristics of grey literature in the past. The Web changed the situation. Dissemination of scientific information and access to the full text of all kinds of documents became easy. Concerning grey literature, the Web was considered as a solution and at the same time, as the final destination. The idea was simple and convincing: increasing availability and accessibility would change the nature of grey literature and, in the end, make it disappear. Grey would turn into white (Artus 2003). This belief was strongly supported by the success of the movement towards open access to scientific information (Suber 2012). The open access principle requires that scientific information be made widely and readily available to society (Willinsky 2005). Defined in 2003 as a "comprehensive source of human knowledge and cultural heritage that has been approved by the scientific community"[I], open access implies that content be openly accessible and this needs the active commitment of each and every individual producer of scientific knowledge.

The reality is different. Not only the definition of grey literature can (and will) survive the Web and open access, (Schöpfel 2010) but also contrary to all expectations and hopes, the Web sometimes increases barriers to scientific information. In spite of the growing success of the open access initiative, a significant part of scientific and technical information remains unavailable on the web or circulates with restrictions. Even in institutional repositories created to provide access to the scientific output of academic organizations, more or less important sectors of the scientific production are missing. The reasons are multiple: lack of awareness, embargoes, deposit of metadata without full text, confidential content and privacy concerns etc. This problem concerns in particular electronic theses and dissertations (ETDs). Many are freely available, but others are under embargo or confidential, restricted to campus access (encrypted or not) or not available at all. While some professionals and scholars are increasingly concerned about the situation (Owen et al. 2009), others welcome the protection of copyright (Hawkins et al. 2013).

Our paper provides empirical evidence on restricted access to American and European ETDs, reviews some published explanations, and then makes a proposal of a conceptual model of independent variables affecting decisions on embargo and on-campus access, together with a table of different degrees of (non) open access to ETDs in institutional repositories (IRs).

The paper builds on a study conducted in Lille between January and April 2013 (Schöpfel & Prost 2013) and contributes to a French-German survey on ETD embargoes carried out by the Institute for Science Networking at the University of Oldenburg and the University of Lille 3.

Evidence

A small but growing number of empirical studies on ETDs reveal figures on access restriction. A survey conducted in winter 2013 produced complementary figures from France, Europe and the United States. Table 1 presents figures from fifteen institutions and service providers, with the surveyed number of theses, the percentage of documents without access restriction, and the part of documents under embargo or restricted to campus-only access.

Institution	Sample	% open access	% restricted access	% on campus	% embargo
ProQuest (US)	500000+	95%	5%	n/a	5%
Texas (Austin)	11539	92%	8%	8%	n/a
PUC Rio de Janeiro (Brazil)	2787	89%	11%	n/a	11%
West Virginia	4600+	85%	15%	15%	0.3%
Florida State	3709	84%	16%	n/a	n/a
ABES (France)	10631	80%	20%	20%	n/a
Lille 1 (France)	833	79%	21%	15,5%	5,5%
Auckland (Australia)	3088	72%	28%	28%	n/a
Lorraine (France)	52	71%	29%	29%	n/a
Maryland	2050	68%	32%	n/a	32%
Valenciennes (France)	35	63%	37%	31%	6%
Liège (Belgium)	191	57%	43%	33%	10%
Virginia Tech	20386	54%	46%	n/a	46%
Amherst	n/a	48%	52%	32%	20%
Lille 3 (France)	124	40%	60%	n/a	n/a

Table 1. Empirical evidence on restricted access to electronic theses and dissertations (ETDs)

Taken together, about 10% of these roughly 550,000 electronic theses are not freely available on the Internet. Without the ProQuest figures, this part with limited access rises to 26%, ranging from 10% to more than 50%. 17% are embargoed for six months to two years or longer while the other 9% can only be accessed on-campus. This panel may not be representative and the results should be interpreted with caution. Nevertheless, they point out that the problem is not limited to a country or region but concerns all institutions with ETDs infrastructures and IRs. Some examples[li]:

At Amherst College, Massachusetts, 32% of PhD theses cannot be accessed from outside of the campus and 20% are under embargo for at least six months (Banach 2011).

At the University of Maryland, 68% ETDs are available without any restrictions. The other theses are under embargo, 21% up to one year and 11% from one to six years (Owen et al. 2009).

ProQuest Dissertation Publishing conducted in 2012 a study on ten years embargo trends (2000-2010) in the ProQuest Dissertations and Theses (PQDT) database. The surveyed corpus of 500,000+ print and electronic theses contained about 25,000 embargoed items (5%). Most of the embargoes are short-term embargoes, for six months to five years, but a small part of theses are under permanent (long term) embargo.

In Brazil, Pavani & Mazzeto (2009) describe access restriction for 11% ETDs on the campus of the Pontifícia Universidade Católica at Rio de Janeiro. About 21% of these files are under embargo for five years or longer.

The University of Liège (Belgium) document server indicates 191 PhD theses for 2012. 108 are freely available on their IR called ORBi (57%). For 33%, the access is limited to the campus; the remaining 10% are embargoed for a non-specified delay.

Since 2006, French universities have progressively switched from the traditional handling of print PhD theses to the new infrastructure of ETDs called STAR, linked to a national gateway “Theses.fr” run by ABES at Montpellier[lii]. From 2006 to 2012, the STAR system processed 10,631 ETDs. 8,737 theses were available on the web without any restrictions (80%) while access to the other 1,894 theses was limited to on-campus availability (20%). STAR does not provide information about embargoes.

Another example from France: from 2008 to 2011, the University of Lille 1 processed 833 ETDs in Science and Technology. Nearly 80% are in open access on their IR. 15% are available on the campus only while the other 5% are under unlimited embargo, based on a decision of the faculty to protect intellectual property and innovation.

Only few data on long-term trends have been published. Based on figures from ProQuest, Hawkins et al. (2013) identified an increasing number of embargoed ETDs. The findings by Owen et al. (2009) can be interpreted in the same way, especially for short-term one-year embargoes. On the other hand, the embargo statistics at West Virginia appear to be relatively stable over time (Hagen 2010), just like the figures between 2008 and 2011 from Lille.

Explanations

Following our review and survey data, experts and professionals explain the access restrictions in different ways, with arguments based on statistics, experience and anecdotic evidence. In a UK survey on mandates for ETDs, 88% of the universities indicated that they allow authors of theses to impose restrictions on access to their work, i.e. the electronic file, for many different reasons. Students, with the agreement of their supervisor, can request an embargo for the following reasons: commercial contract (for instance, funding by an external organisation), patent pending, ethical confidentiality and/or sensitive material (data protection), publication pending and third party copyright (Brown et al. 2010). The same study reveals that restrictions on grounds of third party copyright, data protection or potential risks to personal safety were reported only amongst ETDs (not print support) and that only 60% of the universities allow students to impose restrictions for print theses.

At Brunel University, "while every effort has been made to ensure that embargoing access to theses is not used as 'a panacea against all ills', students are offered the option of a 3-year embargo if they have a publication or patent pending" (Brown & Sadler 2010). Academics of the University of Maryland mention future publication, protection of data or work, student request, proprietary data and patent application as primary reasons for approving of embargoes (Owen et al. 2009).

In France, PhD theses are considered as administrative documents and (except confidential research) must be disseminated, at least on the campus (Schöpfel & Lipinski 2012). Yet, following our survey at Valenciennes (France) PhD students appear sometimes confused by embargo, confidentiality and on-campus options.

In Italy, Arabito et al. (2008) justify embargo options as indispensable for the same reasons: "(…) the free availability of doctoral theses on the web can be jeopardized by thorny copyright issues, which arise in the following cases: use of third party owned materials (…), third parties involved (possible infringement of privacy), patentable discoveries (…), and ongoing publication of data (according to the publisher policy)".

This last argument – expected publication – is by far the most common reason and explains between 1/3 (Owen et al. 2009) and 3/4 (Pavani & Mazzeto 2009) of all embargo decisions. The role of faculty appears to be crucial. At Virginia Tech, nearly half of the students' embargo decisions were taken on advice by faculty while requests by publishers are insignificant (McMillan et al. 2012). Ramirez et al. (2013) confirm that "scholars continue to doubt the viability of publishing opportunities after a dissertation or thesis becomes available electronically in an open access repository. Perceptions and fear, not data, inform many graduate advisors' and graduate students' decisions to restrict access to their ETDs".

Each graduate school has its own guidelines. A recent survey with more than 150 American graduate schools show that nearly 30% of all institutions "either don't allow an embargo at all, or don't tell students (about it at all) where they can find that information readily (…) In their enthusiasm for OA, universities and libraries across the U.S. are cajoling, arm-twisting, or even coercing students into in effect surrendering the copyright to their dissertations and theses, sometimes with the threat that students cannot graduate if they disagree" (Hawkins et al. 2013).

Florida State University Graduate School implemented access restriction - on campus only access – for older, digitized PhD theses: "Since retrospective digitized theses and dissertations did not include retrospective digitized access agreement forms, senior leadership recommended IP restriction for all FSU retrospective digitized theses and dissertations in 2009" (Smith 2009).

Kleister et al. (2013) report how changing the embargo policy at the University of North Texas dramatically dropped down the number of embargoed ETDs, from 80-100 to 20 or less per year. Asking for embargo has always been possible but the burden was on the PhD student to initiate the discussion. From the moment (2007) when this "burden" was replaced by a simple option on the agreement form (as check boxes), the number of embargo decisions was multiplied by more than five. Their conclusion is clear: "The needs of students must be balanced against the institution's needs and goals. Justification for embargo should not be especially onerous, but needs to be more than a mere checkbox on a form…"

At West Virginia University, Hagen (2010) reports that for the period 1998-2010, 85% of the more than 4600 theses are disseminated without any restriction. The part of theses with restricted access decreased from 47% (1998-2000) to 15% in 2010, because the option of encrypted on-campus only access was phased out in 2009 while the part of embargoed ETDs remained stable.

Smith (2009) describes how the Florida State University Graduate School requested campus-community and PDF document security options starting in Fall 2008, and he adds that "since retrospective digitized theses and dissertations did not include retrospective digitized access agreement forms, senior leadership recommended IP restriction for all FSU retrospective digitized theses and dissertations in 2009". Following the published figures, this part of restricted access can be estimated at about 16%.

Only three studies present detailed embargo statistics cut down by scientific disciplines (Owen et al. 2009, Pavani & Mazzeto 2009, ProQuest 2012). Yet, these survey results are not really reliable. Some

disciplines appear to be relatively consistent, such as life and chemical sciences, agriculture and environment, business, some domains of engineering (applied sciences) and public health, all with medium or high rates of embargoes. Pavani & Mazzeto (2009) show that in Science and Technology, pending publications as a reason for embargo concern mostly articles (73%) while in Social Sciences students intend above all publishing a book (57%). Yet, we must be careful with these statistics because of more or less small samples.

People, institutions, reasons and objectives

At first glance, the situation appears rather simple. PhD theses being intellectual work, the student is the only person holding the right to decide about dissemination. Of course, this view is by far much too simplistic. Different actors – people and institutions – can be distinguished who impact more or less the process of decision-making, with different reasons, motivations and objectives. A non-exhaustive list may be helpful to distinguish the different participants in this decision-making process:

- *PhD student: may want to keep the rights to his/her intellectual work; receives advice or orders from the different actors of his/her scientific community*
- *Director of PhD thesis: concerned by quality and reputation, fear of plagiarism.*
- *Jury: concerned by quality and reputation, protection of results.*
- *Community (discipline, staff): supportive or indifferent attitudes towards open access.*
- *Other PhD students: shared concerns about career, evaluation, and plagiarism…*
- *Graduate school: favourable or indifferent towards open access.*
- *University presidency (dean): supporting or not open access policy; concerned with third party rights (confidentiality, copyright infringement).*
- *Academic library: often in favour of open access and running an institutional repository.*
- *Service provider: supportive or indifferent towards open access.*
- *Publishers: opposed or not to open access and publishing of OA theses.*

Figure 1 tries to map these players in a system of decision-making of dissemination and access to ETDs. Each player sets his own goals, fulfils specific functions, plays his particular role, sometimes consistent with others, sometimes in opposition. For instance, PhD students may deposit their non-reviewed papers in open archives “off-campus”, outside of their institution and without any validation or authorization, even when the jury rejects the disclosure.

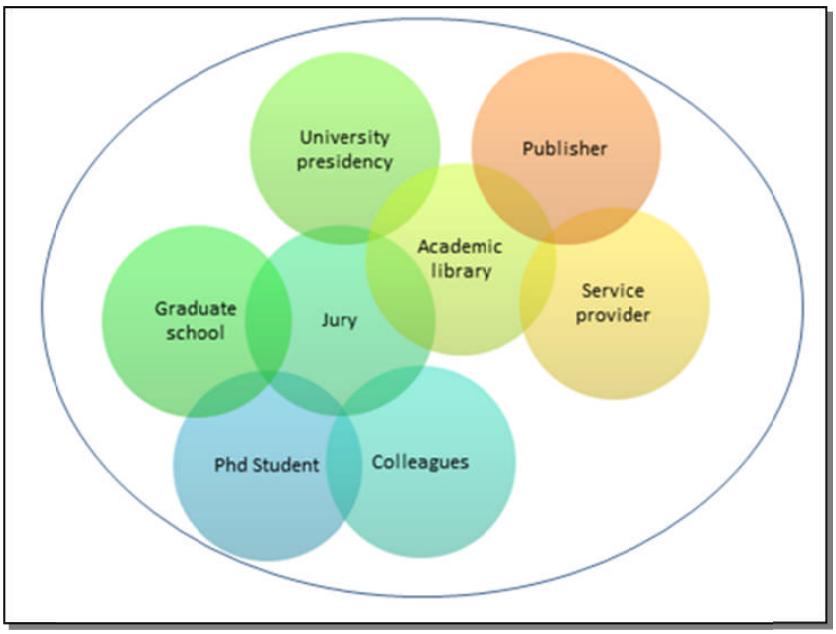


Figure 1: Participants in the decision-making on dissemination of ETDs

All these people, groups and institutions act in different ways, for different reasons, with different objectives and strategies. The literature review and survey results reveal the following components that may be understood as independent variables of the final decision:

- *A publishing project (article, book): if the student intends to publish his results with a scientific publishing house, he/she may be reluctant to disseminate the thesis on the Internet.*
- *Individual knowledge (or ignorance) of publishers' policies towards publishing papers that are already available on the Internet.*
- *Individual attitude toward open access (awareness, ethics, risk avoidance).*
- *Institutional decision on confidentiality and dissemination.*
- *Legal environment (copyright, intellectual property, disclosure of PhD theses).*
- *Institutional open access policy (awareness, risk avoidance).*
- *Institutional workflow of processing ETDs (reference points, opt-ins or opt-outs, easiness).*
- *Protection of third parties' rights (intellectual property, confidentiality, privacy).*
- *The jury members' advice (quality and excellence, awareness of open access).*
- *Tradition and attitudes of the scientific community.*
- *Publishers' acceptance of open access papers: If the publisher does not accept papers or books based on theses openly available on the Internet, his attitude may foster decisions in favor of embargoes.*

Each of these aspects acts in a different way. Some elements may decide on dissemination or non-dissemination, while others are limited to embargo or on/off-campus decisions. Moreover, some are case-by-case decisions while others reflect general attitudes and stable behaviours. Again, a schema may be helpful for global understanding (figure 2):

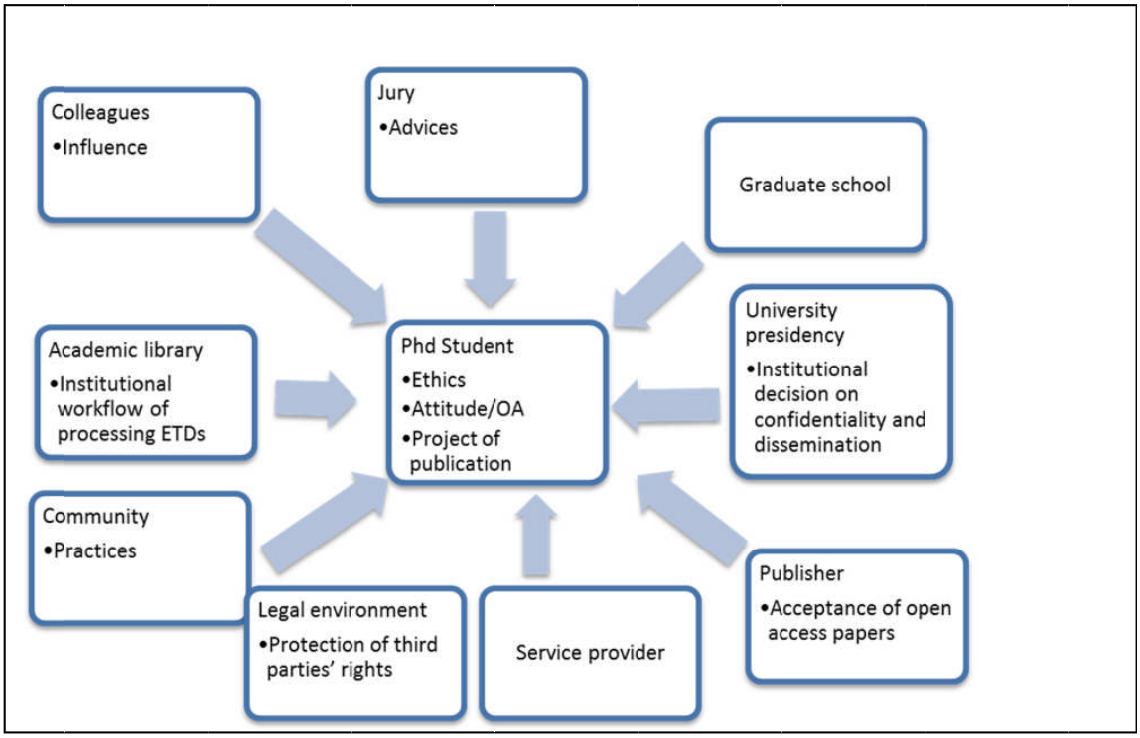


Figure 2: Different aspects of decision-making on dissemination and concealment of ETDs

This model may need empirical confirmation and perhaps, more details. Yet, its central characteristics are the multi-factorial or multivariate approach to the prediction of decisions on dissemination or concealment of ETDs. Even if individual publishing strategies and attitudes towards open access may play a major role, other variables such as personal advice from the PhD director, easiness of decision and references should not be neglected, in particular when discussing ways of improving accessibility and availability of PhD theses.

A model of openness

With regards to accessibility and availability of PhD theses, our analysis showed so far that openness is not a simple, binary concept but that the documents can be more or less open, depending on different variables. Some of those variables are similar to articles published in journals or books, but others are specific to PhD theses. In October 2012, the *Scholarly Publishing and Academic Resources Coalition* (SPARC), "an international alliance of academic and research libraries working to create a more open system of scholarly communication"[liii], released a guide called "How Open Is It" that outlines the core components of open access (e.g., reader rights, reuse rights, copyrights, author posting rights, etc.) across the continuum from "open access" to "restricted access". Compared with our multivariate approach, this *Open Access Spectrum* (SPARC 2012) helps to get a realistic view on the problems of openness, disclosure and concealment of theses. Table 2 shows a possible adaptation of the SPARC guide to the specific conditions of the dissemination of PhD theses.

Reader rights: On site only, or also at distance, via authentication? What about interlibrary loan or document delivery?

Reuse rights: Generous reuse rights (CC-BY licensing) or full copyright protection?

Copyrights: No third party claim or complete concealment (confidentiality) because of sensible or protected results?

Institution rights: In France, by decree, the PhD theses must be disclosed, at least on the campus, except for confidential projects.

Institution policy: In fact, at least two different levels must be distinguished, the global policy of the university (or faculty/department), and the approach of the jury which may, in some cases and sometimes for political reasons, reject open access disclosure via institutional repository even for non-confidential theses.

Posting workflow: Following empirical studies on changes in ETD workflows, we adapt the SPARC component "Automatic posting" to the specific ETD environment. The continuum between open and closed ranges from procedures without embargo options, i.e. where an embargo decision needs a specific individual action (written and argued request), to workflows where open access is available only as an opt-in option while on-campus dissemination is the default option.

Machine readability: The last component is about automatic access and exploitation of the full text and the related data and metadata. Exploitation means: text or data mining, harvesting, or crawling. Our table summarizes the SPARC scale but specifies the existence of supplementary data files (tables, videos, images etc.) that may have been submitted together with the thesis.

Table 2: Degrees of openness for electronic theses in institutional repositories (adapted from SPARC 2012)

Access	Reader Rights	Reuse Rights	Copyrights	Institution Rights	Institution Policy	Posting Workflow	Machine Readability
Open	Free readership rights immediately upon posting	Generous reuse & remixing rights (CC-BY license)	No third party claim	Obligation of free dissemination	OA mandat	No embargo option available in workflow, embargo needs special procedure	Full text, metadata, citations & data, incl supplementary data (annex etc), provided in community machine-readable standard formats through a community standard API or protocol
	Free readership rights after 6 months embargo	Reuse, remixing & further building upon the work subject to certain restrictions & conditions (CC BY-NC & other CC licenses)			Institutional repository		Full text, metadata, citations & data, incl supplementary data (annex etc), may be crawled or accessed through a community standard API or protocol
	Free readership rights after 12+ months embargo	Reuse (no remixing or further building upon the work) subject to certain restrictions and conditions	Third party requires embargo	Obligation of dissemination with restrictions (campus)	OA support	Opt-out if claim for embargo	Full text, metadata & citations may be crawled without special permission or registration
	Free readership rights only on-campus (Intranet)				Interest for OA		Full text, metadata & citations may be crawled with permission
Closed	Total embargo (confidentiality)	No reuse rights beyond fair use/limitations & exceptions to copyright (all rights reserved copyrightà	Third party claims full IP protection (confidentiality, privacy)	No institution rights	No OA policy	Opt-in for free dissemination	Full text & metadata not available in machine-readable format

Back to grey?

A good idea does not necessarily guarantee success. Internet is not synonymous with openness, and the creation of institutional repositories and ETD workflows does not make all items more accessible and available. Sometimes, the new infrastructure even appears to increase barriers to PhD theses.

Different reasons contribute to this unexpected (and most often, unwanted) development, and in a certain way, new technologies and digital infrastructures trigger the tendency for access restrictions. In our first paper, we discussed empirical data in terms of ethics, law, legitimate interests and policy, trade secrets, individual and institutional strategies and workflow-biased decision-making. Our present communication adds a conceptual framework and a differential description of the specific conditions of this part of scientific communication.

Open access is without doubt a valuable and important goal for scientific communication. Yet, scientific and technical information, considered as a part of research behaviour and object of strategic decisions (Roosendaal et al., 2010) always included decisions on concealment and parts of secrecy. Together with copyright and technologies, these individual and institutional decisions contribute today to an unsatisfying and inefficient situation where one part of digital PhD theses are easy to find and to obtain while others remain hidden, embargoed and/or limited to on-campus access. As for open access and institutional repositories in general, one part of the research community is (so far) indifferent or hostile to unprotected dissemination of theses. From the moment the decision on dissemination of ETDs moves from institution to the individual author, we have to deal with these attitudes and opinions.

Openness is not enough for scientific communication. Internet does not change grey literature into white in a mechanical way. Without a minimum of quality and standardization (Dobratz & Scholze 2006), without metadata, referencing, long-term preservation, discovery tools etc., (Schöpfel et al. 2011), and without raising awareness, thorough decision aids and redesigned workflows, perhaps even changes in the legal status of theses, institutional repositories will only provide a partial answer to the question of grey literature in the digital environment. So, back to grey?

Bibliography

S. Arabito, et al. (2008). 'Time to Harvest: Electronic Doctoral Theses in Italy'. In *ETD 2008 11th International Symposium on Electronic Theses and Dissertations , 4 - 7 June 2008, The Robert Gordon University, Aberdeen, UK*.

H. M. Artus (2003). 'Old WWWine in New Bottles? Developments in electronic information and communication: structural change and functional inertia'. In *Fifth International Conference on Grey Literature: Grey Matters in the World of Networked Information, Amsterdam, 4-5 December 2003*.

M. Banach (2011). 'The Benefits of Managing and Publishing ETDs "In House" Using an Open Access Repository'. In *USETDA 2011: The Magic of ETDs…Where Creative Minds Meet. May 18-20, Orlando, Florida*.

S. Bok (1984). *Secrets: On the Ethics of Concealment and Revelation*. Oxford University Press., Oxford.

J. Brown & K. Sadler (2010). 'Vision, impact, success: mandating electronic theses. Case studies of e-theses mandates in practice in the UK Higher Education sector'. Tech. rep., UCL University College of London.

J. Brown, et al. (2010). 'Influencing the Deposit of Electronic Theses in UK HE: report on a sector-wide survey into thesis deposit and open access'. Tech. rep., UCL University College of London, London.

S. Dobratz & F. Scholze (2006). 'DINI institutional repository certification and beyond'. *Library Hi Tech* 24(4):583–594.

J. H. Hagen (2010). 'Managing ETDs and Evolution of the Repository: Twelve Years of Case Studies at West Virginia University'. In *ETD 2010 13th International Symposium on Electronic Theses and Dissertations. June 16-18, 2010, Austin, Texas*.

J. Kleister, et al. (2013). 'Riding the Restriction Wave at the University of North Texas: Comparing IR and ProQuest Embargoes'. In *Third Annual United States Electronic Thesis and Dissertations Association (USETDA) Conference, 2013, Claremont, California, United States*.

G. McMillan, et al. (2012). 'Do Open Access ETDs Affect Publishing Opportunities in the Sciences? '. In *CNI Fall Membership Meeting, December 12, 2012*.

T. M. Owen, et al. (2009). 'ETDs in Lock-Down: Trends, Analyses and Faculty Perspectives on ETD Embargoes'. In *ETD 2009 12th International Symposium on Electronic Theses and Dissertations. June 10-13, 2009, Pittsburgh, Pennsylvania*.

A. Pavani & A. Mazzeto (2009). 'Restricted ETDs and open access'. In *ETD 2009 12th International Symposium on Electronic Theses and Dissertations. June 10-13, 2009, Pittsburgh, Pennsylvania*.

ProQuest (2012). 'Ten Year Embargo Trends'. In *CGS Summer Workshop and New Deans Institute, Boston, Massachusetts, July 7 – July 11, 2012*.

M. L. Ramirez, et al. (2013). 'Do Open Access Electronic Theses and Dissertations Diminish Publishing Opportunities in the Social Sciences and Humanities? Findings from a 2011 Survey of Academic Publishers'. *College & Research Libraries* 74(4):368–380.

H. E. Roosendaal, et al. (2010). *Scientific Publishing: From Vanity to Strategy*. Chandos Publishing, Oxford.

J. Schöpfel & T. A. Lipinski (2012). 'Legal Aspects of Grey Literature'. *The Grey Journal* 8(3):137–153.

J. Schöpfel & H. Prost (2013). 'Degrees of Secrecy in an Open Environment. The Case of Electronic Theses and Dissertations'. *ESSACHESS - Journal for Communication Studies* 6(2 (12)).

J. Schöpfel (2010). 'Towards a Prague Definition of Grey Literature'. In *Twelfth International Conference on Grey Literature: Transparency in Grey Literature. Grey Tech Approaches to High Tech Issues. Prague, 6-7 December 2010*, pp. 11–26, Amsterdam. Text Release.

J. Schöpfel, et al. (2011). 'Open is not enough. A case study on grey literature in an OAI environment'. In *Thirteenth International Conference on Grey Literature: The Grey Circuit. From Social Networking to Wealth Creation. Washington, DC, 5-6 December 2011*, pp. 75–86, Amsterdam. Text Release.

P. Smith (2009). 'ETDs, IRs, Collaborations, and Access at Florida State University'. In *ETD 2009 12th International Symposium on Electronic Theses and Dissertations. June 10-13, 2009, Pittsburgh, Pennsylvania*.

SPARC, PLOS, OASPA (2012). 'How Open Is It? Open Access Spectrum'.

P. Suber (2012). *Open access*. MIT Press, Cambridge MA.

J. Willinsky (2005). *The Access Principle: The Case for Open Access to Research and Scholarship (Digital Libraries and Electronic Publishing)*. MIT Press, Cambridge MA.

All websites accessed in September 2013.

[i] *Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities* of 22 October 2003 http://oa.mpg.de/lang/en-uk/berlin-prozess/berliner-erklarung/

[ii] For more details and examples, see Schöpfel & Prost (2013).

[iii] Gateway to French theses at http://www.theses.fr run by the French Bibliographic Agency of Higher Education ABES http://www.abes.fr

[iiii] http://www.sparc.arl.org/about

For better or for worse
Knowledge output 1944-2001 and effects of the Legal Deposit Act no. 20/2002 and e-publishing on access to GL in Iceland

Stefanía Júlíusdóttir, University of Iceland; School of Social Sciences, Iceland

Abstract

Two studies are presented in this paper. The former was a study on the amount of knowledge output in Iceland, from 1944 to 2001. It covered all publications issued in Iceland in the sampling years, in the following publication genres: all market publications and GL published in Iceland on paper and catalogued in Gegnir, the national union catalogue, international peer-reviewed journal articles bibliographically accessible in Web of Science (WoS), in addition to patent applications, and standards in Iceland. The findings indicated considerable changes in the knowledge output, in particular increases in GL, international peer reviewed journal articles, and patent applications. They also indicated that GL might increasingly have been produced in-house, and published electronically as the 20th century drew to a close; a trend that continued in the 21st century. Publications produced in-house were often not delivered on legal deposit and the e-medium was not covered by the Legal Deposit Act no. 43/1977 and therefore many a time missing in Gegnir. The latter study was aimed at finding how access to GL developed at the beginning of the 21st century. In particular to find what the effects of the Legal Deposit Act no. 20/2002, which covered e-publications, were on legal deposit in the National and University Library of Iceland (NULI). The findings indicated that amendments are necessary to the Legal Deposit Act no. 20/2002, if the access to written knowledge that used to prevail in Iceland is to be maintained.

Introduction

Public access to knowledge and information has been considered one of the prerequisites for functional democracy (Thorhauge, Larsen, Thun, and Albrechtsen, 1997; Lariviére, 2000). In this respect written knowledge, both market publications and GL, is becoming ever more important. Access to written knowledge is vital because no decision is better than the knowledge and information it was based on (Machlup, 1984, Vol. 3, p. 15). Apart from democratic dependence, written knowledge is increasingly needed for every day work. Close to a century ago it was argued that by ignoring GL the progress of science could be harmed (Auger, 1996 (1994), p. 6). This argument is equally valid today when social progress depends on innovation, which in turn depends on access to knowledge and information, in particular GL; because it is less costly to use already established knowledge than having to discover and establish it anew (Georgescu-Roegen, 1999 (1971), p. 22).

Public access to knowledge and information depends on access of the public to bibliographic catalogues that provide information on available material by facilitating retrieval through various kinds of access points (at the minimum author, title and subject searches); whereupon needed material can be accessed in libraries and bookstores. This paper presents a study (2002) that was aimed at assessing the knowledge output in Iceland as manifested in the amounts of market publications, GL, peer reviewed journal articles, patent applications and standards; and a follow-up case study (2012) on the availability of public access, both bibliographic and to the publications themselves.

The idea for these studies was sparked by my experience of working on The *Icelandic national bibliography* in the 1970's and 1980's as a cataloguer; and it was inspired by two books: *A social history of knowledge: From Gutenberg to Diderot* by Peter Burke (2002) and *Shaping written knowledge: The genre and activity of the experimental article in science* by Charles Bazerman (1988). During the time of my work on The *Icelandic national bibliography* it seemed that a large part of the publications issued in the country were not market publications, but rather various kinds of GL, such as government publishing, research reports of various kinds, guidelines for carrying out research, reprints, informational material from political parties, religious congregations, and associations to name some examples of the kinds of GL. All of them were entered in The *Icelandic national bibliography,* provided that they counted at least 5 pages. Therefore bibliographic control of GL, apart from black literature (see below), did not use to be a problem in Iceland, and the publications themselves were available for use at The National Library in Reykjavík and the Quarter Libraries in other parts of the country (Júlíusdóttir, 2006).

In this respect the Icelandic situation differed from that in many other countries where getting hold of GL was difficult because of lack of bibliographic control and hence physical access to the literature itself became problematic (Auger, 1996 (1994); Farace and Schöpfel, 2010).

Findings of the 2002 study indicated greatly increase output of written knowledge during the latter half of the 20th century. Findings of the 2012 study indicated that technological developments that facilitated

in-house production of GL and e-publishing towards the end of the 20th century and the beginning of the 21st century made GL harder to get, contrary to expectations.

Previous work

The author found no studies comparable to the ones presented here. That is studies analysing the development of various kinds of knowledge output within a country, with a focus on market publications on the one hand and GL on the other. In Iceland two studies had been carried out on publications (market publications as well as GL), both focused on how the publication output developed according to the subjects published, without any regard to whether they were market publications or GL (Hjartar, 1968; Eyþórsdóttir, 1997). In addition reports on market publishing in Iceland had been carried out for the Association of Icelandic Publishers (for example Skýrsla fyrir Félag íslenskra bókaútgefenda, 2003), not addressing GL.

Abroad studies on the importance and amount of GL in various subjects have been undertaken. There GL was estimated to be somewhere in the vicinity of 9%, towards the end of the 9th decade of the past century and to be over 20% around the middle of its 10th decade (Debachere, 1995). Other studies focused on the importance and amount of GL in research as it was manifested in citations in research publications (for example Chalabi, 2012; Gentil-Beccot, 2010; MacDonald, Wells, Cordes, Hutton, Cossarini, and Soomai, 2010; Chaves, 2010).

Useful in the present study was Wessels definition of three different types of publications with respect to bibliographic access along with access to the publications themselves. The definitions were the following: white publications, on which bibliographic information is easily accessible by the public. The publications themselves are distributed through traditional distribution channels of market publications and easily available in book stores and libraries. The second category is grey literature, which is neither bibliographically accessible in catalogues open to the public nor available through traditional distribution channels of market publishing, even when they are not confidential or secret publications. The third category is black publications. They are confidential or secret publications, which according to the definition are only meant to be accessible by the few. They are therefore inaccessible bibliographically (not found in catalogues open to the public) and the publications themselves are difficult or impossible to get a hold of (Wessels, 1997).

Research questions

The two studies were conducted at different times; the former was carried out in 2002, and the latter was a follow-up study carried out in 2012. The aim of the former study was to analyse ways of distribution of new knowledge created in Iceland in the period 1944-2001 as it was manifested in different ratios of market publishing, GL, Icelandic patents and Icelandic standards. This was considered important because changes in the ratios of these kinds of knowledge outputs did affect public accessibility to knowledge and information. Access to knowledge and information distributed in market publications was easiest. The research questions in the 2002 study were the following:

1. Had there been changes in the proportion of grey versus market publications in Iceland from 1944 to 2001, and if so then which changes?
2. What was the percentage of knowledge created in Iceland and distributed as the genre of the international peer-reviewed journal article, from 1979 to 2001? The only source available on that issue to the researcher at the time of study was the Web of Science (WoS) database going back to 1979.
3. Had Icelanders attempted to use Icelandic patents and standards as a way of disseminating new knowledge from 1944 to 2001, and if so then to what extent?

Amongst the findings of the 2002 study was an indication that e-publishing had led to decreased bibliographic access and hence decreased access to the e-publications themselves by the authorities as well as the public in Iceland. One of the reasons for this was thought to be that the *Legal Deposit Act no. 43/1977* in force at the time only stipulated deliveries of printed matter and sound recordings. This shortcoming was to be amended in the *Legal Deposit Act no. 20/2002* (taking force on January 1st 2003), which covered a greater variety of media, amongst them e-publications, which were to be delivered to NULI and made accessible to the public.

The aim of the 2012 study 2012 was to find out if the effects of the 2002 legal deposit act had led to increased deliveries of e-publications to NULI and hence made a higher percentage of e-GL, as well as other e-publications accessible both bibliographically and online to the public in the decade that had passed since the 2002 act took force?

Sampling strategy

In the 2002 study the years of sampling were 1944, 1969, 1979, 1989, and 2001. They were chosen with regard to special events likely to have affected research and knowledge output of the nation. The assumption made was that the effects of these events on knowledge creation were realized some years after the event.

The Icelandic Research Council was, founded in 1938-1940. It was the first one of its kind in the Nordic countries (Lúðvíksson, 2002). In 1944, Iceland gained complete independence from Denmark. During the forties a special department (named Atvinnudeild Háskólans), was established at the University of Iceland for research in the areas of agriculture, fishing and industry. The main function of the University of Iceland had, until then, been to prepare officials of the church and state for office, educating priests, lawyers and doctors (Hálfdanarson, Matthíasdóttir and Guðmundsson, 2011; Skúlason, 2000).

In the mid sixties the special department at the University of Iceland, was discontinued and three independent research institutions established in its place by law: The Marine Research Institute, The Agricultural Research Institute and the Industrial Research Institute (*Act on Research in the Interest of Industry no. 64/1965*). At survey time, the number of research institutes, privately and publicly operated had increased considerably.

During the seventies considerable changes took place. Two new faculties were established at the University of Iceland, the Faculty of Social Science and the Faculty of Science (Hálfdanarson, Matthíasdóttir and Guðmundsson, 2011; Skúlason, 2000). The largest increase in the average number of publications in Iceland per year took place during that decade (Fjöldi útgefinna bóka á íslensku, 2006, April, 26th). The bibliography of the yearly output of publishing in Icelandic that since 1944 had been published as a chapter in the *Yearbook of the National Library* (National Library of Iceland, 1945) became a separate publication in two volumes. One contained the bibliography of items printed on paper and another one the bibliography of sound recordings. The volumes for 1979, published in 1980, were the first volumes of the national bibliography to be published using computer technology (National Library of Iceland. National Department, 1975; National Library of Iceland. National Department, 1980). That point in time marked the beginning of general use of computers in Iceland, although they had been in specialized use in the country since the 1950's (Skýrr, 2006).

During the eighties computer technology was taken into use in Icelandic libraries: catalogues in libraries in Iceland became automated and database searches became common in research and special libraries (Einarsson, 1990; Harðarson, 1993; Hauksdóttir, 1996; Júlíusdóttir, 1989, 1991, 1996).

During the last decade of the 20th century legal acts on public access to information and knowledge in primary sources in the public sector were enacted by the Icelandic Parliament (e.g. Administration Procedures Act no. 37/1993; Information Act no. 50/1996; Privacy Protection Act no. 77/2000). They affected both organization of primary sources and access to them. Furthermore, Gegnir was established at the end of that decade (Hauksdóttir, 1996, 2001). The year 2001 marked the beginning of the new millennium. The last issue of The *Icelandic national bibliography* published on paper was issued for that year. Since then bibliographic information on publications issued in Iceland has been made public online in *Íslensk útgáfuskrá* (the *Icelandic National Bibliography*) (http://utgafuskra.is). More and more publications were issued in the digital format and made accessible over the Internet and NULI established an Icelandic Web collection (National and University Library of Iceland, n.d., 2009a, 2009b, 2011).

In the 2012 study access (bibliographic and to the publications themselves) to publications of two research institutions was investigated. They were the Marine Research Institute (MRI) founded in 1965 in accordance with the Act on Research in the Interest of Industry (see above), and the Institute of Economic Studies at the University of Iceland founded in 1989. Both were chosen for their importance to the national economy and because both of them had served in an advisory function to public authorities in Iceland. Both had issued publications on paper and electronically under the 1977 and 2002 legal deposit acts and MRI also under previous legal deposit acts, and both were publishers of important GL.

Data collection

In the 2002 study the following data was collected:

1. Data on publications issued in Iceland in the sampling years. NULI, the legal deposit library kindly, granted a copy of data on these publications in electronic format from Gegnir. Having worked on The *Icelandic national bibliography* I knew that in spite of the time limit given in the Icelandic Legal Deposit Act on delivery of publications (*Legal Deposit Act no. 43/1977*) years could go by before some of the GL was delivered to NULI and was subsequently catalogued in The *Icelandic national bibliography.* Moreover, some of the GL, produced in-house in various institutions and organizations was not delivered to NULI at all. I decided to use electronic data drawn from Gegnir

rather than the paper version of The *Icelandic national bibliography*, in the hope that it would include grey literature such as internal reports available at organizations and institutions catalogued in Gegnir by their staff, but never delivered to NULI and hence not found in the paper publication of The *Icelandic national bibliography*. Using output from Gegnir rather than The *Icelandic national bibliography* on paper was therefore expected to cover more of the GL and to give more reliable results. The data was received as an Excel file. It included: author, title, publisher, classification code, year of publication, and edition. The drawback with using the digital data from NULI was however that a tremendous amount of unanticipated work had to be put into cleaning it up in an attempt to make sure that there was only one record for each published item.

2. Data on items published on the international scholarly market. Number of items published for the international scholarly market, by authors residing in Iceland, was found by searching the Web of Science (WoS), for the sampling years as far back as the WoS went, i.e. for 1979, 1989 and 2001. The searches were carried out on the 15th of July 2002 in the *Science Citation Index Expanded (SCI-EXPANDED)* – 1970 - July 2002, the *Social Sciences Citation Index (SSCI)* – 1970 - July 2002, and the *Arts & Humanities Citation Index (A&HCI)* – 1975 - July 2002 (Web of Science).
3. Information on patent applications in Iceland. Information on the number of patent applications to the Icelandic Patents Office by Icelanders in the sampling years was kindly granted for this research project by the Patents Office in Iceland.
4. Information on Icelandic standards. The information on Icelandic standards issued in the sampling years was kindly granted for this research by the Icelandic Standards Office. Standards were mostly (over 90%) adopted from abroad. Those created in Iceland proved to be too few for inclusion in this project (Rögnvaldardóttir, 2007).

In the summer of 2012 data on the publishing output of the two chosen research institutions was gathered on their home pages where lists of their publications were to be found. Items on the publication list of MRI consisted both of GL and market publications. To get an indication of the effects of the 2002 legal deposit act on how bibliographic access and in many cases also access to the publications themselves had developed, the items on the lists were searched for in Gegnir.
Neither of the chosen institutions used Gegnir as their library catalogue. If the items on the lists were found in Gegnir they would therefore have been catalogued by another library most likely by NULI as legal deposit items. The searches in Gegnir were title searches with and without subtitles. Moreover keyword searches were also used when the title searches gave no results. Occasionally the year of publication in the catalogue record found in Gegnir was different from that found for the same item in the publication list of the institution. When that happened the publication year in Gegnir was used in the analyses.

Data analysis

Data from NULI, the Icelandic Patent Office, and WoS was analysed by use of the bibliometric method (Diodato, 1994; Hertzel, 1987; Pritchard, 1969) in order to find out if there had been changes in the proportions of each type of knowledge output under investigation. The data from the Icelandic Patent Office was not analyzed further according to subgenres. Calculations were made of the ratio of market publications at home and abroad (as manifested in WoS) of knowledge created in Iceland versus the ratio of such knowledge and information disseminated as GL, and as patent applications in Iceland in the sample years.
Published items were analysed as market publications when they were published by companies that had publishing for profit as their main function and as GL when published by parties that had something else than publishing for profit as their main function. My work as a cataloguer at The *Icelandic national bibliography,* made it possible for me to analyse the issuing bodies of publications in Iceland into these two types at sight. Information on knowledge created in Iceland and published abroad was only available at WoS and only for the sampling years 1979, 1989 and 2001. Calculations of the data for this 57 year period from 1944-2001 were done in Excel (Júlíusdóttir, 2006).

Problems and shortcomings

In the 2002 study, I assumed, with regard to production techniques that information for the study was available on almost all 1944 publications, that the same was probably true for the 1969 ones, and even most of the 1979 ones. I did, however, suspect that changes in the methods of production since around 1990 (that made in-house production of publications ever easier), would have resulted in a larger percentage of publications being delivered late or not at all to NULI. Such publications would not have been found by the data collection methods used in this study and were therefore not included, even when they were not intended to be inaccessible or black by the publisher. Thus, one of the main

shortcomings of the study as a whole was the fact that to obtain all the data and information needed was very difficult, if not impossible, due to the difficulty of finding out what had been published, particularly in the latter sampling years. This applied to GL published in Iceland and also to knowledge created in Iceland and published abroad as market publications. Therefore, the findings of such a study could at the best give an indication of the developments as also argued by Eyþórsdóttir (1997) and Hjartar (1968).

Another shortcoming was that the 2002 study had to be limited to paper publications because before January 1st 2003 only items printed on paper and sound recordings were to be delivered on legal deposit (Legal Deposit Act no. 43/1977). No information was, therefore, available on e-publications in Iceland, at the time of data collection in 2002. The third major shortcoming was that it was likely that some of the 2001 publications had not yet been delivered to NULI, and that some of the delivered ones had not yet been catalogued in 2002 when the data was collected. Therefore, a smaller percentage of 2001 publications would have been available for the study than for the other sampling years. This applied in particular to GL.

The fourth major shortcoming was that when cleaning the digital data from NULI (for example deleting duplicate (or multiple) entries for the same publications), the author noticed that some publications issued in years other than the sampling years of the study were included. These were deleted but their appearance, (in the data that should only have contained records from the defined sampling years), led one to suspect that likewise some of the publications of the years under study were missing if their publication year had in the same manner not been correct in the system.

Furthermore, no reliable statistics on sale of books published in Iceland existed. It was, in short neither known exactly how many titles had been issued, nor how many copies were sold of each title or in total in the country each year (Harðardóttir, personal telephone communication, June 12th 2002).

Only data in the SCI-Expanded Index, SSCI, and A&HCI at the WoS was available on the international appearance of scholarly publications by authors residing in Iceland; this limited the items worked with and did by no means cover the whole of the publications published abroad and created by those who resided in Iceland. It was known at the time of data collection that Icelandic literature was increasingly published abroad, but statistical data on the publishing of Icelandic literature abroad proved impossible to get in 2002 from those holding their distribution rights in Iceland (Júlíusdóttir, 2006). Although not covering the output of Icelandic literature abroad by foreign publishers, it was hoped, that the study would give an indication of the developments of the scientific activities on the international market by knowledge creators residing in Iceland.

Only patents applied for to the Icelandic Patent Office were included. Patent applications from Iceland, to patent offices abroad bypassing the Icelandic Patent Office (if any), were not included.

Findings

Table 1 provides an overview of the information sought in research questions 1-3 above. It shows that the output of written knowledge items addressed here escalated in Iceland in the period from 1944 to 2001; the increase was almost eightfold; from 404 to 3189 see Table 1, Line 7. However, the increase of each genre as a percentage of the total output differed. Thus, GL increased more than ten times while the combined market publications issued in Iceland and abroad increased almost six times, and the market publications issued in Iceland were only three and a half times more numerous in 2001 than in 1944 (see Table 1, Lines 1, 3, 4). Publications created in Iceland retrievable in WoS were under 4% of the total output of knowledge items in 1979, and over 14% in 2001 (see Table 1, Lines 2, 7). Patent applications were seven times as many in 2001 as in 1944; they were 1.7% of the total of the written knowledge genres addressed here in 1944, and 1.5% in 2001; while their numbers multiplied their percentage of the whole decrease in the time period under study.

The findings of the 2002 study, furthermore, indicated that the advent of e-publishing had lowered the percentages of GL catalogued in Gegnir. This resulted in a lack of public access, and sometimes also of access by the authorities to these publications. One of the reasons was that, even though GL publications might be open on the Internet, they could not be found due to lack of bibliographic cataloguing, especially when the subject terms sought were not in the title of the item.

Table 1: Developments of the knowledge output in Iceland 1944-2001 of market publishing, grey literature, the genres of international peer-reviewed journal articles and patent applications to the Icelandic Patent Office

Year of publication	1944	1969	1979	1989	2001	Line
No issued by commercial publishers	204	357	558	884	727	**1**
No issued by commercial publishers abroad, WoS	NA	NA	64	122	452	**2**
Total issued by commercial publishers	204	357	622	1006	1179	**3**
No issued by grey literature publishers	193	530	1021	862	1961	**4**
Total issued	397	887	1643	1868	3140	**5**
No of patent applications	7	26	9	18	49	**6**
Total output in Iceland inc. WoS	**404**	**913**	**1652**	**1886**	**3189**	**7**
						9
% issued by commercial publishers	50.5	39.1	33.8	46.9	22.8	**10**
% issued by commercial publishers abroad, WoS	NA	NA	3.9	6.5	14.2	**11**
% issued by commercial publishers in total	50.5	39.1	33.6	53.3	37.0	**12**
% issued by grey literature publishers	47.8	58.1	61.8	45.7	61.5	**13**
% of patent applications	1.7	2.8	0.5	1.0	1.5	**14**
% of total in DDC 800 (literature) including WoS	37.0	31.0	25.0	25.0	19.0	**15**

Already in 1944 GL accounted for 47.8% of total publishing in the country and was up to 61.8% of the total publishing in 1979. Since then GL-items as a percent of the whole of written knowledge catalogued in Gegnir decreased and were only 61.5% in 2001 (see Table 1, Line 13) (Júlíusdóttir, 2006). Moreover the percentage of literature (classified in 800 in *Dewey decimal classification system*) decreased steadily. It went from 37% in 1944 to 19% in 2001 of the total knowledge output (see Table 1, Line 15). At the turn of the century the total numbers of publications had increased much more than the number of the inhabitants (see Figure 1). This is considered to be an indication of increased dependence of the working culture of Icelandic society on written knowledge; and that research was becoming a way of making a living for many people in Iceland. Not to witness an increase in the percentages of the output of GL at the turn of the 21st century, when research activity in the country had increased greatly could not be right. It is argued here that this was an indication of in-house production of GL (for example research reports and other kinds of GL) and as the century drew to a close increased use of e-GL which was not covered by the legal deposit act and therefore probably not catalogued in Gegnir (see Figure 1 and Table 1). Standards created in Iceland were too few for inclusion in the study (see above).
The publishing output of market publishers in 1989 was unusually high. This finding was confirmed both by the editor of the *Icelandic national bibliography* and the manger of the Icelandic Publishers Association.

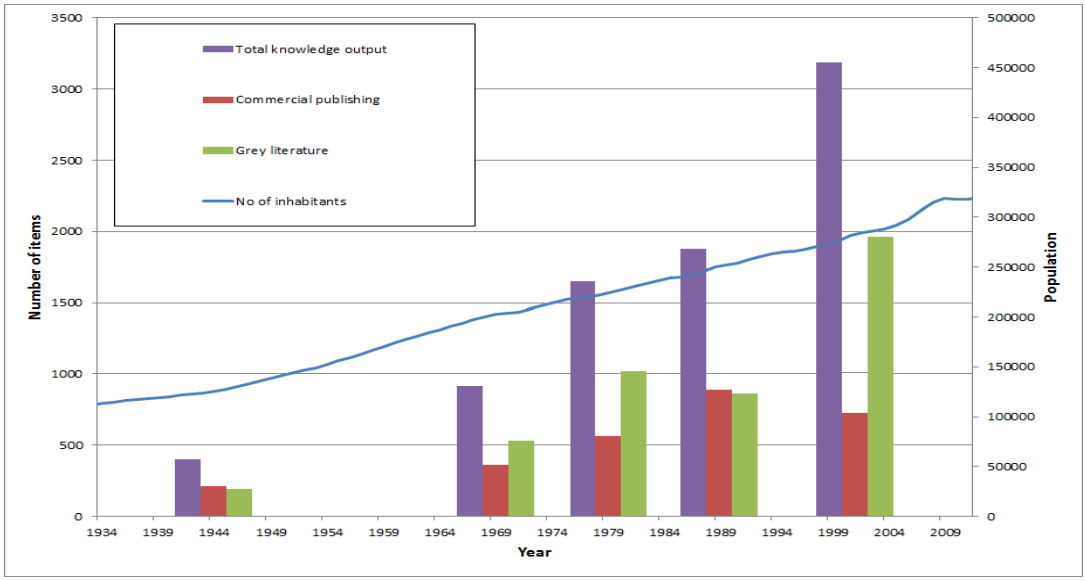


Figure 1: Development of total knowledge output in Iceland, market publications (issued in Iceland and abroad found in WoS), GL and patent applications in Iceland are depicted in the columns. The line shows the increase in the population for comparison.

Effects of the Legal Deposit Act no. 20/2002 on access to GL

The suspicion that e-publishing had affected bibliographic access to publications in Iceland called for a second study, which was undertaken in the summer of 2012. To find that out information on cataloguing of publications of two research institutions was gathered in Gegnir. These institutions were chosen for their importance to the national economy and because both of them served an advisory function to public authorities. They were the Marine Research Institute and the Institute of Economic Studies at the University of Iceland. The findings indicated that the advent of e-publishing had lowered the percentages of their publications catalogued in Gegnir. Resulting in a lack of public access and sometimes also of access by the authorities to these publications, even though they might be open on the Internet bibliographic access needed to locate them was not available (see Figures 2 and 3). Since the data collection was carried out in the summer of 2012 a few publications issued by these two institutions have been catalogued in Gegnir (as found in searches on November 14th 2013).

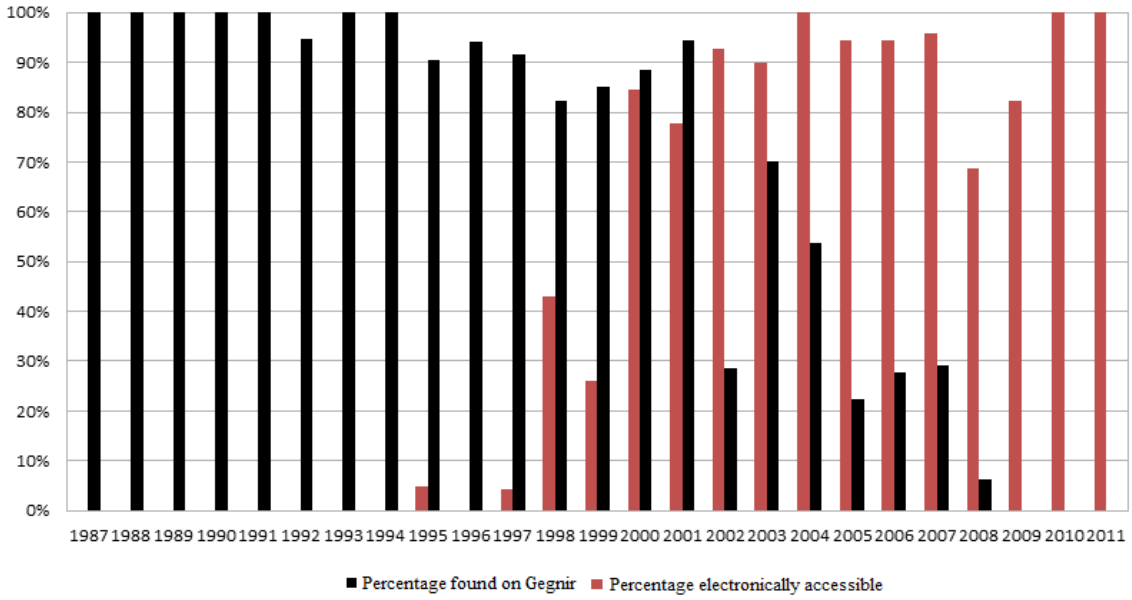


Figure 2: Percentages of publications of the Institute of Economic Studies at the University of Iceland catalogued in Gegnir in the summer of 2012 and thus available bibliographically and physically for use by the public.

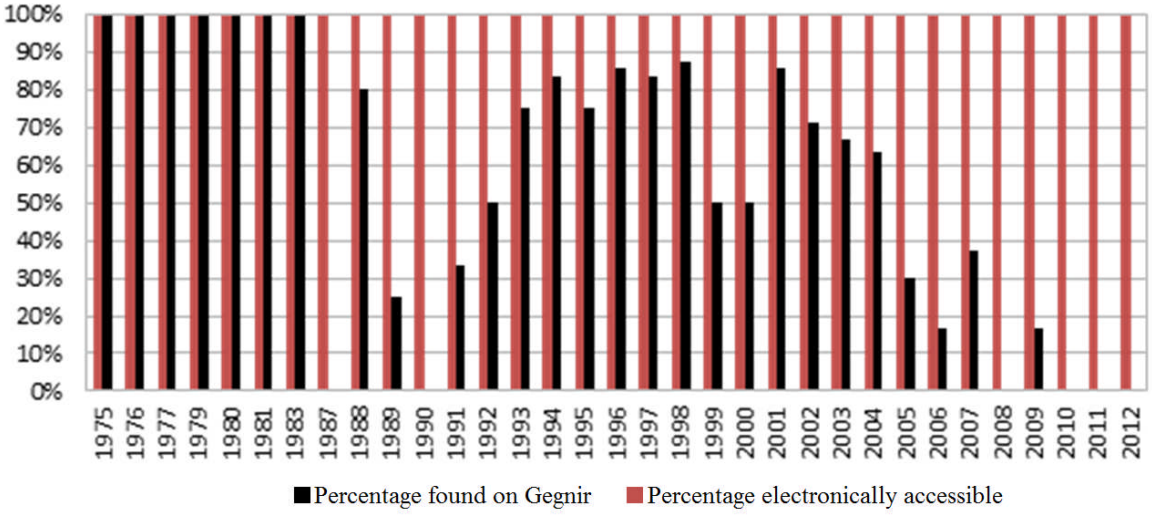


Figure 3: Percentages of publications of the Marine Research Institute catalogued in Gegnir in the summer of 2012 and thus available bibliographically and physically for use by the public.

At MRI all items on the publication list on the homepage had been scanned and were available as pdf documents to the public (see Figure 3). A similar kind of access was available to some of the publications on the list of the Institute of Economic Studies, although the e-items there were mostly items published in this century (see Figure 2). Contrary to expectations, bibliographic access to GL issued by the two institutions investigated did not improve after the 1st of January 2003. This meant that even finding known items on subjects that did not appear in their titles could be almost impossible, because Google searches proved imprecise and even useless in some such cases. Having to open and read many items on a long list of publications to find out which item was the right one was very time consuming to say the least. According to staff at NULI there was not a backlog of publications waiting to be catalogued; such a situation would have made matters right in the end, when all had been catalogued. The only explanation remaining was that the publications of these important institutions had simply not been delivered to NULI. Among the reasons for this development was that the staff of the two institutions claimed not to have been aware of their legal duty to deliver their publications to NULI.

Discussion

The aim to provide the public with access to important publications was apparent in the *Act on Reading Societies and Educational Films no. 57/1937* (reading societies were the forerunners of public libraries in Iceland); and subsequently in acts on public libraries up to 1976 (the first was the *Public Libraries Act no. 42/1955*). According to a stipulation in these acts government publications were delivered for free to reading societies and public libraries receiving public support. One of the requirements for public support of a reading society was that it had at least 10 (ten) members and that the local people contributed equal funds to the operation of the reading society. Hence government publications should have been widely available to the public.

Up to the time when technological developments came to facilitate production of publications directly by the issuing bodies, all publications in Iceland, apart from the black were white according to Wessels definitions (1997), in the sense that they were catalogued in The *Icelandic national bibliography* and available for consultation by the public at The National Library in Reykjavík, the three Quarter Libraries in the west, north, and east of the country, and quite possibly at the issuing bodies themselves. Users would, however, have had to go these places to get hold of the GL. The advent of e-publishing was expected to make access to GL easier through dissemination on home pages of the issuing bodies (Farace and Schöpfel, 2010; Wessels, 1997); in particular the access to contemporary GL (Júlíusdóttir, 2006). It could hardly be expected that the same GL would be accessible on the home pages forever.

The finding of the 2002 study that close to half of all publications catalogued in Gegnir were GL in Iceland in 1944, and that this percentage of GL of the whole of knowledge output indicated that legal deposit deliveries were in relatively good order. The Icelandic situation certainly was quite different from that in some other countries, judging by the estimates that abroad GL made up 9% of publishing by the end of the 9th decade and that this had risen to over 20% by the middle of the 10th decade of the past century (Debachere, 1995). It would seem that these percentages were bound to be an underestimation when compared to the percentage of GL in Iceland, which was a rather backwards country in 1944 with few research institutions and only one university, which mostly had the role of educating public officials. When that situation changed the percentage of GL went up above 60 in 1979.

The increases in the numbers of higher educational institutions and increased research activity in the country, is manifested in the increase in GL (although less than expected after 1979), publications created in Iceland retrieved in WoS, and patent applications. Contrary to expectations, access to GL issued in Iceland did not improve subsequent to the *Legal Deposit Act no. 20/2002* taking force January 1st 2003. The reason why staff of the issuing bodies of GL was unaware of their duty to deliver their publications to NULI, may be that NULI had not informed sufficiently on the duty to deliver legal deposit copies. Although the 2002 act covered a greater number of media than the 1977 act, its scope is narrower than that of previous such acts in Iceland, in the sense that in the area of e-publishing it only stipulates legal deposit of internet publications and not intranet publications. This results in a situation where the same publication is to be delivered on legal deposit if published on paper or disseminated on the internet, but not if disseminated on an intranet. In Iceland publications on research work are typically disseminated as GL during developmental stages many a time only on intranets and the final results are published on the international scholarly market. The effects of the *Legal Deposit Act no. 20/2002* are that entire research projects may be carried out in Iceland without delivery of and hence access to any publication on them being available in Iceland to authorities and the public presently or in the future. Further research is needed to establish the effects of this stipulation in the 2002 act.

Conclusions

This development calls for an amendment to the *Legal Deposit Act no. 20/2002*. Its revision should be based on the traditional spirit of access of the public and authorities to publications issued in Iceland, GL as well as market publications. As this was written e-repositories are operated in the country. Rafhladan.is is an e-repository operated by NULI covering a wide subject area. Examples of other repositories are the institutional repositories skemman.is for The Agricultural University of Iceland, The Iceland Academy of the Arts, The National and University Library, The University of Akureyri, The University of Bifröst, The University of Iceland and The University of Reykjavik. Skemman.is houses students' digital theses and dissertations in addition to articles and other research material from the universities' academic staff. Hirsla.is is a repository for The National University Hospital. Such repositories are far from covering all the GL that should be available for access in Iceland, and the bibliographic access to items disseminated there is limited. The access to the publications housed there has to be catalogued bibliographically and be linked to the national union catalogue to be most useful. Rules have to be set on which items to preserve for the future as it is highly unlikely that all can be maintained for future use, as was the case when fewer items were published.

References

Auger, C. P. (1996 (1994)). *Information sources in grey literature* (3. edition). London: Bowker.

Bazerman, Charles. (1988). *Shaping written knowledge: The genre and activity of the experimental article in science*. Madison: University of Wisconsin Press.

Burke, Peter. (2002). *A social history of knowledge: From Gutenberg to Diderot: Based on the first series of Vonhoff Lectures given at the University of Groningen (Netherlands)* (Reprinted). Cambridge: Polity Press.

Chalabi, Lydia. (2012). Characteristic and use of grey literature in scientific journals articles of Algerian Researchers: Case study of University of Science and Technology Houari Boumediene (Physics, Chemistry and Computer Sciences). In Farace, Dominic and Frantzen, J. *GL14 Conference Proceedings*. Amsterdam: TextRelease, 2013.

Chaves, Todd A. (2010). Grey literature in Karst research: The evolution of the Karst information portral (KIP). In Farace and Schöpfel (editors), *Grey literature in library and information studies* (pp. 181-197). Berlin: Walter de Gruyter GmbH & Co. KG.

Debachere, M-C. (1995). Problems in obtaining grey literature. *IFLA journal, 21* (2), 94-98.

Diodato, Virgil P. (1994). *Dictionary of bibliometrics*. New York: Haworth Press.

Einarsson, Eiríkur Þ. (1990). Nýting erlendra gagnabanka. [Use of foreign bibliographic databases]. In *Upplýsingar eru auðlind* (pp. 259-269). Reykjavík: STUPP.

Eyþórsdóttir, Hildur G. (1997). Íslensk bókaútgáfa í 30 ár, 1967-1996. [Icelandic publishing during 30 years, 1967-1996]. *Ritmennt, 2*, 76-80.

Farace, Dominic J. and Schöpfel, Joachim (Eds.). (2010). *Grey literature in library and information studies*. Berlin: De Gruyter Saur.

Fjöldi útgefinna bóka á íslensku. (2006, April 26th). *Landskerfi bókasafna. Fréttasafn*. Retrieved May 12th 2006 at: http://www.landskerfi.is/sja.php?id=206.

Gentil-Beccot. (2010). The driving and evolving role of grey literature in high-energy physics. In Farace and Schöpfel (editors), *Grey literature in library and information studies* (pp. 155-166). Berlin: Walter de Gruyter GmbH & Co. KG.

Georgescu-Roegen, Nicholas. (1999 (1971)). *The entropy law and the economic process*. Cambridge, Mass.: Harvard University Press.

Grey Net International. (2006). *Grey literature network service*. Retrieved August 3rd 2006 at: http://www.greynet.org/.

Hálfdanarson, Guðmundur, Matthíasdóttir, Sigríður and Guðmundsson, Magnús. (2011). *Aldarsaga Háskóla Íslands 1911-2011*. [Centennial history of the University of Iceland]. Reykjavík: Háskólaútgáfan. Karlsson, Gunnar (Ed.).

Hauksdóttir, Sigrún (2001). *Consortium of Icelandic libraries*. Retrieved February 14th 2006 at: http://www.bokis.is/iod2001/papers.html.

Hauksdóttir, Sigrún. (1996). Gegnir. [Gegnir: The automated library system of the National and University Library]. *Bókasafnið, 20*, 64-67.

Harðarson, Hrafn. (1993). Tölvuvæðing í bókasöfnum á Íslandi. [Automation in Icelandic libraries]. *Bókasafnið, 17*, 17-22.

Hertzel, Dorothy. (1987). Bibliometrics, history of the development of ideas. In *Encyclopedia of library and information science* (Vol. 42, supplement 7, pp. 144-219). New York: Marcel Dekker.

Hjartar, Ólafur F. (1968). Íslenzk bókaútgáfa 1887-1966. [Publishing in Iceland 1887-1966]. In *Landsbókasafn Íslands. Árbók 1967. [National Library of Iceland. Yearbook 1967*

Júlíusdóttir, Stefanía. (1989). Plans for co-operative automation in Icelandic libraries. *Scandinavian Public Library Quarterly, 22*(1), 21-26.

Júlíusdóttir, Stefanía. (1991). The state of automation in Icelandic libraries: Past, present and future: An overview. *In OPAC´s and the User: Proceedings of the third Anglo-Nordic Seminar* (pp. 59-62). Esbo: Nordinfo.

Júlíusdóttir, Stefanía. (1996). Tölvuvæðing íslenskra bókasafna: Könnun á stefnumörkun og stöðu mála. [Automation in Icelandic libraries: Policies and practices]. *Bókasafnið, 20*, 54-63.

Júlíusdóttir, Stefanía. (2006). Changes in publishing and LIS work in Iceland. In Hauksson, Úlfar (Ed.), *Rannsóknir í félagsvísindum VII: Félagsvísindadeild* (pp. 65-78). Reykjavík: Félagsvísindastofnun Háskóla Íslands.

Lariviere, Jules. (2000). *Gudelines for legal deposit legislation* (a revised, enlarged and updated editon of the 1981 publication by Jean Lunn IFLA Committee on Catloguing). Paris: IFLA. Retrieved January 4th 2012 at http://archive.ifla.org/VII/s1/gnl/legaldep1.htm.

Lúðvíksson, Vilhjálmur. (2002). *[Presentation given at the Annual Meeting of the Icelandic Research Council, by Vilhjálmur Lúðvíksson the Managing Director of the Icelandic Research Council].* Paper presented and distributed on paper at the Annual Meeting of the Icelandic Research Council 2002, Reykjavík.

MacDonald, Bertrum H., Wells, Peter G., Cordes, Ruth E., Hutton, Gregory R.G., Cossarini, Danielle M., and Soomai, Soonai S. (2010). The Use and Influence of Information Produced as Grey Literature by International, Intergovernmental Marine Organizations: Overview of Current Research. In Farace and Schöpfel (editors), *Grey literature in library and information studies* (pp. 167-180). Berlin: Walter de Gruyter GmbH & Co. KG.

Machlup, Fritz. (c1984). *The economics of information and human capital* (Vol. 3). Princeton: Princeton University Press.

National and University Library of Iceland. (2009a, July). *Fréttir 29.7.2009: Vefsafn opnað fyrir prófanir. [News 29.7.2009: Web collection opens for trial*]. Retrieved April 27th 2011 at: http://landsbokasafn.is/index.php/news/128/15/Vefsafn-opnad-fyrir-profanir.

National and University Library of Iceland. (2009b, September). *Fréttir 29.9.2009: Vefsafn opnað. [News 29.7.2009: Web collection opens*]. Retrieved April 27th 2011 at: http://landsbokasafn.is/index.php/news/135/15/Vefsafn-opnad.

National and University Library of Iceland. (n.d.). *Rafhlaðan. [Collection of electronic publications issued in Iceland].* Retrieved May 25th 2013 at: http://rafhladan.is/.

National and University Library of Iceland. (2011). *Vefsafn. [Web collection].* Retrieved April 27th 2011 at: http://vefsafn.is/.

National Library of Iceland. National Department. (1975). *Icelandic national bibliography 1974*. Reykjavík: National Library of Iceland.

National Library of Iceland. National Department. (1980). *Icelandic national bibliography 1979*. Reykjavík: National Library of Iceland.

Pritchard, Alan. (1969). Statistical bibliography or bibliometrics? *Journal of Documentation, 25*(4), 348-349.

Rögnvaldardóttir, Guðrún. (2007). Staðlaráð tvítugt, eða þannig: Brot úr sögu Staðlaráðs. [The Standards Organisation becomes twenty]. *Staðlamál, 11*(3), 2.

Skúlason, Páll. (2000). *Vísindi og fræði á aldamótaári. [Science and research at the millennium*. Presentation delivered on January 26th 2000]. Retrieved January 26th 2000 at: http://www.visindiogfraedi.hi.is/pallsk.html.

Skýrr. (2006). *Skýrr*. Retrieved July 20th 2006 at: http://www.skyrr.is.

***Skýrsla** fyrir Félag íslenskra bókaútgefenda: Áfangaskýrsla.* (2003). [*A Report for the Association of Icelandic Publishers: An interim report*]. Reykjavík: Instutute of Economic Studies at the University of Iceland.

Thorhauge, Jens, Larsen, Gitte, Thun, Hans-Peter and Albrechtsen, Hanne. (1997). *Public libraries and the information society. Luxembourg*: European Commission.

Wessels, R. H. A. (1997). Het belang en toegankelijkheid van grijze literatuur [The importance and accessibility of grey literature]. Informatie professional, 1(3), 28- 30, 32.

Scholarly publishing behaviour in Slovakia … Are we ready for repositories?

Jaroslav Šušol, Miriam Ondrišová, Jana Ilavská, and Marcela Kopecká,
Comenius University in Bratislava, Faculty of Philosophy, Slovak Republic

Abstract
Institutional repositories are regarded as one of major conceptual and technological revolutions in the field of scholarly publishing, enabling quick and widespread dissemination of research results. In order for the repositories to become this useful and powerful tool, it is necessary that they get full political and financial support from the academic institutions' managements as well as a wide acceptance from the academics, university teachers and researchers.
At the Comenius University in Bratislava, the oldest and the largest university in Slovakia, we are preparing a project of new institutional repository. In order to find out about the prospects of such an endeavor, we plan to make an enquiry on its feasibility from both the above-mentioned points of view – 1. state of the art of repositories implementation in Slovakia and 2. the acceptance of electronic publishing channels within the complex of manifestations of publishing behavior.
One of the principal approaches towards research in publishing behavior is the application of quantitative methods of scientometrics or bibliometrics that are usually put into a broader context. Scientometrics is not a new discipline, but very current at the moment in Slovakia, vividly discussed in the academic circles due to the fact that some of its methods and indicators are being recently used as an important part of university management and of allocation of financial resources from the level of ministry of education.
We will apply both sociological and scientometrics methods on publishing data from Comenius University in our research. It should lead to defining the patterns of publishing behavior of authors in the academic environment, with respect to the particularities in humanities and social sciences, natural sciences and technical disciplines.

Keywords: institutional repository, publishing behavior, electronic publishing, online publishing, bibliometrics, scientometrics

1. Introduction

Research data and research papers are frequently cited as some of primary examples of grey literature. Their accessibility is often limited to a low number of copies, especially, of course, in case of printed volumes. Institutional repositories are regarded as one of major conceptual and technological revolutions in the field of scholarly publishing, enabling quick and widespread dissemination of research results. In order for the repositories to become this revolutionary tool, it is necessary that they get full political and financial support from the academic institutions' managements as well as a wide acceptance from the academics, university teachers and researchers.
At the Comenius University (CU) in Bratislava we have started a project of building a scientific park in the domain of molecular medicine, environmental medicine and biotechnologies in 2013. Several supporting project activities are devoted to social and humane aspects of biotechnologies (legislation, ethics, economics). One of these socio-cultural streams of research is focused on analysis of users and usage contexts of institutional repositories and creation of the CU institutional repository model.
In the framework of this research goal we plan to make an enquiry on its feasibility from both the above-mentioned points of view, and to look at:

1. state of the art of repositories implementation in Slovakia,
2. the acceptance of electronic publishing channels within the complex of manifestations of publishing behavior, including the researchers' willingness and readiness to share data and resources.

In this paper we present the results of some of our pilot research efforts into the domain of publishing behavior and repositories in Slovakia. Preliminary characteristics of publishing culture/reality in different scientific disciplines, or rather broader discipline areas, were compiled on the basis of bibliographic data contained within CU Library database of CU employees' publications. Application of quantitative methods of scientometrics or bibliometrics can provide us with an exact picture of distribution of individual bibliographic parameters within the database and define some of the features of publishing behavior at Comenius University.

2. Institutional repositories in Slovakia

Institutional repository is a complex, multi-structured device whose primary role lies in collecting, archiving and disseminating digital publications and data coming from a particular, usually institutionally defined environment. This phenomenon is closely connected with such concepts and technologies as scholarly communication and publishing, digital library and open access.

Traditional approach to classification of repositories takes into account various aspects of their content and its geographical limits, recognizing e.g. subject-based repositories, research repositories, national repositories or institutional repositories. Predominantly, building of institutional repositories is connected with academic institutions where repositories are designed to hold the full-texts of the complete intellectual output of institutions' individual members. In this respect, institutional repository fulfills one more important role – the role of presentation, marketing the institution in the digital space.

Institutional repositories and open access (OA) can have various practical manifestations, depending mostly on the nature of material that is being archived and disseminated which, consequently, has an impact on its economic and, possibly, legislative functioning. Two principal streams of open access publishing have been defined as the Gold model and the Green model (Craig 2007, Houghton 2013). The Gold model is based on a traditional journal as a vehicle for publication, with possible changes in the financial model, which is dominantly saturated by authors or sponsors. Within the Green model, the authors usually post their manuscript into an electronic archive (institutional or subject-based), as a pre-print (in the form that was submitted to a journal) or as a post-print (a final copy of the peer-reviewed article).

From the very beginning of the development in the area of online publishing and repositories it was clear that one of the main advantages of these new technologies was the increased visibility, or rather, better findability of the material and publications that were published digitally on the net. Some of the early research results in this area (Lawrence 2001, Harnad 2004) even showed that there is a correlation between free online availability of publications and higher citation counts.

More recent studies do not see this relation so clearly. Some authors have identified three possible reasons that could account for the observed citation differences between OA and non-OA articles. They were termed an *Open Access postulate* (authors are more likely to read, and thus cite, articles available freely), a *Selection Bias postulate* (the most prominent, and thus most citable, authors are more likely to make their works freely available), and an *Early View postulate* (the period between the early posting of an article, either pre-print or post-print, and the appearance of the cognate published journal article allows for earlier accrual of citations) (Craig 2007).

We looked through the state of building of the institutional repositories in Slovakia via questionnaire survey within research project VEGA 1/0429/10 *Information ecology of the academic information environment* (Steinerová 2012). Questionnaire had been addressed to all Slovak universities, colleges and the Slovak Academy of Sciences (total number of addressees - 43). We received 27 completed questionnaires (response rate 69.2%). All significant academic institutions had been among respondents. Our findings were not very positive, only 14 institutions reported to have some form of institutional repository - 2 institutions have been building standalone institutional repository and 13 institutions are using local storage with web access to cover some of repository functions (one of the institutions has both, standalone repository as well as local storage). Twelve institutions have not any kind of repository yet, but they are planning to build it in relatively short time. One institution has neither repository nor plans to build it. It is interesting, that in majority of cases (18) it was the library (often in cooperation with IT department) which initiated or is initiating the creation of repository within institution, while administration and management did not show a lot of interest in this matter. The survey detected the most common problems and barriers that institutions used to face, among them especially copyright and other related legislation, lack of cooperation among relevant departments within institution (management, library, IT department, archive, publishing house, etc.), reluctance to share research data out of official and reputable publication channels, individual traits (computer skills, age).

In Slovakia, the state of the art of repositories was predominantly influenced by the Act on Higher Education No. 131/2002, and its subsequent amendments. Since 2010, due to the efforts connected with building a system of plagiarism/originality control, all the universities in Slovakia are obliged to collect all qualification works and final theses (bachelor, master, dissertation, etc.) in electronic format and send them into a state-wide Central Register of Theses (http://www.crzp.sk). Similarly, each university has to build a database focused on bibliographical registration of publications of its employees. This database, however, is not supposed to contain digital copies of the publications (full texts). Again, all the data from individual databases are merged into the Central Register of Publications (http://www.crepc.sk).

These two types of registers that are being built at all universities in Slovakia can be seen as a good starting point in all efforts and plans to develop institutional repositories. Although the register of publications at each university usually contains only bibliographic descriptions of books and articles published by university researchers, each academic library is obliged to collect hard copies of these documents (full text in case of articles, principal pages in case of books, proceedings and collections of papers). So, the organizational system of getting hard copies of documents is in place, the task now should be to turn it into digital environment.

3. Publishing behavior

Publishing behavior represents an important part of *information behavior* of information users. While information behavior in the broadest sense is usually defined as a complex of various ways of interaction between a human as an actor of information processes and the information, the publishing behavior is a specific component of this phenomenon, that occurs on the production side of information and communication spectrum. It is embodied in various ways and approaches that the authors apply when it comes to publishing, and especially to selecting and using diverse communication and publication channels.

It is possible to explore the publishing behavior using various approaches, quantitative and qualitative methods, analyzing the subjective interpretation of reality among authors themselves (sociological or ethnological methods based on questionnaires, surveys or interviews) or measuring hard data in the form of publication outputs, production. Although there are researchers in this area who like to stress that there are some methodological drawbacks to these enquiries. For example, Michels and Schmoch (2013) stipulate that "it has been confirmed in the social sciences that subjects under observation change their behavior, making the genuine observation of their behavior almost impossible. This phenomenon, called Hawthorne effect, is more pronounced in cases where the results of the observation have consequences for the observed subjects".

One of the most important impulses for the development of publishing behavior research arrived in 1960s, with the application of computing technology into the processes of secondary information storage and retrieval. Research projects concentrated mostly on specific conditions and problems of particular scientific discipline (Cox 1993, Luukkonen 1992, Swan 2008, Kyvik 2003, Rey 1998, Rockwell 2000, Björk 2000 etc.).

The situation in economics since 1970s was analyzed by Cox (1993). He presented various aspects that can be studied in connection with the production and distribution of economics literature, like contributions of individual institutions to the economics journals and their impact on the development of their reputation, searching for „doctoral roots" of individual authors, relative quality of journals and their influence on economics literature. As Cox pointed out, „the knowledge of the publication behavior of individuals can serve many useful purposes. For instance, if it is assumed that this behavior of individuals is stable over time, it will help assess the likelihood of multiple publications by individuals in the future. In addition, such knowledge provides valuable information about the implicit strategies of economic researchers in selecting their research outlets, the behavioral aspects of joint work (i.e., co-authored paper), and the relationship between the publication count and an author's stage in the career cycle." (Cox 1993).

At the beginning of 1990s, Luukkonen (1992) sought to verify the assumption that in their publishing behavior the scientists tend to select prestigious publishing channels, seeking an optimal level in the hierarchy of publications. The interviews with researchers in three areas (zoology, biomedicine, automation and control technology) suggested that when it comes to selecting the publishing channel, the Finnish scientists value equally the criteria of reward and communication functions of publishing. The real publishing behavior of biomedical scientists, however, shows their leaning towards higher appreciation of publication prestige (Luukkonen 1992). More recent research from Great Britain suggests that the majority of scientists prefer the journal titles with relevant reader community. This choice is not always in line with the highest impact factor journals in a given discipline. Publishing in journals with high impact factor usually brings advantages in the form of higher formal criteria research assessment, however, the reach to the right audience brings reward in terms of recognition by peers (Swan 2008).

There were also attempts to map the development of publishing behavior on a wider interdisciplinary scale. Kyvik's research focused on assessing the evolution of publishing behavior within 2 decades and analyzed changes in publication patterns at the universities in Norway. Based on 3 questionnaire surveys from 1982, 1992 a 2001 he concluded that co-authorship became a more common way of publishing, the extent of publishing oriented towards an international audience has increased, the scientific article in an international journal consolidated its position as a dominant type of publication and the number of publications per academic researcher increased. One of the most intriguing conclusions states that the

publication patterns in social sciences get closer to those in natural sciences (Kyvik 2003). A similar research, although restricted to the field of agronomy, was carried out in Spain. The analysis of articles published in national and international journals between 1980 and 1995 revealed 2 leading tendencies – migration of works towards SCI journals and increased use of books and monographs as channels of publication of research works (Rey 1998).

In 2000, trustworthiness of electronic publishing was put into focus in the research that took place at the universities in Canada. The results showed that only 16% of respondents published their articles in a peer-reviewed electronic resource, while as many as 61% used other, traditional communication channels (Rockwell 2000). The Swedish survey from the same year concentrated on the situation in building engineering and concluded that both the students and professors consider reviewing to be an important part of electronic publishing processes. When comparing traditional and electronic regimes of publishing, 75% of respondents said that the speed of publishing on the web was more important for them than sticking to formal publishing procedures (Björk 2000).

The literature, however, is not always absolutely positive about electronic publishing, it also reflects the drawbacks of using new technologies for the purposes of scholarly publishing. As the Nature points out - „Scientific productivity, as measured by scholarly publication rates, is at an all-time high. However, high-profile cases of scientific misconduct remind us that not all those publications are to be trusted – but how many and which papers? Given the pressure to publish, it is important to be aware of the ways in which community standards can be subverted. Our concern here is with the three major sins of modern publishing: duplication, co-submission and plagiarism.“ (Errami 2008).

First major survey of this type in Slovakia was done in 2002 by means of a questionnaire among the users of academic libraries. It focused, among other things, on the authors' relationship towards traditional and electronic publishing technologies. The results supported the assumption that the acceptance of electronic communication channels for publishing the output of scientific research depends on particular discipline, with strong unbalance between social sciences and humanities on one side and natural and technical disciplines on the other (Šušol 2004).

A combination of qualitative and quantitative methods was applied in the survey of publishing preferences of PhD students at the Faculty of Arts, Comenius University in Bratislava. One of the key issues was the authors' attitude towards traditional and electronic publishing. The conceptual analysis of the background of individual author's attitudes or choices, revealed 3 most frequent categories – *time*, *availability* and *reputation.* First 2 categories (time and availability) are being mostly connected with electronic publishing. The results also indicate that the refusal of electronic publishing channels is caused rather by their non-acceptance in the authors' closest peer environment, not that much by the fact that the authors themselves would not realize the overall circumstances, advantages and disadvantages of electronic publishing (Steinerová et al. 2006).

The above-mentioned examples and results demonstrate that the publishing behavior is a multi-dimensioned research domain, as the behavior of information process actors is influenced by various factors. The attitudes, opinions, preferences, that together define publishing behavior of authors in academic area, are not only subject to evolution in time, but have also certain local or national particularities.

One of the principal approaches towards research in publishing behavior, of course, is the application of quantitative methods of scientometrics or bibliometrics that are usually put into a broader context. We will eventually apply both sociological and bibliometrics methods on publishing data from Comenius University in our research. It should lead to defining the patterns of publishing behavior of authors in the academic environment, with respect to the particularities in humanities and social sciences, natural sciences and technical disciplines. There are several variables that can be calculated and compared in this way, e.g.:

- publishing in foreign language / Current Contents Connect (CCC) journals
- usage of foreign (or foreign language) resources / electronic resources
- types of publications
- co-authorship
- methods and quantity of citing, referring
- attitude towards publishing in the network environment / open access regime.

At the moment, though, we will only present results of some of our preliminary analyses, aimed at finding connections between electronic publications, or rather electronic/online versions of publications on one side, and their availability and citation rate, on the other.

4. Comenius University database of publication outputs analysis

Bibliographical registration of publication outputs of academic staff in Slovakia is regulated by Library Act No. 183/2000 Coll. and Act No. 455/2012 on Higher Education Institutions and Directive of Ministry of Education No. 456/2012 Coll. Registration process distinguishes between different types of publications (monograph, chapter, journal article, review, conference paper, proceedings, etc.) and uses different criteria (originality, territoriality, communication channel, target audience, etc.) to categorize publication outputs within overall 83 categories. Some of the categories are considered to be of higher quality or value (e.g. A – scientific works), as they publish the results of original scholarly research, usually validated by the quality of publication channel (CCC journal, reviewed collection of papers etc.). Bibliographic processing of publications includes also registration of citations to the publications (excluding authors' self-citations), there are 8 categories of citations (e.g. foreign/domestic registered in citation indexes in Web of Science or Scopus, not registered in citation indexes, reviews, art critiques).
The pilot phase of bibliometric analyses was carried out at the Comenius University (CU) in Bratislava on the data of publications register that is available via iPortal of the CU Academic Library (http://alis.uniba.sk:8000/cgi-bin/gw_49_3_8/chameleon?skin=epc).
In October 2013, the CU database of publication outputs, which has been systematically built since the beginning of 1990s, contained 226 453 records of various types of publications and 212 787 citations. The distribution of records among faculties in the whole retrospect of the database is as in Table 1.

Table 1 CU database of publications outputs overview

Faculty[1]	Number of records/publications	Number of citations	Number of cited publications
FMED	50 388	49 873	8 696
JFMED	19 340	22 380	3 698
FPHARM	13 415	12 872	2 574
FLAW	8 991	5 464	1 563
FPHIL	39 311	35 529	8 758
FNS	38 674	72 923	11 379
FMPH	21 022	41 288	4 726
FSPORT	11 094	12 923	2 744
FEDU	17 036	12 334	3 050
FEVTH	2 439	545	244
FRCTH	4 898	2 269	885
FM	3 990	3 242	975
FSEV	2 529	1 250	643
CU Total*	**226 453**	**212 787**	34 968

* Total number of publications/citations is not sum of faculties because of co-authorship among authors from different CU faculties.

From the digital repository point of view, we were interested in the typological structure of publications. We focused especially on publications which were published recently - during the period of years 2011-2013 (till October 2013). Within these 3 years CU authors published 25 281 publications (about 9.000 publications per year), in the structure described in Table 2.
When we analyzed given data by faculties, we had noticed differences in preferred publication channels between scientific disciplines (distinction between hard and soft sciences), see Graph 1. It has been proved, in compliance with similar analyses published abroad, that social sciences and humanities' outputs differ from those from natural sciences, medicine and technology. We can observe higher preference of monographic literature (scientific monographs, textbooks, chapters) in soft sciences; hard sciences mostly prefer scientific journals as communication channels. Quite interesting is the distribution of conference papers, which represent 22% of soft sciences publications and only 8% of hard sciences publications.

[1] CU faculties (13): Faculty of Medicine, Jessenius Faculty of Medicine in Martin, Faculty of Pharmacy, Faculty of Law, Faculty of Arts, Faculty of Natural Sciences, Faculty of Mathematics, Physics and Informatics, Faculty of Physical Education and Sport, Faculty of Education, Evangelical Lutheran Theological Faculty, Faculty of Roman Catholic Theology of Cyril and Methodius, Faculty of Management, Faculty of Social and Economic Sciences

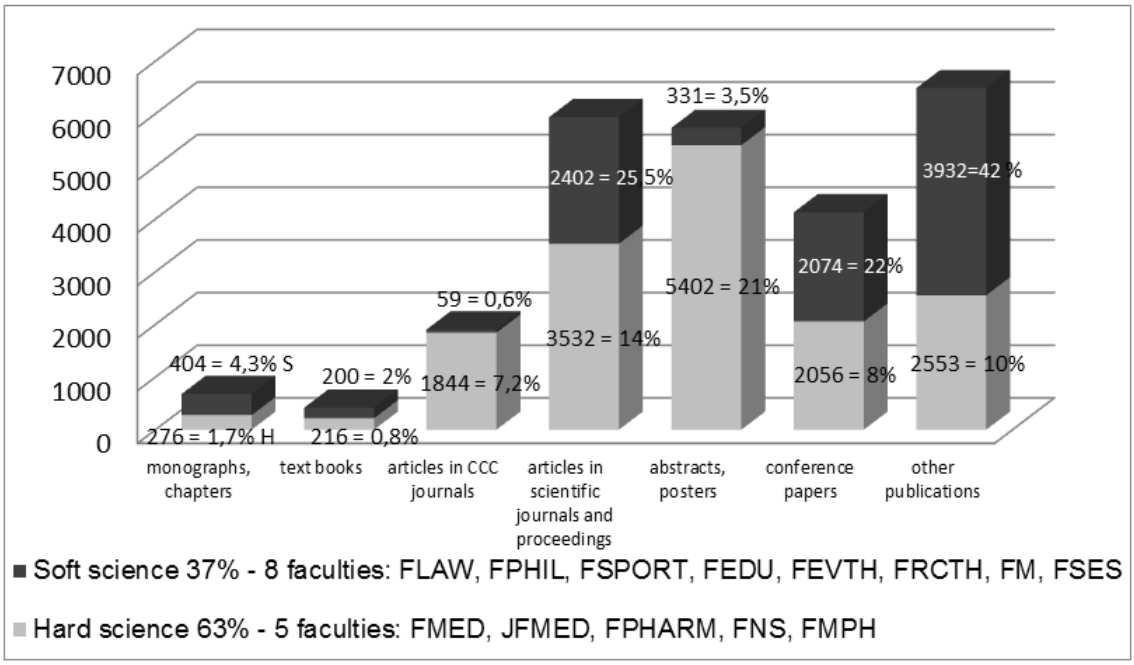
Table 2 CU publications 2011-2013 – typological structure

Typological structure of CU publications	Number of publications	% of all publications
monographs, chapters	688	2.69%
text books	441	1.65%
articles in CCC journals	1 911	7.53%
articles in scientific journals and proceedings	5 958	23.47%
abstracts, posters	5 649	22.68%
conference papers	4 120	16.34%
other publications	6 514	25.65%
Total	**25 281**	100%

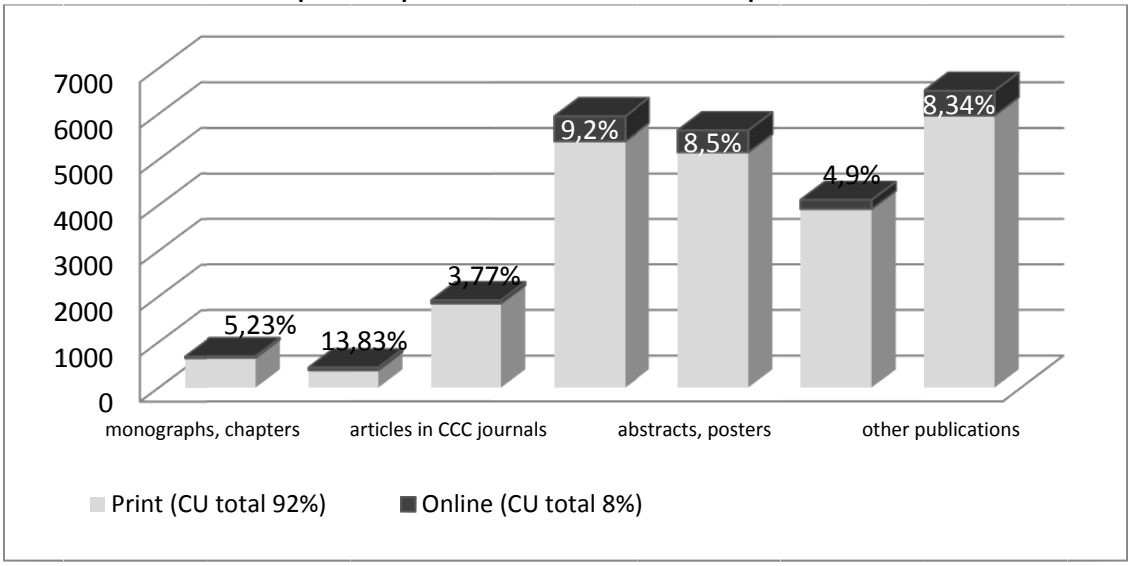
Furthermore, we wanted to know the percentage of publications published online in electronic form, as these could be directly used for repository needs (after solving copyright matters, of course). Only 8% of CU publications from 2011-2013 had been published online (for their distribution among faculties see Graphs 3, 4 and Table 3).

Our analyses showed that Faculty of Mathematics, Physics and Informatics (FMPH), Faculty of Arts (FPHIL) and both medical faculties (FMED, JFMED) belong among faculties with highest rate of online publications. In case of FMPH it is certainly also due to quite long tradition of publishing preprints via well-known specialized web portals such as arxiv.org and cern.ch. Medical faculties publish a lot of abstracts which are often available via different web sites, medical portals and databases (PubMed, bmj.sk). FPHIL online publications are represented especially by domestic conference proceedings, peer-reviewed collections of papers and articles in Slovak online journals.

Graph 1 CU publications 2011-2013: different faculties = different publication channels



Graph 2 CU publications 2011-2013: online publications



Graph 3 CU publications 2011-2013: online publications by faculties

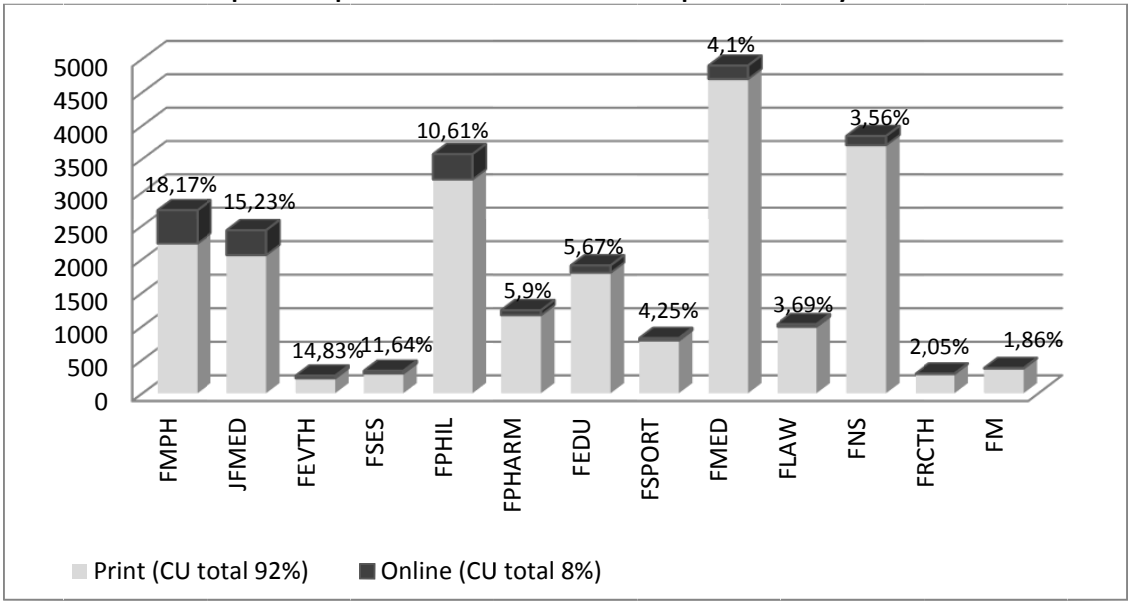


Table 3 CU publications 2011-2013: share of online publications by faculties

CU faculties	Number of online publications	% of all online publications
FMPH	496	25,73%
FPHIL	379	19,66%
JFMED	370	19,19%
FMED	201	10,43%
FNS	137	7,11%
FEDU	108	5,60%
FPHARM	73	3,79%
FSES	39	2,02%
FEVTH	39	2,02%
FLAW	38	1,97%
FSPORT	35	1,82%
FM	7	0,36%
FRCTH	6	0,31%

Almost 17% of all CU online publications are accessible from CU domain (uniba.sk) and its component parts, they are distributed on 48 various locations (e.g. stella.uniba.sk, motiv.fns.uniba.sk, ii.fmph.uniba.sk, moodle.uniba.sk, etc.). In addition, CU operates several local storages with purpose to store and make available different documents in electronic form, especially theses (> 52 400), digitalized collections of papers published by CU (> 400), textbooks (> 50). Such distribution of documents complicates their discoverability as well as preservation. And we managed to prove, that better visibility and discoverability of publications achieved by online publishing channels could increase average citation number. It is obvious especially in following publication categories: scientific domestic conference proceedings and collections of papers; articles in domestic journals, textbooks, articles in CCC journals, articles in foreign scientific journals.

5. Conclusions

Establishing an institutional repository is still quite complicated in Slovak academic institutions, although the benefits of the repository for academic individuals (e.g. increased dissemination and impact of the research outputs, provision of archiving) as well as for the institution itself (effective marketing, better support of education and research, central archives) are well known.

Comenius University in Bratislava feels the need to build complex institutional repository, as up to now it had been using only locally distributed storages enabling access to some of its electronic publications. Institutional repository could be built within CU scientific park and then it could serve as an institutional showcase for research. But effective operation of the repository requires first of all well thought out institutional information strategy, collaboration between researchers/authors, departments, faculties, libraries and other component parts of university. Last but not least it is necessary to deal with copyright and legislation problems.

Acknowledgements

This work was supported by the Slovak Research and Development Agency under the contract No. APVV-0208-10, the project of the “University Scientific Park” of Comenius University in Bratislava ITMS 26240220086 and project KEGA 133UK-4/2013 MIKS - Mediálne a informačné kompetencie pre znalostnú spoločnosť.

References

BJÖRK, Bo-Christer and Ziga TURK. 2000. A Survey of the Impact of the Internet on Scientific Publishing in Construction IT and Construction Management. In: *The Journal of Information Technology in Construction (ITcon)* [online]. Rotterdam: International Council for Research and Innovation in Building and Construction, 2000, vol. 5, pp. 73-88 [ret. 2013-11-04]. ISSN 1874-4753. Available from: http://www.itcon.org/2000/5/paper.htm

BUTLER, Linda. 2008. Using a balanced approach to bibliometrics: quantitative performance measures in the Australian Research Quality Framework. In: *Ethics in Science and Environmental Politic. Ethics Sci Environ Polit [online]*. Oldendorf (Luhe): Inter-Research. June 2008, vol. 8, no.11, pp. 83-92, 2008 [ret. 2013-11-04]. ISSN 1611-8014. Available from: http://www.int-res.com/articles/esep2008/8/e008p083.pdf

COX, Raymond A. K., CHUNG, Kee H. and Albert A. OKUNADE. 1993. Publishing behavior of individuals and most prolific authors in the economics literature. In: *Quarterly Journal of Business and Economics* [online]. Lincoln: University of Nebraska-Lincoln College of Business Administration. June 1993, vol. 32, no. 3, pp. 32-42 [ret. 2013-11-04]. ISSN 1939-8123. Available from: http://www.allbusiness.com/finance/404806-1.html

CRAIG, Iain D., PLUME, Andrew M., MCVEIGH, Marie E., PRINGLE, James and Mayur AMIN. 2007. Do Open Access Articles Have Greater Citation Impact? A critical review of the literature. In: *Journal of Informetrics* [online]. Amsterdam: Elsevier BV. July 2007, vol. 1, iss. 3, pp. 239-248 [ret. 2013-11-04]. ISSN 1751-1577. Available from: http://dx.doi.org/10.1016/j.joi.2007.04.001

ERRAMI, Mounir and Harold GARNER. 2008. A tale of two citations. In: *Nature* [online]. London: Nature Publishing Group, 2008, vol. 451, pp. 397-399 [ret. 2009-08-20]. ISSN 0028-0836. Available from: http://www.nature.com/nature/journal/v451/n7177/full/451397a.html

HARNAD, Stevan and Tim BRODY. 2004. Comparing the impact of Open Access (OA) vs. non-OA articles in the same journals. In: *D-Lib Magazine* [online]. Reston: Corporation for National Research Initiatives. June 2004, vol. 10, no. 6 [ret. 2013-11-04]. ISSN 1082-9873. Available from: http://www.dlib.org/dlib/june04/harnad/06harnad.html

HOUGHTON, John and Alma SWAN. 2013. Planting the Green Seeds for a Golden Harvest: Comments and Clarifications on "Going for Gold". In: *D-Lib Magazine* [online]. Reston: Corporation for National Research Initiatives. January/February 2004, vol. 19, no. 1/2 [ret. 2013-11-04]. ISSN 1082-9873. Available from: http://www.dlib.org/dlib/january13/houghton/01houghton.html

KIMLIČKA, Štefan. 2001. Význam spracovania a prezentácie publikačnej činnosti na vysokej škole - úloha akademických knižníc. In: *Zborník z odborného seminára Evidencia publikačnej činnosti, citácie a medzinárodné štandardy, ktorý sa uskutočnil 30. januára 2001 v Nitre* [online]. Nitra: Slovenská poľnohospodárska knižnica pri SPU v Nitre. February 2001 [ret. 2013-11-04]. Available from: http://www.slpk.sk/eldo/epc/1.htm

KYVIK, Svein. 2003. Changing trends in publishing behavior among university faculty, 1980-2000. In: *Scientometrics* [online]. Budapest: Akademiai Kiado Rt. September 2003, vol. 58, iss. 1, pp. 35-48 [ret. 2013-11-04]. ISSN 1588-286. Available from: http://www.springerlink.com/content/t779563774467186/

LAWRENCE, Steve. 2001. Free online availability substantially increases a paper's impact. In: *Nature.* May 2001, vol. 411, no. 6837, pp. 521. ISSN 0028-0836.

LUUKKONEN, Terttu. 1992. Is scientists' publishing behavior rewardseeking? In: *Scientometrics* [online]. Budapest: Akademiai Kiado Rt. June 1992, vol. 24, no. 2, pp. 297-319 [ret. 2013-11-04]. ISSN 1588-286. Available from: http://www.springerlink.com/content/r5583r778831t1t4/

MICHELS, Carolin and Urlich SCHMOCH. 2013. Impact of bibliometric studies on the publication behaviour of authors of authors. In: *Scientometrics* [online]. Budapest: Akademiai Kiado Rt. April 2013 [ret. 2013-11-04]. ISSN 1588-286. Available from: http://link.springer.com/article/10.1007/s11192-013-1015-7

Predpis č. 456/2012 Z. z. Vyhláška Ministerstva školstva, vedy, výskumu a športu Slovenskej republiky o centrálnom registri evidencie publikačnej činnosti a centrálnom registri evidencie umeleckej činnosti.

REY, J., MARTÍN, M.-J., PLAZA, L., IBÁÑEZ, J. J. and Isabel MÉNDEZ. 1998. Changes in publishing behavior in response to research policy guidelines. The case of the Spanish Research Council in the field of agronomy. In: *Scientometrics* [online]. Budapest: Akademiai Kiado Rt. January 1998, vol. 41, Iss. 1-2, pp. 101-111 [ret. 2013-11-04]. ISSN 1588-286. Available from: http://www.springerlink.com/content/q744436514g14743/

ROCKWELL, Geoffrey and Lynne SIEMENS. 2000. *The Credibility of Electronic Publishing : Report on Responses to the Questionnaire* [online]. Nanaimo: Malaspina University. 2000 [ret. 2013-11-04]. Available from: http://web.mala.bc.ca/hssfc/Final/QuestionnaireR.htm

SKALKA, Ján and Libor VOZÁR. 2008. Centrálny register publikačnej činnosti. In: *ITlib. Informačné technológie a knižnice* [online]. Bratislava: Centrum VTI SR. 2008, no. 2, pp. 20-24 [ret. 2013-11-04]. ISSN 1336-0779. Available from: http://itlib.cvtisr.sk/archiv/2008/2/centralny-register-publikacnej-cinnosti.html?page_id=1279

STEINEROVÁ, Jela and Jaroslav ŠUŠOL. 2005. Library users in human information behavior. In: *Online information review* [online]. Bingley: Emerald Group Publishing Ltd. 2005, vol. 29, no. 2, pp. 139-156 [ret. 2013-11-04]. ISSN 1468-4527. Available from: http://www.emeraldinsight.com/journals.htm?articleid=1464932&show=html&ini=CNKI

STEINEROVÁ, Jela, ŠUŠOL, Jaroslav and Mirka GREŠKOVÁ. 2006. Information behavior in relevance judgements. In: *Využívanie informácií v informačnej spoločnosti.* Bratislava: Centrum VTI SR, 2006, pp. 29-40. ISBN 80-85165-92-9.

STEINEROVÁ, Jela (edt.). 2012. Informačná ekológia akademického informačného prostredia. Bratisalva: Univerzita Komenského, 2012. ISBN 978-80-223-3178.

SWAN, Alma. 2008. *Key Concerns Within The Scholarly Communication Process. Report To The Jisc Scholarly Communications Group* [online]. Truro: Key Perspectives Ltd, March 2008, 65 p. [ret. 2013-11-04]. Available from: http://www.jisc.ac.uk/media/documents/aboutus/workinggroups/topconcernsreport.doc

ŠUŠOL, Jaroslav. 2004. Publikačné preferencie autorov vo vzťahu k tradičným a elektronickým zdrojom - situácia v oblasti vedeckej komunikácie. In: *Knižničná a informačná veda 20 = Library and information science 20*. Bratislava: Univerzita Komenského, 2004, pp. 37-57. ISBN 80-223-1993-7.

Zákon č. 183/2000 Z. z. o knižniciach, o doplnení zákona Slovenskej národnej rady č. 27/1987 Zb. o štátnej pamiatkovej starostlivosti a o zmene a doplnení zákona č. 68/1997 Z. z. o Matici slovenskej v znení neskorších predpisov.

Zákon č. 455/2012 Z.z. Zákon, ktorým sa mení a dopĺňa zákon č. 131/2002 Z. z. o vysokých školách a o zmene a doplnení niektorých zákonov v znení neskorších predpisov, a ktorým sa mení zákon č. 455/2004 Z. z. o zriadení Akadémie ozbrojených síl generála Milana Rastislava Štefánika, o zlúčení Vojenskej leteckej akadémie generála Milana Rastislava Štefánika v Košiciach s Technickou univerzitou v Košiciach, o zriadení Národnej akadémie obrany maršala Andreja Hadika a o zmene a doplnení niektorých zákonov v znení zákona č. 144/2008 Z. z.

An International Journal on Grey Literature, 2005-2014

ISSN 1574-1796 (Print) ISSN 1574-180X (PDF/Online Password)

Tenth Anniversary Special Offer!

TGJ Volume 10, Issues 1-3, 2014	☐**Print** ☐**PDF** ☐**Password = €240 subscription**
TGJ *27 Thematic Issues* 2005-2013	☐**Print** ☐**PDF** ☐**Password = €540 set price**
TGJ Back Issues 2005-2013	☐**Print** ☐**PDF = € 80 per issue**

TGJ Volume 9, Number 1, Spring 2013	☐ Adapting New Technologies for Grey Literature
TGJ Volume 9, Number 2, Summer 2013	☐ Tracking Grey Literature Across Disciplines
TGJ Volume 9, Number 3, Autumn 2013	☐ Improving Grey Literature through Innovation
TGJ Volume 8, Number 1, Spring 2012	☐ Social Networking and Grey Literature
TGJ Volume 8, Number 2, Summer 2012	☐ Data Frontiers in Grey Literature
TGJ Volume 8, Number 3, Autumn 2012	☐ Managing Change in Grey Literature
TGJ Volume 7, Number 1, Spring 2011	☐ Transparency in Grey Literature
TGJ Volume 7, Number 2, Summer 2011	☐ System Approaches to Grey Literature
TGJ Volume 7, Number 3, Autumn 2011	☐ Research and Education in Grey Literature
TGJ Volume 6, Number 1, Spring 2010	☐ Government Alliance to Grey Literature
TGJ Volume 6, Number 2, Summer 2010	☐ Shared Strategies for Grey Literature
TGJ Volume 6, Number 3, Autumn 2010	☐ Research on Grey Literature in Europe
TGJ Volume 5, Number 1, Spring 2009	☐ Paperless Initiatives for Grey Literature
TGJ Volume 5, Number 2, Summer 2009	☐ Archaeology and Grey Literature
TGJ Volume 5, Number 3, Autumn 2009	☐ Trusted Grey Sources and Resources
TGJ Volume 4, Number 1, Spring 2008	☐ Praxis and Theory in Grey Literature
TGJ Volume 4, Number 2, Summer 2008	☐ Access to Grey in a Web Environment
TGJ Volume 4, Number 3, Autumn 2008	☐ Making Grey more Visible
TGJ Volume 3, Number 1, Spring 2007	☐ Grey Standards in Transition and Use
TGJ Volume 3, Number 2, Summer 2007	☐ Academic and Scholarly Grey
TGJ Volume 3, Number 3, Autumn 2007	☐ Mapping Grey Resources
TGJ Volume 2, Number 1, Spring 2006	☐ Grey Matters for OAI
TGJ Volume 2, Number 2, Summer 2006	☐ Collections on a Grey Scale
TGJ Volume 2, Number 3, Autumn 2006	☐ Using Grey to Sustain Innovation
TGJ Volume 1, Number 1, Spring 2005	☐ Publish Grey or Perish
TGJ Volume 1, Number 2, Summer 2005	☐ Repositories - Home2Grey
TGJ Volume 1, Number 3, Autumn 2005	☐ Grey Areas in Education

Check the appropriate box(es) above, complete your name and address information. Return via fax, email, or post. An invoice will then be forwarded directly to you!

Your Name, Address, and Email

TextRelease, Javastraat 194-HS, 1095CP Amsterdam NL, T/F +31(0)20.3312420, info@textrelease.com

BDSP: a unique initiative to archive and disseminate French grey literature on public health

François Petitjean, Marie-Laure Charron, Séverine Ferron, and Christiane Stock
La Banque de données en santé publique and INIST-CNRS, France

Abstract
BDSP (Public Health Databank) is a network of institutions initiated 20 years ago by the French ministry of health. Its aim is to collect and provide access to documents on public health topics, with specific attention to grey literature published in French.
40 organizations participate in this open and collaborative network, including the stakeholders of the French health system and from other French-speaking countries. It is managed by a team of 4 members of the EHESP School of Public Health.
The network produces a multiservice portal which includes a bibliographic database with more than 470,000 records and 64,000 fulltext documents online. It also offers a multilingual glossary and a thesaurus in public health. The website receives about 3000 visits per day.
BDSP is a unique product by its "construction", its contents and its scope. Its flexible organization has allowed for many adaptations and extensions over the past 20 years, aiming to cover all fields on public health. The bibliographic database offers access to an important number of grey documents published by the main French organizations involved in public health, research units or government agencies, at both a national and regional level. The documents include expert's reports, summary reports, policies, theses and dissertations. Participation in the BDSP network allows its members to archive their grey publications and data, to increase their visibility, to share current awareness tasks and exchange bibliographic records or documents as well as best practices. Working in a network further helps to deal with obstacles in managing grey literature such as insufficient recognition of the documents compared to journal articles, difficulties to obtain authorizations for the dissemination (in particular for government agencies) and to keep up with the rapid evolution of knowledge.
At present the network explores ways to extend its cooperation with other French speaking countries and to share grey literature in French.

BDSP – a network

BDSP (Public Health Databank) is a network created by information professionals in 1993 at the request of the French ministry of health in order to provide access to the documentation on public health and especially to grey literature in French. The network is composed of 40 data producers, including stakeholders from the French health system and other French-speaking countries: hospitals, universities, NGO's, research institutions, state agencies. It is open and collaborative, guided by formal agreements and strongly based on reciprocity of input and services.

Reciprocal exchange of services is the driving force for input to be added to the portal to keep it going, services such as the creation of records on documents and events or monitoring hundreds of serials. In return its members benefit from document archiving facilities, increased international visibility, access to all other input, citable records, forums for sharing technical expertise, shared tools, etc.

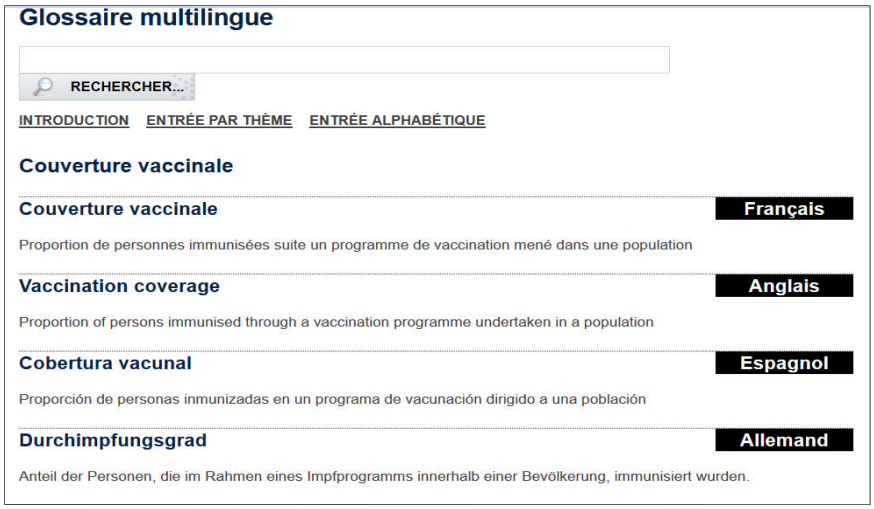
Two committees (librarians of the network and heads of the member institutions) are running the BDSP network.

A team of 4 people from the EHESP School of Public Health animate the network and design and update the portal. The team covers skills in information science, computer programming and public health.

bdsp.ehesp.fr – the multiservice information portal

The BDSP network produces a portal providing a variety of services to researchers and practitioners in public health. The most prominent service is its bibliographic database with over 473,000 records and 64,000 full-text documents (14%)[1]. 400 collections of serials and monographs are examined on a regular basis.

The multilingual glossary[2] was developed by a European expert group in 1996 at the initiative of the European Community. Its present version includes 400 concepts in public health and is available in 4 languages (English, French, German, and Spanish) on the BDSP website since 2003.



BDSP also developed its own thesaurus[3] on public health in 1992[4] and has updated it on a regular basis. Its latest edition (2007) includes 12,825 terms (7144 descriptors), grouped in 57 microthesaurus subsets. It has been mapped to other thesauri and controlled vocabularies (e.g. PASCAL). The graphic display of the online version allows easy navigation between terms and its relations. Hyperlinks make it simple to find the corresponding records in the bibliographic database. The thesaurus can be downloaded on request and free of charge and can be included in other databases.
The reviewed directory[5] of websites on public health is maintained by the members of the network. Created in 2001, it includes more than 1000 entries on resources, websites and stakeholders in public health in France and in Europe. Descriptions range from simple references to detailed information on a site, following pre-established criteria. The user interface provides access through simple search as well as browsing by themes.
The portal further comprehends dedicated areas for job postings (including internships and calls for project), for conferences and meetings announcements, a blog and a newsfeed. BDSP users thus have free access to an important amount of information on public health in different forms.

The bibliographic database

The BDSP database has over 473,000 bibliographic references, with an annual increase of about 15,000 records.[6] Input is added by its members through file transfer on a regular basis. BDSP uses an in-house format for its database, but provides conversion interfaces for each partner for the import and export of records. Journal article references (many of them in English) make up the majority of the entries.
Each record includes a bibliographic description, an abstract (mostly in French), keywords according to the BDSP thesaurus as well as French and English keywords from the controlled PASCAL[7] vocabulary. Further information is displayed with hyperlinks, facilitating new searches on authors, corporate authors, document types, and to access the full text, if available. It is possible to start with a term in the thesaurus and to list the corresponding records. Search facilities include simple search as well as advanced search criteria such as keywords, periodical, publication dates, document types, etc. Initially broad searches can be narrowed down with the help of a word cloud or with different filters. It is possible to limit search to records with full text attached. The user interface provides features to export the records (500 entries maxi per export) to Reference Manager or to Zotero, in different formats.

Grey literature in the bibliographic database

The BDSP database provides access to an important number of unique grey documents produced by the main French stakeholders in public health. In November 2013 grey literature represents 11% (or 51,000 documents) of the bibliographic records. More than 2000 records are added every year. The following document types are included:

Type	Subtype	Percentage
Reports	Policy, expert, inspectorate, official documents	46.5%
Academic degrees	Theses, dissertations, doctoral theses	42%
Conferences	Proceedings, papers, oral presentations	11.5%

Grey documents are mostly written in French (89%), being produced in a large part by the members of the network. The main producers of grey literature are universities (theses), government agencies (expertise reports) and national or regional health observatories.

All French doctoral theses with a topic on public health are referenced in the database by a university member of the network. In addition, about 6400 master theses[8] from the EHESP School of Public Health are included.. From 1999 onwards only master theses selected by the EHESP jury and for which the author has given his agreement are available in full text.
About 2600 full text documents are available.
26% of the grey literature citations in the BDSP database provide an access to the full text (13,500 records). The rate for full text related to grey literature is far higher than for other document types (12%). The document is either deposited in the BDSP archive, for those members who don't have their own repository, or a link to a distant repository is added in the metadata record. The table below shows the importance of the BDSP archive as a "home" for grey literature.

Type	Nb of records	Full text available	FT deposit in the BDSP archive	Link to FT on adistant server
Grey literature	51487	26.2%	5167	8366
Other	421862	11.8%	10444	39235

Difficulties with regard to grey literature (GL) and leads for progress
Several obstacles appear with regard to archiving issues and the dissemination of grey literature:
Collecting grey documents requires the identification and mobilization of GL producers, as opposed to journals material. This applies in particular to reports. Librarians of the regional health agencies for example point out their difficulties in obtaining their in-house production.
This lines up with the poor recognition of grey literature or factual data by the health professionals.
Due to ignorance of intellectual property rights, many services and agencies won't disseminate their production to a wider public, but keep it in intranets in order to be on the safe side. The same difficulties apply with authors when librarians try to obtain agreements to disseminate the full-text.
The grey literature typology used in the database could be more detailed in order to avoid cataloguing errors and to improve consistency (e.g. expert's reports with ISBN).

In spite of these difficulties, several opportunities for progress emerge:
- The development of sites dedicated to certain publics or professional groups, providing added value to their productions and the possibility to share them,
- The reference to the national archives and to the obligation to deposit documents produced by the national services,
- The development of « best practices in public health » should be an important source for productions to be published,
- Working on the typology of grey documents may result in specific actions according to the documents and the users.

Network organization, added value and return on investment
The BDSP network is unique by its organization. The open and collaborative network allows for efficiency and flexibility. Pooling and sharing production of services which are useful for all have made it possible to extend and adapt the BDSP portal for the past 20 years. Today it covers all fields on public health (52 topics are identified/used). Meetings and the participation in an electronic forum enable exchanges of professional practices, highly appreciated by the information professionals. The network adds value to in-house productions and events, and furthers the dissemination of information to its member structures. The members of the network are the stakeholders for the smooth working of the BDSP. The success and long existence of the network is closely linked to the fact that members contribute and get a return on investment in different ways. As mentioned before, the bibliographic database offers archiving facilities for the documents of its producers, providing added value services such as perennial links to the full text and thus the citability of documents.
This is even truer for grey literature: 38% of the full text grey documents are deposited in the BDSP archive. Access to grey literature is thus made easier for the member organizations of the network.
International visibility is increased by providing keywords in both French and English. The database is indexed by Google. In exchange for providing input to the database the members could, according to the agreements, receive the same number of records produced by other members.

Usage of the bibliographic database and the website
The BDSP website is well frequented. In 2012 an average of 3000 visits per day (1 million per year) were made by 1500 users per day; the average visit took about 3 minutes and 17 seconds with 5 page views. The bibliographic database is the most frequented part of the portal: 1173 visits per day in 2012 by 860 visitors (i.e. 39% of the total visits).

It is difficult to establish a typology of users due to a lack of recent data. Network members think however that public health professionals are the most frequent users, including when they are still studying or in professional training. BDSP users are located in France (87%) or in other French-speaking countries (6%). Specific data on the usage of grey literature is not available.

Perspectives

The BDSP must develop constantly in order to maintain its interest for its users. One way to evolve is to increase its coverage and to fetch new collaborations. Indeed BDSP intends to open its network to all French speaking countries. Swiss and Moroccan Institutions are already members. Several ways of achieving this are being explored:

The network seeks to associate the documentation centres in public health from African countries. At present 16,000 records in the bibliographic database deal with African topics.

The BDSP portal could take into consideration the specific health problems of emerging or developing countries and add areas dedicated to a specific country to the portal.

It is also necessary to promote the BDSP to health professionals and researchers in other French-speaking countries (e.g. a cooperation project with the « Santécom » database of Quebec).

Thus BDSP reinforces its role as federator and aggregator for scientific and professional information in the public health domain and as main access point for French language grey literature on this topic.

Network members currently discuss a new project to adapt the services offer in order to better answer the needs of health professionals, researchers and students. Working on the interface between data and users should enable information to be transferred into knowledge and skill development.

Several ideas are being explored:

- Create dedicated areas for geographic areas (French regions, French-speaking countries)
- Elaborate and disseminate thematic syntheses,
- Inform on training courses, link to distant learning courses,
- Provide areas for best practice guides and contribute thus to continued training, linked to the concept of evidence-based public health.

References

VIGNON (Marie-Edith). Banque de données en santé publique : un réseau d'information sur la santé publique. – In : LA SANTE DE L'HOMME 2010/11-12;(410):35-. http://www.bdsp.ehesp.fr/Fulltext/432226/

DENEVERT (Daniel). Le réseau BDSP à l'horizon 2014 : quel avenir pour les portails d'information en santé publique ? 2009/01. http://www.bdsp.ehesp.fr/Fulltext/395524/

VIGNON (M.E.). La recherche d'information en santé publique : la Banque de données en santé publique (BDSP). - Revue médicale de l'assurance maladie 2001/04-06;32(2):161-166

VIGNON (Marie-Edith). Des outils d'indexation variés pour alimenter en réseau une base de données, l'exemple de la BDSP. - DOCUMENTALISTE. SCIENCES DE L'INFORMATION 2007;44(1):93-95. http://www.bdsp.ehesp.fr/Fulltext/356079/

VIGNON (Marie-Edith). Banque de données en santé publique : un réseau d'information sur la santé publique. - LA SANTE DE L'HOMME 2010/11-12;(410):35-. http://www.bdsp.ehesp.fr/Fulltext/432226/

PETITJEAN (François) CHARRON (Marie Laure), FERRON (Séverine), MENARD (Daniel) ; La banque de données en santé publique : comment répondre aux besoins d'information des professionnels et des chercheurs ? Poster, 16ièmes Journées des sciences de la santé, Bobo-Dioulasso, Burkina Faso, mai 2012

[1] Data as of November 2013
[2] http://asp.bdsp.ehesp.fr/Glossaire/
[3] http://asp.bdsp.ehesp.fr/Thesaurus/
[4] http://fulltext.bdsp.ehesp.fr/Adbs/DSI/2007/44/93-95.pdf
[5] http://asp.bdsp.ehesp.fr/Webs/
[6] Data as of November 2013
[7] PASCAL : database produced by Inist-CNRS (http://www.inist.fr)
[8] See here for a list of diplomas : http://www.ehesp.fr/formation/

Technology Transfer Support on National Level

Ľubomír Bilský and Miroslav Kubiš
Slovak Centre of Scientific and Technical Information, CVTI SR, Slovakia

Introduction

One of the recent tasks and responsibilities of the Slovak Centre of Scientific and Technical Information (SCSTI) has been to provide technology transfer support to public research organisations on a national level.

For this purpose, the SCSTI has been mandated by the Ministry of Education, Science, Research and Sport of the Slovak Republic to implement a national project, co-financed from the EU structural funds within the Operational Programme Research and Development, entitled the National Infrastructure for Supporting Technology Transfer in Slovakia – NITT SK. The project is focused on building and securing the operation of the national system supporting the technology transfer processes in Slovakia and is implemented over the period June 2010 – December 2014 with a total budget of €8.2 million.

The transfer of knowledge from science to industry, for the benefit of society, will be supported through the complex mechanism, which is currently being developed. A key player in the whole system will be the *National Technology Transfer Centre*, which will co-operate closely with local technology transfer centres established at universities and public research institutions. An important part of the system will be the *National patent fund,* from which financial support for the processes of protecting and commercialising intellectual property will be provided to public research institutions. The financial support will include payments for patent-filing and maintenance (PCT, USPTO, European and national applications), patent attorneys services, patent translations, partner search, marketing, negotiating and licensing. Even at the current moment, patent-filing fees are covered from the state budget in money administered by the SCSTI, while external support services are provided within the NITT SK project. This is a great benefit for the Slovak public and state universities and the Slovak Academy of Sciences, which have been facing a lack of financial means for these purposes.

Concerning the technology transfer activities that are being provided through the NITT SK project, these cover the entire process of transferring scientific knowledge into the practice, starting from intellectual property protection and ending with its commercial utilisation. Services are provided by external high-level experts on relevant technology transfer issues from the respective scientific field. All costs are covered from the NITT SK budget and are free to the scientific community. Specific services include the assessment of intellectual property commercialisation potential, state of the art analyses, patent application preparation and filing, technology marketing, partner searches, negotiation support, preparation of licensing agreements, spin-offs creation, follow-up of royalties' payment and so on. Since 2012, when this support started, some patent applications have already been supported within the NITT SK project.

Last year, the *National portal for technology transfer* was launched within the NITT SK project. This provides a comprehensive overview of technology transfer issues providing services to researchers, students, companies and local TT offices. Useful items provided include sample agreements, internal guidelines and other methodological materials related to technology transfer. It also affords a list of competencies of Slovak R&D institutions and links to relevant Slovak, foreign and international organisations, groups and initiatives.

1. Technology transfer

An important part of the research activities at universities and R&D institutions is finding applications for their inventions and outcomes from research in practical life for the benefit of the society and for the benefit of the institutions and individuals involved in the research. Society can benefit from these inventions e.g. through new jobs created, improved standard of living, health and social care, protection of the environment and cultural heritage and knowledge gained. The benefits for research institutions and researchers could be of an immaterial character (e.g. partnering with private sector companies and collaboration with other research institutions at national or international level, enhancing their reputation with professionals and the general public, increase in the number of students enrolled to study at universities involved in these activities, competitive advantage) and of a material character (e.g. leveraging resources, extending research budgets as a result of participation in research programmes funded by the private sector, additional income for research institutions and inventors from patented and licensed technologies).

The transmission of the new knowledge latent in an invention into practical life proceeds via a somewhat complex process of technology transfer. The definitions and concepts of technology transfer have been discussed in many different ways and they are surrounded by many different interpretations

and views depending on the institutions' objectives, research background, stakeholders, research areas and disciplines and underlying perspective (theories) [1,2]. One straightforward definition is that technology transfer is the process of transferring scientific findings from one organisation to another for the purpose of further development and commercialisation [3].

2. Brief outline of background situation in technology transfer in Slovakia

Technology transfer, albeit not known under this name, existed in the former Czechoslovakia. Within its external limitations it was relatively well-established with a network of specialised departments, particularly in research institutes, providing support services. The Slovak Technical Library in Bratislava, the predecessor of the Slovak Centre of Scientific and Technical Information, was the sole institution in Slovakia with a complete collection of Czechoslovak patent documents and a large collection of patent documents from abroad. That this institution celebrates its 75th anniversary this year simply has to be stated at this occasion.

The situation in technology transfer from research institutions changed substantially in Slovakia under the new economic conditions following the establishment of the independent Slovak Republic in 1993. Since then, there has been a need for a systematic and modern approach to technology transfer, in particular from the state-subsidised research institutions (public and state universities, the Slovak Academy of Sciences and public research institutes), in order to make use of the public resources in the most effective manner. The academic community has always been a boundless source of knowledge, experience and skills, imagination and invention, curiosity and speculation, enthusiasm and dedication. However, these positive features, to some extent at least, have been overshadowed by the fact that the awareness of intellectual property rights within the academic community is somewhat low, laws related to intellectual property have frequently been breached with no apparent subsequent consequences and patent application and patent maintenance fees are too high (min €5,365.5 per 20 years) for research institutions to cover; in addition, investigators have tended to work individually rather than institutionally for the private sector who has the funding means and there has been a lack of information and motivation on the part of inventors to be more involved in the research in which they would function more as representatives of a research institution rather than as individuals; also the infrastructural pre-requisites for conducting high-quality research (e.g. instrumentation, software) have been poor. As a result, actual technology transfer has been scarce as also have technology transfer offices and the number of production units as potential collaborators or recipient of applied research outputs have been reduced. The NITT SK project has been put in place to enhance this situation.

3. National Infrastructure for Supporting Technology Transfer in Slovakia – NITT SK project

The major features outlined in chapter 2 characterised the status in technology transfer prior to implementation of the National Infrastructure for Supporting Technology Transfer in Slovakia – NITT SK, the national project, in June 2010. The project with a total budget of €8.2 million is co-financed from the EU structural funds within the Operational Programme Research and Development. Its implementation period is from June 2010 up to December 2014.

Within the implementation of this project, the SCSTI as the principal investigator of the project, has to address and minimise the negative features summarised above, support development of the existing positive features and create the conditions which will facilitate intellectual property protection, technology transfer in its complexity and support the establishment of technology transfer centres or similar departments at those universities that have a high potential for technology transfer. These are important organisational units within the structure of research institutions, since technology transfer is costly, time-demanding and requires expertise in product or technology development, regulatory approval, marketing and sales, all of which are activities different from those expected of researchers.

It is not a coincidence that the SCSTI was mandated by the Ministry of Education, Science, Research and Sport of the Slovak Republic and has been viewed as the institution with the full competence to complete the project successfully. There has been a long-standing tradition in the SCSTI in specialising in patent and corporate literature and other categories of grey literature [4] and also in providing some librarian services related to intellectual property protection and technology transfer, e.g. conducting preliminary state-of-the art search of patent documents and grey literature, translating technical and specialist texts, consulting with and advising inventors on issues of an administrative character and collaborating with the Industrial Property Office in Slovakia. The staff have been trained and educated in these particular fields. They are experienced specialists and some members of the staff are in training as patent attorneys. A more detailed account of their activities can be found in [5]. Some specific tasks, however, as they are defined in the NITT SK project, can only be fulfilled by externally contracted professionals.

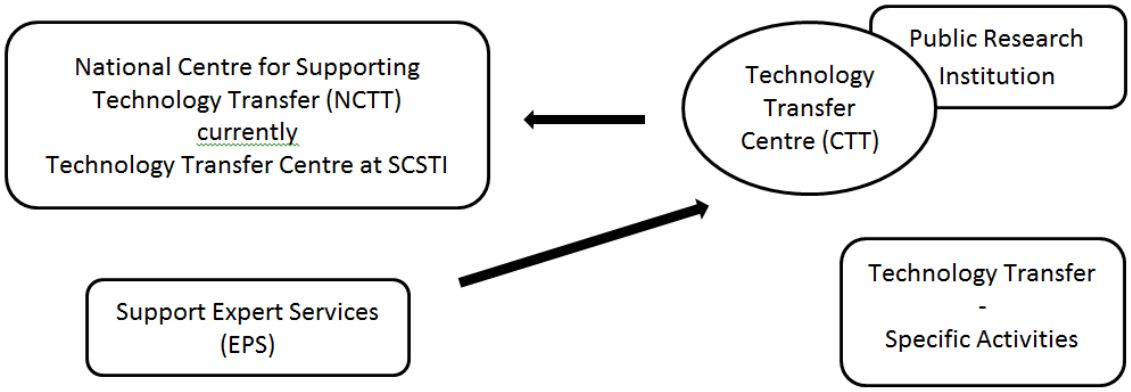
The strategic goal of the NITT SK project is to establish and implement the National System for Supporting Technology Transfer from state-funded research institutions (public and state universities, the Slovak Academy of Sciences and public research institutes) which are the target group of the project. Three specific objectives are fundamental to achieving this strategic goal.

4. NITT SK – Specific objective 1

4. 1 National System for Supporting Technology Transfer in Slovakia and National Technology Transfer Centre

Specific objective 1 is to establish and operate the Technology Transfer Centre at the SCSTI, i.e. to provide the Technology Transfer Centre at the SCSTI with human resources and operational capacities and provide auxiliary services in technology transfer to research institutions across the whole of Slovakia. This is the core objective directly related to technology transfer. It creates the conditions and environment for technology transfer in Slovakia to be a meaningful, successful and sustainable process. Its major role is to design and put into operation a mechanism for the *National System for Supporting Technology Transfer in Slovakia (NSPTT)* [6]. The key subject in the NSPTT will be the *National Technology Transfer Centre (NCTT).* The idea is that the NCTT within the NSPTT will provide public research institutions with all the support services needed to protect intellectual property originating from public research institutions (not individuals) and the subsequent commercialisation of legally protected intellectual property. The NCTT will collaborate closely with local technology transfer centres established at public research institutions. The concept of the NCTT is currently being developed, its most appropriate legal form considered, and its statutes, job positions and processes proposed. The intention is that some of the support expert services will be provided by the NCTT staff in collaboration with a provider, a successful applicant following a public procurement process. The NCTT will also be responsible for maintaining the *National Patent Fund* from which financial support will be channelled to cover patent-filing and maintenance fees (PCT, USPTO, European and national applications), fees for services of patent attorneys, patent translations, partner search, marketing, negotiating and licensing. It is expected that, in addition to the state subsidy, the NCTT members will contribute a certain sum from their royalties to this fund. All issues concerning the NCTT and National Patent Fund are still under consideration. At this stage, some of the tasks of the NCTT are being fulfilled by the Technology Transfer Centre at SCSTI which was established on 21st March 2011.

A schematic representation of activities and relations between the National Technology Transfer Centre, Technology Transfer Centre at SCSTI and local Technology Transfer Centres at public research institutions is provided in the diagram below:



4. 2 National Portal for Technology Transfer

The third component is the *National Portal for Technology Transfer* (NPTT) which came into operation in 2012. It aims to be a comprehensive resource of information and issues related to intellectual property protection and technology transfer, serving researchers, students, companies and local technology transfer offices alike.

4.2.1 NPTT – access point to support services

The NPTT is an important channel through which the specific support expert services (EPS) can be ordered and communicated (see also the diagram above). The EPSs form a bundle of highly specialised services of two types – those related to intellectual property protection and those related to technology transfer. Examples of such services are advising on strategy for intellectual property protection, assessment of intellectual property commercialisation potential, contract-drafting and negotiation, state-of-the art analyses, patent application preparation and filing, technology marketing, partnering, setting up spin-off companies, follow-up of royalties' payment.

To date, 42 requests for support expert services have been submitted from the Slovak Academy of Sciences, Comenius University, Žilinská University in Žilina, the Slovak University of Technology in Bratislava, the Technical University in Košice, the Academy of Fine Arts in Bratislava and the Animal Production Research Centre in Nitra. The services in most frequent demand are assistance with selecting the industrial property of the research institution's portfolio which is suitable for protection (comprising commercial potential assessment, identification of industrial property, industrial property protection feasibility study), preparation and filing of patent and utility model applications, drafting and assessment of agreements, advising in strategy for intellectual property commercialisation and partnering. Due to their highly specialised character, these services are outsourced. The group of services provided by SCSTI specialists comprises search in technological databases (ad hoc or monitoring), search in intellectual/industrial property databases (ad hoc or monitoring, state-of-the art, patentability, competitors, citation, grey literature, etc.) and consultation and assistance in these areas. The first two services are completed by dispatching the respective report to a researcher.

All types of these services are provided to public research institutions without charge with the expert services following the conclusion of an Agreement for the Provision of Support Expert Services. Currently, agreements with six universities, six institutes of the Slovak Academy of Sciences and one agreement with a public research institute have been concluded. The Agreements with three more universities are in the post-negotiation phase shortly prior to concluding. Some EPS services (e.g. patent-filing and maintenance fees) are funded from a special item in the state budget allocated to the SCSTI outside of the NITTSK project budget. To date, patent applications originating from the Slovak Academy of Sciences and one pending patent application from the Technical University in Košice, have been funded from this source.

4.2.2 NPTT – source of template documents

The content of the NPTT website is varied; it covers all possible topics in technology transfer and intellectual property and some categories of grey literature. Templates such as *sample agreements* http://nptt.cvtisr.sk/sk/poskytovane-vzorove-materialy/dohody-zmluvy.html?page_id=539:

- Confidentiality (Non-disclosure) Agreement
- Consulting Agreement
- Research Contract
- Licence Agreement

and sample *internal guidelines*:
http://nptt.cvtisr.sk/sk/poskytovane-vzorove-materialy/smernice.html?page_id=538

- Guideline on Shares in Intellectual Property
- Intellectual Property Fees Guideline
- Guideline on Intellectual Property Administration and Documents Handling
- Intellectual Property Guideline
- Industrial Property Protection Guideline
- Guideline on Protection of Non-registered Industrial Property Exempt from Author´s Rights
- Author's Rights Guideline
- Non-disclosure and Confidentiality Guideline
- Intellectual Property Licensing Guideline
- Guideline on Intellectual Property Assignment Agreement

can be accessed and downloaded from the NPTT website.

4.2.3 NPTT – source of methodological materials

The NPTT is also intended as a point of education providing access to methodological materials which are another example of the grey literature category. Up to the present, the following researcher's guides are available:

- Intellectual Property and Confidentiality Guide [7]
- Establishment and Operation of University's Technology Transfer Centre Including a Proposal for a Standard Model of Operation [8]
- Academic Materials and Publishing Guide (in preparation) [9]
- Research Contracts Guide (in preparation) [10]

and a series of four educational brochures entitled *Intellectual Property and Technology Transfer* from specialist seminars [11-14].

4.2.4 NPTT – access point to databases and information resources

The NPTT provides access to various databases (e.g. Intellectual Property Offices world-wide, databases of products and services, databases of technologies, experts in technology transfer), information

resources (e.g. collections of books on technology transfer and intellectual property and related topics, the purchase of which is funded from the NITT SK projects; journals and e-books from ebrary). Access is also provided to basic information on the research competences of Slovak public research institutions intended especially for networking purposes.

5. NITT SK – Specific objective 2

The aim of Specific objective 2 is to propose, set up and administer for the scientific community an Integrated System of Services (ISS) using information and communication technologies to control access to special scientific databases, integrated software applications and other electronic information resources and services which promote and facilitate technology transfer.

The scientific community will be provided with easy, user-friendly and efficient access via a common platform of Integrated System of Services to high-capacity data storage and high-speed hardware facilities equipped with modern scientific software applications so the researchers can process data from their experiments and store large data sets in a safe environment with guaranteed maintenance. Scientific software applications are purchased by the SCSTI or by the research institutions themselves. Currently, the MATLAB computing software is available to the scientific community. Research-related databases administered by the SCSTI, e.g. SK CRIS (Slovak Current Research Information System) [15], CREPČ (Central Register of Publication Activity), CREUČ (Central Register of Artistic Activity), SciDAP (Central Databases of Slovak Electronic Information Resources on Research and Development), CRZP (Central Repository of Theses and Dissertations) [16] will be integrated in the ISS in due course.

6. NITT SK – Specific objective 3

Specific objective 3 is to raise awareness in the scientific and academic community of intellectual property rights protection and technology transfer and also of the importance of popularising science. Activities within this objective include:

- participation in organising various events of an educational character, such as training, seminars and conferences for representatives of research institutions, scientists, inventors, PhD students and students at universities where they can acquire the basics of intellectual property rights and their protection, technology transfer and present their achievements in applied research,
- dissemination of information on the current status and news in technology transfer and intellectual property rights in Slovakia and in the world by editing and publishing e-TTbulletin. It aims to cover many aspects of the dynamically developing area of technology transfer. This journal is the only e-journal of its kind in Slovakia. The electronic version of the bulletin is available on http://ttb.cvtisr.sk/bulletiny-z-roku-2013.html?page_id=106. The printed version of the bulletin, which is published quarterly, contains some selected and summarised articles from e-bulletin http://nptt.cvtisr.sk/sk/ttb-transfer-technologii-bulletin.html?page_id=282. There have been five issues published up to date. Another means of communicating with the wider public and popularising science is publishing articles or interviews in the Hospodárske noviny newspaper or the Quark journal.
- participation in organising and co-organising research and development- and technology transfer-popularising events, exhibitions displayed on the premises of the SCSTI SR, media programmes on the TA3 TV channel, Regina and VIVA radio stations. The audio or video records of these events are also publicly available on the NPTT website.

Activities proceeding within the individual specific objectives, and within Specific objectives 1 and 3 in particular, are coordinated and often complement each other.

7. NITT SK project as a source of grey literature

Like many other projects, the NITT SK project is a rich source of various categories of grey literature. They have been published as direct outputs of the implementation of the NITT SK project; some categories of grey literature, however, are indirect products of services provided within the NITT SK project. In addition to those already referred to in sub-chapters 4.2.2 and 4.2.3, some further examples of the grey literature categories produced in the course of implementation of the NITT SK project are listed below:

- analytical studies, state-of-the art reviews, questionnaires and reports compiled in the analytical phase of the project which were part of the activities for all three specific objectives of the project. These studies were fundamental for the development of subsequent activities in the NITT SK project [18-27],
- internal methodological materials [28],
- books of abstracts and books of abstracts and presentations from the *Conference NITT SK – Technology Transfer in Slovakia and Abroad (2011– 2013)* [29-32],
- periodical project- monitoring reports,

- decree [33] which replaces a previous guideline on registration of publication activity at universities in Slovakia,
- reports from business trips within the project,
- agreements and contracts concluded with public research institutions and those drafted as part of EPS services,
- promotional materials for conferences, seminars, workshops and exhibitions,
- reports related to industrial property searches as part of EPS services,
- industrial property documentation – applications or patents, utility models, etc. as results of implementation of the NITT SK projects,
- datasets of Slovak Technical Standards.

Activities in the NITT SK project with all its support services will continue to induce knowledge and technology transfer which will consequently lead to an increased production of grey literature. Some categories of grey literature, in particular patent applications and patents granted resulting from research activities at public research institutions, are used as one of qualitative indicators in the evaluating innovative and technology transfer abilities of these institutions [28].

Conclusions

The support activities and services within the NITT SK project are aimed at provison on a national level. So far, 42 requests for support expert services have been processed from 13 public research institutions located in Bratislava, Žilina and Košice, hence encompassing the whole of Slovakia. Three more research institutions from universities in Zvolen, Nitra and Košice will soon be added to this group with more universities and research institutes showing an interest in joining in to capitalise on the benefits and advantages afforded by the NITT SK project. Taking into account that the specialised support expert services have only been in place from March 2013, these figures are high and expectations for the future are promising. A barrier of earlier scepticism on the part of researchers is rapidly dissolving with positive examples attracting more research institutions. This nationwide approach appears to be optimal. To the best of our knowledge, a similar system for supporting technology transfer on a national level with a central Technology Transfer Office is being introduced and developed only in the Republic of Ireland [34].

The trend towards integrating intellectual resources to boost its global competitiveness, harness its knowledge base, enhance its economic position and tackle the great societal challenges of the 21st century is evident in the recent activities of the European Union. Innovation has been the key issue in the Europe 2020 strategy for growth and jobs and mechanisms to strengthen knowledge transfer offices in public research organisations through trans-national collaboration in particular. The European TTO CIRCLE network was established with the aim of linking the major European public research organisations in order to play a collective role in driving changes to the Technology Transfer landscape in Europe, and to boost innovation in Europe through a set of initiatives, including: fostering the use of their knowledge portfolio; sharing best practices, knowledge and expertise; performing joint activities; establishing informal channels of communication with policymakers; organising training programmes; and developing a common approach towards international standards for the professionalisation of technology transfer [35]. This can be seen as another valuable source for networking, knowledge transfer and support not only in establishing technology transfer centres in Slovakia but also in underlying the importance of the National Technology Transfer Centre in Slovakia as a reliable spokesperson and representative of Slovak technology transfer offices.

Where grey literature is concerned, industrial property documentation will be used even more both intensively and extensively as one of the indicators for evaluating the innovation capabilities of research institutions and countries as proposed in Innovation Union Scoreboard 2013 [36].

Acknowledgement

The authors wish to acknowledge the financial support received for the National Infrastructure for Supporting Technology Transfer in Slovakia – NITT SK, national project, from the European Regional Development Fund within the framework of the Research and Development Operational Programme 2007-2013. The authors also wish to thank Eliška Klimentová, Ľubomír Kucka, Olinka Števková and Marián Trgala for their valued comments.

References

[1] SAZALI, A. W. Defining the Concepts of Technology and Technology Transfer: A Literature Analysis [online]. *International Business Research*. The Canadian Center of Science and Education: January 2012, vol. 5(1), 61-71 [viewed 6 November 2013]. ISSN 1913-9004. Available from: http://www.ccsenet.org/journal/index.php/ibr/article/view/13847/9501

[2] CHOI, H. J. Technology Transfer Issues and a New Technology Transfer Model [online]. *The Journal of Technology Studies*. Epsilon Pi Tau, Inc., 35(1), Fall 2009 [viewed 6 November 2013]. ISSN 1541-9258. Available from: http://scholar.lib.vt.edu/ejournals/JOTS/v35/v35n1/choi.html

[3] ASSOCIATION OF UNIVERSITY TECHNOLOGY MANAGERS. About Technology Transfer [viewed 6 November 2013]. Available from: http://www.autm.net/Tech_Transfer/2979.htm

[4] SCHÖPFEL, J. Towards a prague definition of grey literature. *The Grey Journal* [online]. GreyNet. Spring 2011, vol. 7(1), 5-18 [viewed 5 November 2013]. Available from: http://ezproxy.cvtisr.sk:2278/eds/pdfviewer/pdfviewer?sid=aea83b7d-bf53-4d65-979c-205a7a3fbd1f%40sessionmgr113&vid=2&hid=110

[5] HARACHOVÁ, M. and Ľ. KUCKA. PATLIB Centre: why are we grey?. In: *Abstracts from 15th Conference on Grey Literature* [online]. Amsterdam: GreyNet, in press [viewed 6 November 2013]. Available from: http://www.textrelease.com/images/Abstracts_Session_Two.pdf

[6] NOSKOVIČ, J., SMEJA, M. and O. ŠTEVKOVÁ. *Návrh Národného systému pre podporu transferu technológií* [online]. Bratislava: CVTI SR, 2012 [viewed 6 November 2013]. Available from: http://nitt.cvtisr.sk/buxus/docs/Navrh_na_NSPTT_FINAL.pdf

[7] CENTRUM VEDECKO-TECHNICKÝCH INFORMÁCIÍ SR. Duševné *vlastníctvo a zachovanie mlčanlivosti. Príručka pre výskumníkov* [online]. Bratislava: CVTI SR, 2013 [viewed 6 November 2013]. Available from: http://nptt.cvtisr.sk/buxus/docs/DV_a_zachovanie_mlcanlivosti.pdf

[8] BRIGHTON, D., ĎUROVČÍKOVÁ, H. and K. MÜLLEROVÁ. *Metodický postup na zriadenie a prevádzku centra transferu technológií na vysokých školách, vrátane návrhu štandardného modelu fungovania* [online]. Bratislava: CVTI SR, 2012 [viewed 6 November 2013]. Available from: http://nptt.cvtisr.sk/buxus/docs/Metodicky_postup_na_zriadenie_a_prevadzku_CTT.pdf

[9] CENTRUM VEDECKO-TECHNICKÝCH INFORMÁCIÍ SR. A*kademické materiály a publikovanie. Príručka pre výskumníkov.* Bratislava: CVTI SR, 2013.

[10] CENTRUM VEDECKO-TECHNICKÝCH INFORMÁCIÍ SR. *Zmluvy o výskume. Príručka pre výskumníkov*. Bratislava: CVTI SR, 2013.

[11] ADAMOVÁ, Z., KLINKA, T. and K. MÜLLEROVÁ. *Duševné vlastníctvo a transfer technológií 1* [online]. Bratislava: CVTI SR, 2012 [viewed 6 November 2013]. ISBN 978-80-89354-07-8. Available from: http://nptt.cvtisr.sk/buxus/docs///Dusevne_vlastnictvo_a_transfer_technologii_1.pdf

[12] ADAMOVÁ, Z. et al. *Duševné vlastníctvo a transfer technológií 2* [online]. Bratislava: CVTI SR, 2013 [viewed 6 November 2013]. ISBN 978-80-89354-09-2. Available from: http://nptt.cvtisr.sk/buxus/docs///Dusevne_vlastnictvo_a_transfer_technologii_2.pdf

[13] ADAMOVÁ, Z. and J.KRATĚNOVÁ. *Duševné vlastníctvo a transfer technológií 3* [online]. Bratislava: CVTI SR, 2013 [viewed 6 November 2013]. ISBN 978-80-89354-12-2. Available from: http://nptt.cvtisr.sk/buxus/docs///Dusevne_vlastnictvo_a_transfer_technologii_3.pdf

[14] ADAMOVÁ, Z. et al. *Duševné vlastníctvo a transfer technológií 4*. Bratislava: CVTI SR, 2013.

[15] ZENDULKOVÁ, D. and J. NOGE. The SK CRIS system as a source of unique information about scientific activities and their outcomes. In: *Abstracts from 15th Conference on Grey Literature* [online]. Amsterdam: GreyNet, in press [viewed 6 November 2013]. Available from: http://www.textrelease.com/images/Abstracts_Session_Three.pdf

[16] ŽITŇANSKÁ, M. et al. *"Gold Value Offer" to Science and Research in Slovakia – is presented by the NISPEZ project and related activities.* In: *Abstracts from 15th Conference on Grey Literature* [online]. Amsterdam: GreyNet, in press [viewed 6 November 2013]. Available from: http://www.textrelease.com/images/Abstracts_Poster_Session.pdf

[17] NOGE, J. and J. KRAVJAR. Central Repository of Theses and Dissertations – A Unique Umbrella Solution for a Significant Category of Grey Literature in the Slovak Republic. In: *Abstracts from 15th Conference on Grey Literature* [online]. Amsterdam: GreyNet, in press [viewed 6 November 2013]. Available from: http://www.textrelease.com/images/Abstracts_Poster_Session.pdf

[18] LINCZÉNYI, R. et al. *Štúdia I. Analýza súčasného stavu v oblasti transferu technológií v SR a definovanie koncepčných východísk v oblasti TT pre vytvorenie Národného systému podpory TT na Slovensku* [online]. Bratislava: CVTI SR, 2011 [viewed 5 November 2013]. Available from: http://nitt.cvtisr.sk/buxus/docs/Studia_I_o_TT_NITT_SK.pdf

[19] Sipko, S. et al. *Štúdia II. Základné koncepty, nástroje a prístupy k transferu technológií vo svete - prehľadová štúdia* [online]. Bratislava: CVTI SR, 2011 [viewed 5 November 2013]. Available from: http://nitt.cvtisr.sk/buxus/docs/Prieskum_poziadaviek_AO_na_sluzby_TT.pdf

[20] CENTRUM VEDECKO-TECHNICKÝCH INFORMÁCIÍ SR. *Vyhodnotenie dotazníkového prieskumu pre definovanie požiadaviek akademickej obce na dodávku externých podporných služieb v procese transferu technológií v rámci implementácie národného projektu NITT SK* [online]. Bratislava: CVTI SR, 2011 [viewed 5 November 2013]. Available from: http://nitt.cvtisr.sk/buxus/docs/Prieskum_poziadaviek_AO_na_sluzby_TT.pdf

[21] SHEARMAN, A. et al. *Hodnotenie potenciálu vedeckých pracovísk v oblasti výskumu, vývoja a transferu technológií s využitím Centrálneho registra evidencie publikačnej činnosti* [online]. Bratislava: CVTI SR, 2011 [viewed 5 November 2013]. Available from: http://nitt.cvtisr.sk/buxus/docs/NITTSK_studia_IV_akt_1_1_CREPC.pdf

[22] KUBIŠ, M et al. *Analytická štúdia I - východiská pre vytvorenie Integrovaného systému služieb* [online]. Bratislava: CVTI SR, 2011 [viewed 5 November 2013]. Available from: http://nitt.cvtisr.sk/buxus/docs/Akt_2_1_STUDIA_I_zeditovana_final.pdf

[23] KUBIŠ, M et al. *Analytická štúdia II - požiadavky na dodávku špecifických produktov a outsourcingových služieb* [online]. Bratislava: CVTI SR, 2011 [viewed 6 November 2013]. Available from: http://nitt.cvtisr.sk/buxus/docs/Akt_2.1_STUDIA_II_zeditovana_final.pdf

[24] PUTALOVÁ, A. et al. Štúdia I. *Identifikovanie možných nástrojov propagujúcich aktívne využívanie inštitútov ochrany práv duševného vlastníctva vedeckou komunitou v SR* [online]. Bratislava: CVTI SR, 2011 *[viewed 6 November 2013]*. Available from: http://nitt.cvtisr.sk/buxus/docs/Studia_I_final.pdf

[25] MURÍN, G. et al. Štúdia II. *Možnosti využitia klasických prostriedkov masmediálnej komunikácie pre TT* [online]. Bratislava: CVTI SR, 2011 [viewed 5 November 2013]. Available from: http://nitt.cvtisr.sk/buxus/docs/Studia_II_final.pdf

[26] ORAVCOVÁ, A. et al. *Možnosti využitia najnovších internetových nástrojov na efektívne pôsobenie smerom k vedeckej komunite* [online]. Bratislava: CVTI SR, 2011 [viewed 6 November 2013]. Available from: http://nitt.cvtisr.sk/buxus/docs/Studia_III_final.pdf

[27] TÓTHOVÁ, D. et al. *Identifikácia požiadaviek na integrovaný redakčný systém pre webové sídlo Národný portál pre transfer technológií a pre tvorbu a vydávanie elektronického bulletinu venovaného problematike výskumu, vývoja a transferu technológií* [online]. Bratislava: CVTI SR, 2011 [viewed 6 November 2013]. Available from:http://nitt.cvtisr.sk/buxus/docs/Studia_IV_NITT_3_1_final_20110531.pdf

[28] BILSKÝ, Ľ. et al. *Metodika hodnotenia kvality vedeckých pracovísk v oblasti výskumu, vývoja, inovácií a transferu technológií* [online]. Bratislava: CVTI SR, 2012 [viewed 6 November 2013]. Available from: http://nitt.cvtisr.sk/buxus/docs/METODIKA_-Hodnotenie_kvality_VV_pracovisk_1.pdf

[29] CENTRUM VEDECKO-TECHNICKÝCH INFORMÁCIÍ SR. *Zborník abstraktov z podujatia Konferencia NITT SK 2011 Transfer technológií na Slovensku a v zahraničí. Book of Abstracts from Conference NITT SK 2011 Technology Transfer in Slovakia and Abroad.* Bratislava: CVTI SR, 2013. ISBN 978-80-89354-19-1.

[30] CENTRUM VEDECKO-TECHNICKÝCH INFORMÁCIÍ SR. *Zborník abstraktov z podujatia Konferencia NITT SK 2011 Transfer technológií na Slovensku a v zahraničí. Book of Abstracts from Conference NITT SK 2011 Technology Transfer in Slovakia and Abroad*. Bratislava: CVTI SR, 2013. ISBN 978-80-89354-17-7.

[31] CENTRUM VEDECKO-TECHNICKÝCH INFORMÁCIÍ SR. *Zborník abstraktov a prezentácií z podujatia Konferencia NITT SK 2011 Transfer technológií na Slovensku a v zahraničí. Book of Abstracts and Presentations from Conference NITT SK 2011 Technology Transfer in Slovakia and Abroad*. Bratislava: CVTI SR, 2013. ISBN 978-80-89354-16-0.

[32] CENTRUM VEDECKO-TECHNICKÝCH INFORMÁCIÍ SR. *Zborník abstraktov z podujatia Konferencia NITT SK 2013 Transfer technológií na Slovensku a v zahraničí. Book of Abstracts from Conference NITT SK 2013 Technology Transfer in Slovakia and Abroad.* Bratislava: CVTI SR, 2013. ISBN 978-80-89354-18-4.

[33] *Vyhláška MŠVVaŠ SR č.456/2012 o centrálnom registri evidencie publikačnej činnosti a centrálnom registri evidencie umeleckej činnosti.*

[34] GOVERNMENT OF THE IRISH REPUBLIC. *Putting public research to work for Ireland Policies and procedures to help industry make good use of Ireland's public research institutions* [online]. Enterprise Ireland, 2013 [viewed 11 November 2013]. Available from: http://www.enterprise-ireland.com/EI_Corporate/en/Research-Innovation/Companies/IPP-Putting-public-research-to-work-for-Ireland.pdf

[35] THE EUROPEAN COMMISSION. Enterprise and Industry. *Innovation Union Scoreboard* [online]. 20 September 2013 [viewed 11 November 2013]. Available from: http://ec.europa.eu/enterprise/policies/innovation/facts-figures-analysis/innovation-scoreboard/index_en.htm

[36] THE EUROPEAN UNION. *Innovation Union Scoreboard 13* [online]. The European Union, 2013 [viewed 11 November 2013]. ISBN 978-92-79-27583-8. Available from: http://ec.europa.eu/enterprise/policies/innovation/files/ius-2013_en.pdf

A Social Networking Research Environment for Scientific Data Sharing: The D4Science Offering

Massimiliano Assante, Leonardo Candela, Donatella Castelli,
Francesco Mangiacrapa, Pasquale Pagano
Istituto di Scienza e Tecnologie dell'Informazione - Italian National Research Council, Pisa, Italy

Abstract
Modern science calls for innovative practices to facilitate research collaborations that span institutions, disciplines, and countries. Scientists and practitioners are called to produce enhanced forms of scientific communication thus to make it possible for others to identify errors, to support, reject or refine theories and to reuse data for further understanding and knowledge. This scenario can benefit a lot from Internet-based services aiming at providing individual users having at their disposal scarce resources with interaction-oriented facilities, i.e., social computing, and elastic access to facilities on demand, i.e., cloud computing. In this paper, we present the D4Science approach that by leveraging social computing and cloud computing realises an integrated web-based working environment where scientists have at their fingertips what is needed to accomplish a scientific investigation. In particular, we present the large array of collaboration-oriented facilities obtained by integrating social networking paradigms with virtual research environments.

Introduction
In the last decade, the data intensive science [1] paradigm acquired popularity thus giving raise to a growing demand for effectively dealing with "*big data*" [2]. Moreover, the emphasis on *science democratisation* is gaining momentum. Scientists are called to produce enhanced forms of scientific communication based on publication of "*comprehensive scientific theories*" – including the data and algorithms they are based on – as to make it possible for "*others to identify errors, to support, reject or refine theories and to reuse data for further understanding and knowledge*" [3] (we name these forms of communication *comprehensive research products*). These trends are occurring not only in the context of the "big" sciences (e.g. physics, astronomy, earth observation) but can also be found in the *long tail* of science [4], i.e. the large amount of relatively small laboratories and individual researchers who do not have access to large-scale dedicated IT yet have the potentiality to produce the bulk of scientific knowledge.
This calls for innovative IT research environments potentially providing every scientist with the instruments he/she needs to conduct a complete scientific investigation by benefitting from state-of-the-art practices across the boundaries of institutions, disciplines, and geographic regions.
Recent technological advances in *social computing* [5] and *cloud computing* [6] can largely contribute to the realisation of IT research environments favoring these new trends. As a matter of fact they enable internet-based services to provide individual users having scarce resources at their disposal with *(a)* interaction-oriented facilities (social computing), and *(b)* elastic access to facilities on demand (cloud computing). For social computing, social networks like LinkedIn or Twitter are revolutionising the way people use and exploit the Internet to communicate, and crowdsourcing is perceived as an innovative problem-solving strategy [7]. For cloud computing, be it infrastructure, platform or software as a service [8], people use it almost every day even without being aware of it, e.g. Google's gmail, Apple's iCloud, Dropbox.
In this paper we present the social networking research facilities offered by the D4Science infrastructure [9]. D4Science is an IT infrastructure that provides its users with a comprehensive set of data management facilities offered *as-a-Service*. The D4Science social networking research facilities complement the infrastructure offering by providing a scalable set of services promoting the cooperation among users, e.g. allowing to share news on recently produced comprehensive research products, start discussions on such objects, rate and reuse such objects for the production of new ones. Such facilities are implemented by benefitting from the underlying infrastructure to realise an elastic behaviour, i.e. the services realising them exploit the infrastructure to acquire computing capacities on demand.
The main contribution of this paper is to describe the set of social networking research facilities offered by the D4Science infrastructure while documenting how these facilities have been actually implemented by relying on a cloud-oriented offering of resources.

The D4Science Social Networking Facilities

The D4Science infrastructure [9] is an *Hybrid Data Infrastructure* (HDI) [10] built and operated by relying on gCube [11,12], a software system specifically conceived to simplify the realisation and management of such a type of infrastructures. An HDI (and its enabling technology) is characterised by three distinguishing features:

- it is a "system of systems", i.e., it federates existing systems (including Grid and Cloud) with the goal to supplement (while not supplanting) the services offered by these constituents by nicely integrating them into an organised and managed whole;
- it offers a comprehensive yet extensible set of data management facilities as-a-Service operating on a rich array of data typologies ranging from papers to tabular data, maps, and comprehensive research products;
- it enables *Virtual Research Environments* (VREs) [13,14], i.e. web-based, community-oriented, comprehensive, flexible and secure working environments conceived to serve the needs of specific application domains.

These three distinguishing features build one on top of the other to promote an economy-of-scale oriented model, i.e. VREs are realised by relying on the set of facilities offered as-a-Service (e.g. it is possible to reuse a given service across multiple VREs) while the mechanisms for effectively and efficiently offering facilities as-a-Service relies on the federated systems offering (e.g. it is possible to automatically deploy a new service to outsource a computational intensive service task by relying on a federated Cloud infrastructure).

In the remainder of the paper, we will focus on the social networking facilities only. However, a very brief description of some key services is included to provide the reader with an overview of the whole D4Science. D4Science provides services for *(i)* management of the whole infrastructure including services deployment, hosting nodes management, systems federation, and VREs creation and dismissal; *(ii)* data access and storage including CRUD[1]-oriented, efficient and large scale management of files, semi-structured data, tabular data, geospatial data, and biodiversity data; *(iii)* large-scale data processing including a rich array of facilities ranging from environments supporting the distributed execution of user-defined workflows by transparently relying on available computing platforms [15] to environments offering off-the-shelf data-mining algorithms, e.g. [16].

The collaboration, cooperation, and knowledge transfer in VREs are crucial and so is the sharing of scientific data and comprehensive research products. Some previous studies on knowledge management show that the success of knowledge transfer does not lie only in communication systems or documents, but also in social relationships [17]. This is the reason that brought us to design and experiment a *social networking area* to complement the VRE offering in terms of available facilities. This functional area aims at promoting knowledge transfer by providing users with a "clear overview" of what is happening around them, in their research environments.

The D4Science social networking facilities manifest in a number of applications made available through a thin client (namely a web browser) and relying on the HDI. These facilities are conceptually close to the common facilities promoted by social networks – e.g. posting news, commenting on posted news – yet adapted to deal with the settings previously described, namely to promote large scale collaboration and cooperation on comprehensive scientific products, data sets, theories and tools.

Three main entities characterise the social networking facilities: *data*, i.e. any information object managed via a VRE tool, *application*, i.e. any tool made available via a VRE, and *user*, i.e. any member of a VRE. The resulting environment connects: *(i)* (scientific) data with applications and users so that it is possible to share data and make them *easily accessible* for other users of the environment; *(ii)* applications (of a VRE) with their users so that it is possible to be informed on what is being produced by users across VRE applications; *(iii)* every user with other user activities so that it is possible to look at what is being shared or posted by others and see if any of these posts can be of interest.

Overall, the D4Science social networking facilities (cf. Figure 1) provide their users with: *(a)* a continuously updated list of events / news produced by users and applications (*Home Social*), *(b)* a folder-based file system allowing for managing complex information objects in a seamless way (*Workspace*), *(c)* an email-like facility for exchanging messages with selected co-workers (*Messages*), *(d)* a list of happenings organised by date, e.g. publication of a comprehensive research product, comments on a research product (*Notifications*), *(e)* a settings area where the user can configure diverse aspects characterising the system behaviour including his/her data and notification preferences (*Personalisation*).

[1] Create, Read, Update and Delete (CRUD) are the basic functions a storage should offer.

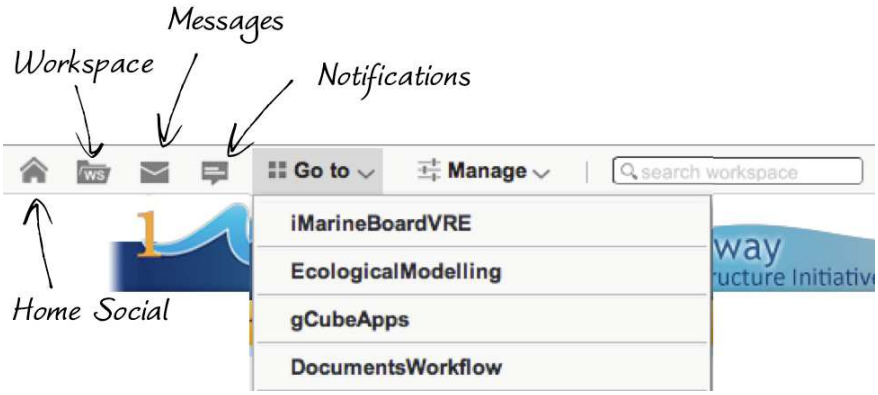


Figure 1. Dockbar

Home Social

The Home Social consists of two facilities.

The *News Feed* lists users' and applications' updates and makes these available to every user according to his/her preferences. Moreover, it enables users to comment, subscribe or re-share these updates, and – in case of application posts – to make users access the subject of the post directly in the application that created it, e.g. to see a data set or an application.

The *Share Updates* enables (a) users to post updates or interesting links to others (this could be done with any social network) and (b) applications to post updates such as the availability of a new product or facility.

For instance, Figure 2 shows a news about a just produced AquaMaps object [16,18], i.e. a scientific product consisting of a species distribution model and a number of maps resulting from its projection. Moreover, there is a news about a recent algorithm made available by an application (the Statistical Manager).

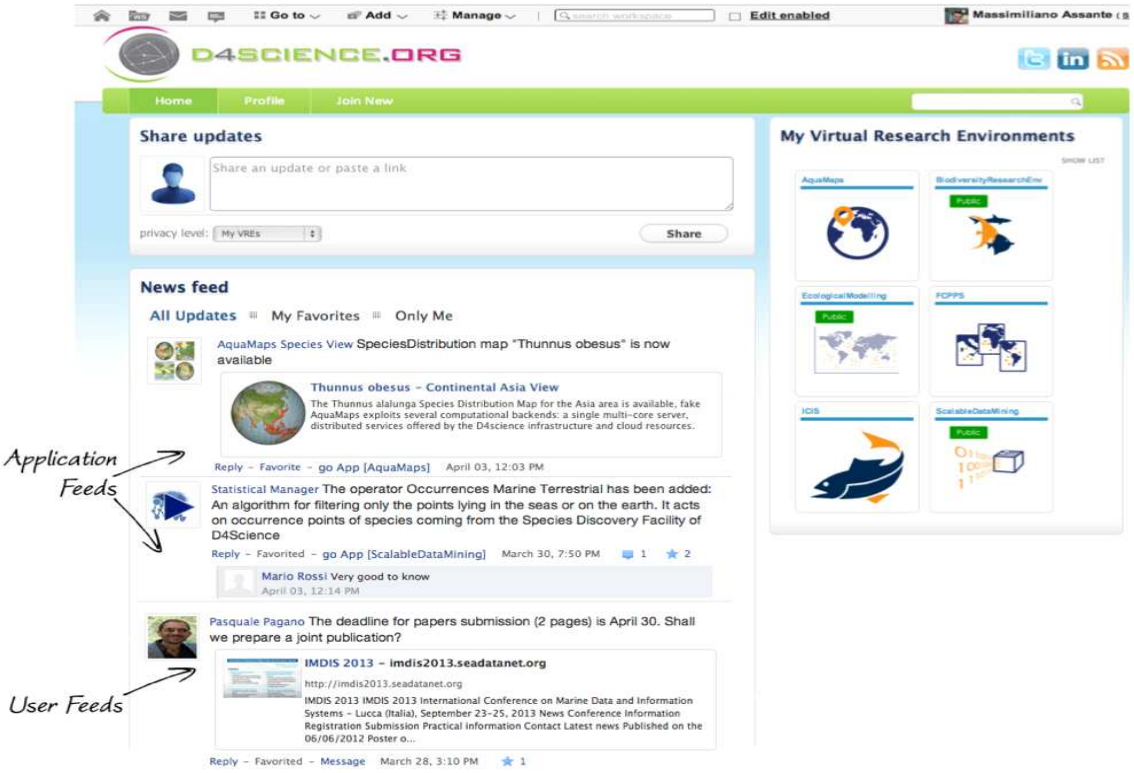


Figure 2. Home Social Facilities

Workspace

The Workspace resembles a folder-based file system any user is familiar with, where the added value is represented by the type of information objects it can manage in a seamless way. It supports items ranging from binary files to information objects representing, for instance, tabular data, workflows, species distribution maps, time series, and comprehensive research products. Through it, sharing of data is fostered to immediate availability of results, workflows, annotations, documents and the like.

Sharing is fundamental since users need to work collaboratively on the same data and rely on common research materials. Sharing can be performed per folder and it is invite-based. Any shared folder and, in

turn, its content including sub folders is shared with the other users of the system. The users involved in the sharing are alerted by the notification mechanism previously introduced, which will further be explained in the following.

Besides making its users capable of organising and sharing very different items, the workspace provides them with other useful features: *(i)* the Search feature allows users to search for their files by name and within the files' textual content (full-text search), *(ii)* the Smart Folders feature allows users to create virtual folders that do not physically exist but are a result of a previously saved search. Of course the content of a Smart Folder changes in time depending on the current content of the user workspace. This is because whenever a user opens a smart folder the related saved search is performed again, live, against the existing content of the workspace; *(iii)* the Accounting feature instead allows every users partaking into a shared folder to see the operations performed within that particular shared folder they belong to. These operations are the ones available over the content of a folder, such as upload of a new file, editing, deletion, read etc. Any user of a shared folder can see what the other users of the shared folder have done.

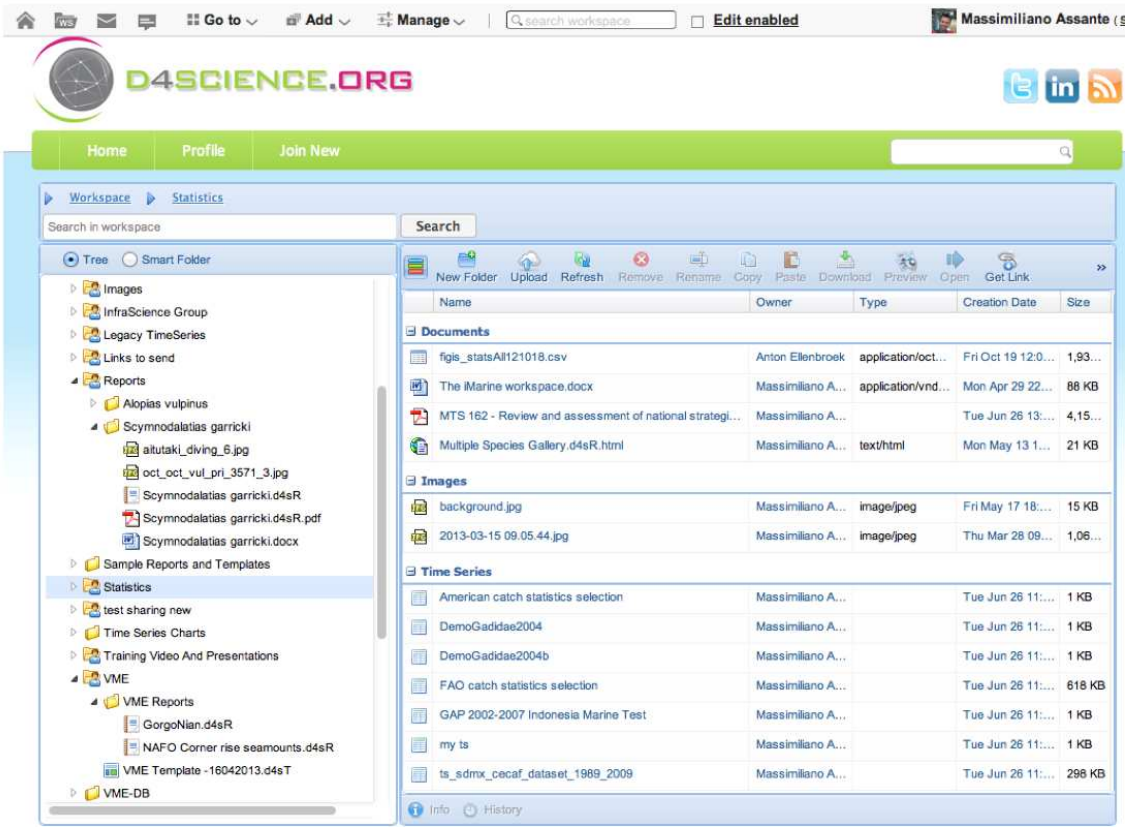


Figure 3. Workspace

Messages

Messages is an email-like application for exchanging private messages with the other users of the system. In this case the real added value is represented by the integration with the other social networking facilities. In particular, since it supports workspace item attachments, users can attach not only a simple binary file to a message, rather they could attach complex multi-parted information objects. For instance, the action of sending a very large dataset in this case would be immediate and would not require bandwidth consumption at all. The user could attach this very large dataset, maybe having size of tens of gigabytes, to his private message by selecting it from his workspace and send it to another user. Behind the scenes a local copy of this file would be performed and only its pointer passed to the receiving user.

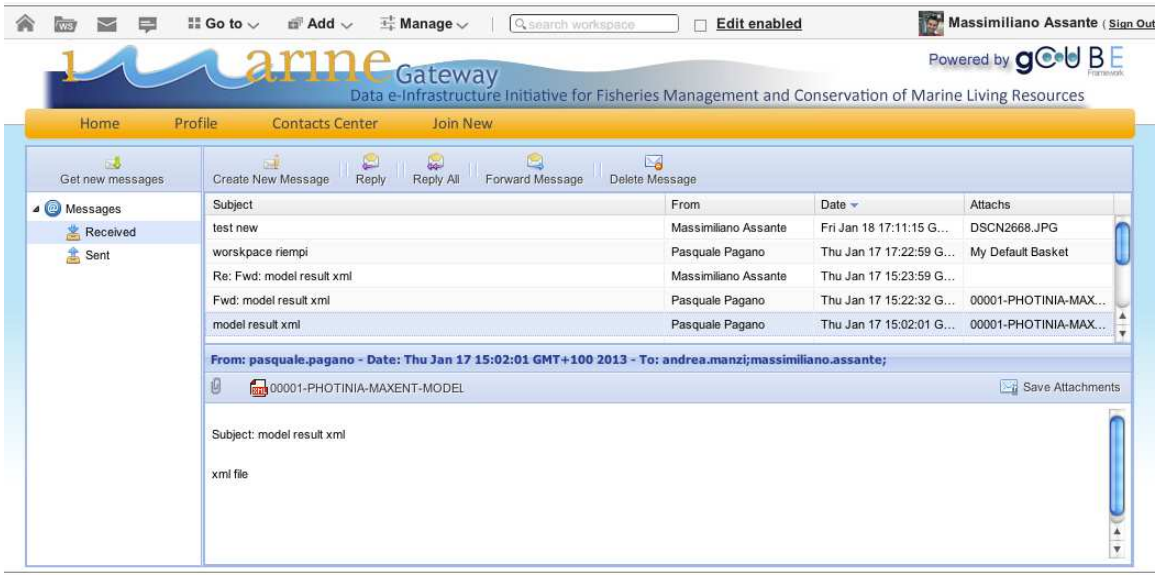


Figure 4. Messages

Notifications

The notifications page contains a list of happenings organized by date in reverse chronological order. Indeed, every user of the system receives notifications from VRE applications and other users he/she is interested in. In Figure 5 we can see an example of the Notifications Page. For each single notification the user is provided with a link to the subject of the notification. For instance if a notification is about an updated item in one of his workspace shared folder he could click right next to this notification and be taken directly on the workspace shared folder involved in the change. The same applies for user posts, messages, calendar events etc.

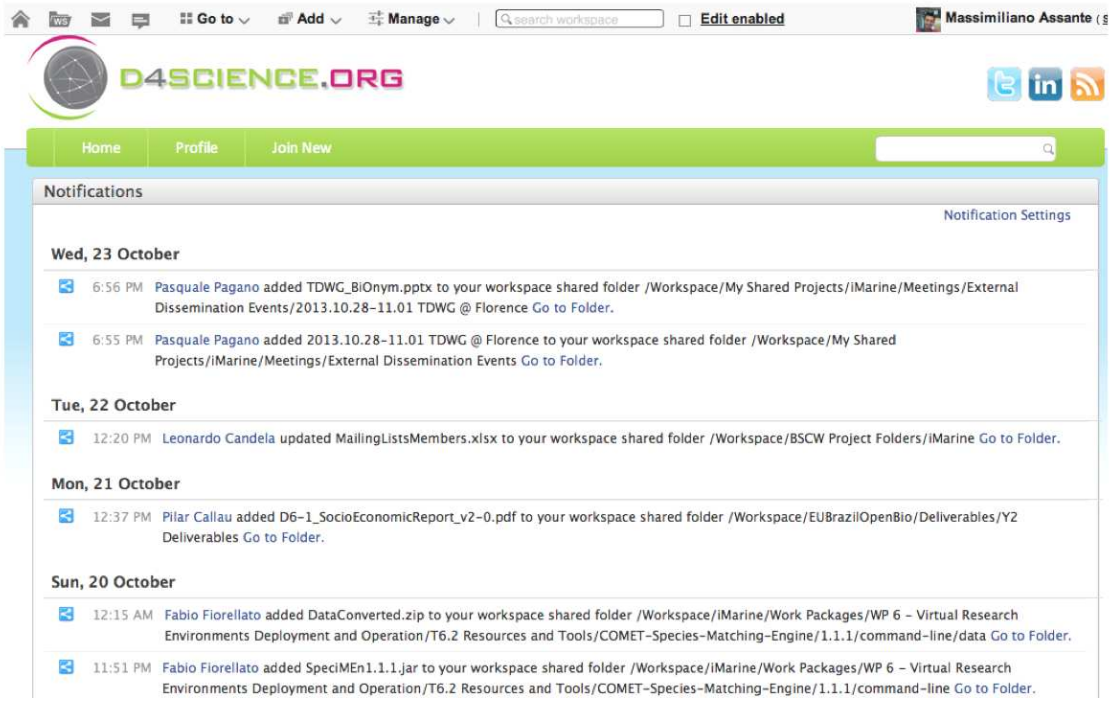


Figure 5. Notifications

Personalisation

Personalization is the settings area where users can configure diverse aspects characterizing the system behavior regarding notification preferences. The way a user gets notifications is very fine grained. For each type of notification users might receive, it is possible to specify how to receive it and through which "channel" to be notified. Figure 6 shows a partial list of the Personalization user interface. The notifications are classified per category so that the ones belonging to the workspace sharing are grouped together as well as the ones belonging to the Social Networking etc.

For each notification type it is present *(i)* a number of checkboxes allowing users to choose the notification channels, *(ii)* an ON/OFF button allowing enabling/disabling the notification, and (iii) a textual explanation of the notification type.

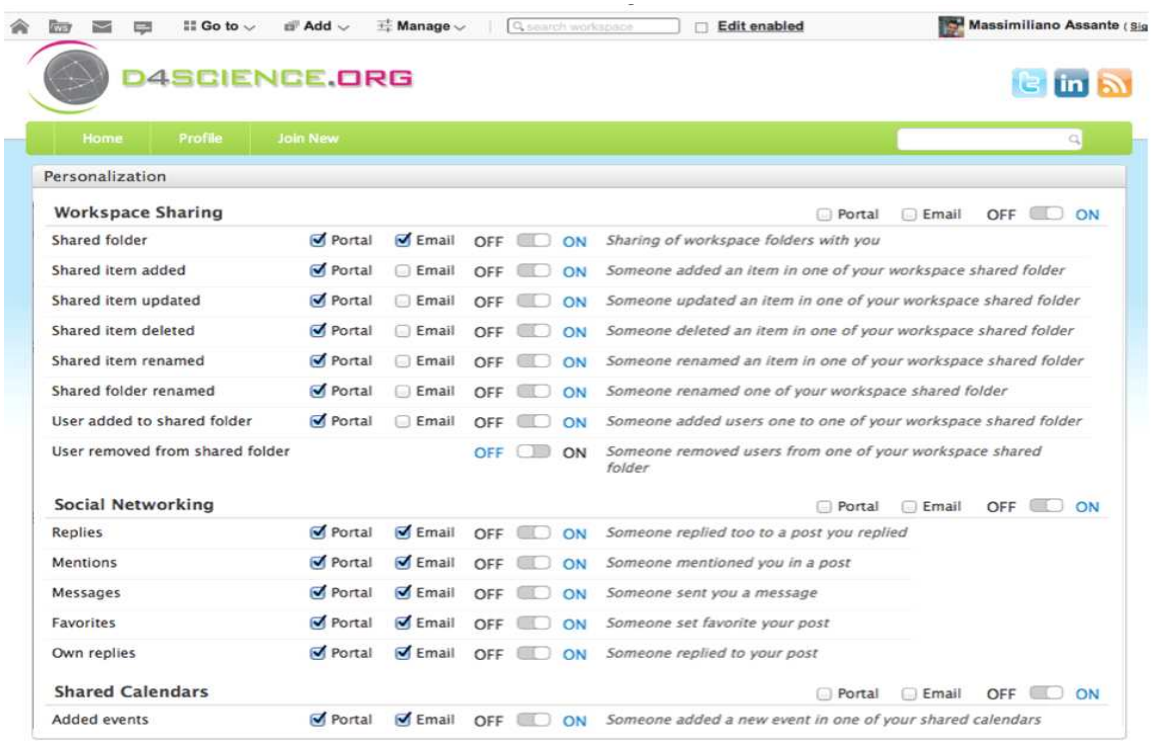


Figure 6. Personalisations

Realising The Social Networking Research Environment In Practice

As for the user interface, the environment was built over the Web portal technology. Indeed, it is composed by Web based interfaces that are world-wide accessible. It exploits a component-oriented approach (Java Portlet), where each functionality described in the previous section has been mapped onto a "piece" of web UI as displayed in Figure 1. This figure shows the dockbar portlet, a "control panel" that is always present on the top of the page when a user is working on the research environment. The dockbar aims at (i) making VRE members able to access the social facilities introduced in the previous section, (ii) giving one-click access to the VRE(s) a user subscribed to, and (iii) providing users with the capability to search in the past news as well as in the workspace. Figure 2 shows the *Home Social* page, a core page of the research environment that is displayed when a user logs in: this portal page contains the 2 portlets delegated to the production (Share Updates) and consumption (News Feed) of user and application updates plus the "My Virtual Research Environments", a portlet delegated to the listing of the VREs a user subscribed to.

As for the back-end, the social networking facilities exploit the Social Networking Framework, hereafter named gCube SNF. This framework is composed by a number of Java libraries designed to exploit the resources accessible via the D4Science infrastructure. By embracing the HDI approach, gCube SNF is capable to exploit several different storage back-end technologies, including the most common NoSQL data stores. In particular, a Column store and a Document store are exploited to manage the communication, collaboration and cooperation facilities, and the storage and sharing of heterogeneous information objects, respectively.

The Column store is instantiated in the D4Science infrastructure via an Apache Cassandra cluster. It is used by the gCube SNF for handling the facilities partaking in the Home Social, including the Notifications system. Scalability, high-availability and high throughput offered by this Column store are the capabilities required to support the raising needs of the word-wide distributed scientific communities currently exploiting D4Science. gCube SNF includes a Java library, named Social Networking Library, that abstracting over the datastore provides all the required logic to support publication of posts, post replies and/or likes, as well as the "smart retrieval" of these posts per user or VRE. This library also supports selective notifications on the activities performed by users over these Posts.

For the Document store, D4Science offers a wide range of opportunities by means of *JackRabbit*, i.e. a hierarchical content store with support for structured and unstructured content, full text search,

versioning, and transactions, *MongoDB*, i.e. a document-oriented store giving support for replication and high availability, automatic sharding, and Map/Reduce, and *Terrastore*, i.e. a distributed document store with support for elastic exploitation of resources, per-document consistency, collection-based interface, custom data partitioning. gCube SNF exploits most of the available technologies and their intrinsic capabilities to deliver a distributed virtual file system with support for the heterogeneous information objects managed by the D4Science applications and users. It provides organization of the information objects in (virtual) directories; manages the typical metadata of a file system, e.g. last modified time, creation time, owner, access permissions, and other file attributes such as mime-type, read-only, etc; supports for accounting of space and usage, e.g. number of read and write operations. In essence, by combining different technologies it offers support for replication, high availability, high accessibility via WebDav and HTTP-based access, and controlled and secure sharing. By doing this, it embeds all typical Dropbox features in a distributed infrastructure and extends them to include the support for virtual complex scientific products and workflows storing, sharing, and accounting.
All the Java libraries composing the gCube SNF belong to the gCube software system, released under the terms of the European Union Public License (EUPL v1.1). The source code, the binary and the documentation are all freely available. Further details are available at the gCube system public wiki [19].

Conclusion

The D4Science social networking facilities represent an innovative approach that nicely complements the VRE offering and boost cooperation and collaboration. These facilities are conceptually borrowed from social networks yet adapted to deal with the scientific data practices. For these reasons, their characteristic of relying on HDI facilities is fundamental, e.g. the capability to effectively share ``big data'' or scientific products is a must that is hard to satisfy without having scalable and elastic access to tools and computing capabilities. These facilities are flexible enough to be exploited in contexts ranging from cultural heritage [20] to biodiversity [18] communities of practice.
Among the envisaged enhancements of such facilities, it is worth to cite the introduction of recommendations, e.g. to notify a user on a potentially interesting VRE, application or dataset by relying on users posts, interests and workspace content [21].

Acknowledgements The work reported has been partially supported by the iMarine project (FP7 of the European Commission, FP7-INFRASTRUCTURES-2011-2, Contract No. 283644) and the EUBrazilOpenBio project (FP7 of the European Commission, FP7-ICT-2011.EU-Brazil, Contract No. 288754).

References

1. T. Hey, S. Tansley, and K. Tolle, Eds., *The Fourth Paradigm: Data-Intensive Scientific Discovery*, Microsoft Research, 2009.
2. L. K. Stapleton, “Taming Big Data,” *IBM Data Management Magazine*, vol. 16, no. 2, pp. 12–18, 2011.
3. G. Boulton, P. Campbell, B. Collins, P. Elias, D. W. Hall, G. Laurie, O. O’Neill, M. Rawlins, D. J. Thornton, P. Vallance, and M. Walport, “Science as an open enterprise,” The Royal Society, Final report, June 2012.
4. P. Murray-Rust. (2008, January) Big Science and Long-tail Science. Peter Murray-Rust Blog. [Online]. Available: http://blogs.ch.cam.ac.uk/pmr/2008/01/29/big-science-and-long-tail-science/
5. F.-Y. Wang, K. Carley, D. Zeng, and W. Mao, “Social computing: From social informatics to social intelligence,” *Intelligent Systems, IEEE*, vol. 22, no. 2, pp. 79–83, 2007.
6. M. Armbrust, A. Fox, R. Griffith, A. D. Joseph, R. Katz, A. Konwinski, G. Lee, D. Patterson, A. Rabkin, I. Stoica, and M. Zaharia, “A view of cloud computing,” *Communications of the ACM*, vol. 53, no. 4, pp. 50–58, Apr. 2010.
7. A. Doan, R. Ramakrishnan, and A. Y. Halevy, “Crowdsourcing systems on the World-Wide Web,” *Commun. ACM*, vol. 54, no. 4, pp. 86–96, Apr. 2011.
8. Foster, Y. Zhao, I. Raicu, and S. Lu, “Cloud Computing and Grid Computing 360-Degree Compared,” in *Grid Computing Environments Workshop, 2008. GCE ’08*, 2008.
9. D4Science.org. (2013) D4science infrastructure. [Online]. Available: http://www.d4science.org
10. L. Candela, D. Castelli, and P. Pagano, “Managing big data through hybrid data infrastructures,” *ERCIM News*, no. 89, pp. 37–38, 2012.
11. L. Candela, D. Castelli, and P. Pagano, “gCube: A Service-Oriented Application Framework on the Grid,” *ERCIM News*, no. 72, pp. 48–49, January 2008.
12. L. Candela, D. Castelli, and P. Pagano, “Making Virtual Research Environments in the Cloud a Reality: the gCube Approach,” *ERCIM News*, no. 83, pp. 32–33, October 2010.
13. L. Candela, D. Castelli, and P. Pagano, “Virtual research environments: an overview and a research agenda,” *CODATA Data Science Journal*, vol. 12, pp. GRDI75-GRDI81, 2013.
14. M. Assante, P. Pagano, L. Candela, F. De Faveri, and L. Lelii, “An approach to virtual research environment user interfaces dynamic construction,” in *Proceedings of the International Conference on Theory and Practice of Digital Libraries (TPDL 2011)*, 2011.
15. M. M. Tsangaris, G. Kakaletris, H. Kllapi, G. Papanikos, F. Pentaris, P. Polydoras, E. Sitaridi, V. Stoumpos, and Y. E. Ioannidis, “Dataflow processing and optimization on grid and cloud infrastructures,” *IEEE Data Eng. Bull.*, vol. 32, no. 1, pp. 67–74, 2009.

16. L. Candela, D. Castelli, G. Coro, P. Pagano, and F. Sinibaldi, “Species distribution modeling in the cloud,” *Concurrency and Computation: Practice and Experience*, vol. -, pp. –, 2013.
17. S. J. Yang and I. Y. Chen, “A social network-based system for supporting interactive collaboration in knowledge sharing over peer-to-peer network,” *International Journal of Human-Computer Studies*, vol. 66, no. 1, pp. 36–50, 2008. [Online]. Available: http://www.sciencedirect.com/science/ article/pii/S1071581907001139
18. L. Candela and P. Pagano, “The D4Science Approach toward Grid Resource Sharing: The Species Occurrence Maps Generation Case,” in *Data Driven e-Science - Use Cases and Successful Applications of Distributed Computing Infrastructures (ISGC 2010)*, S. C. Lin and E. Yen, Eds. Springer, 2011, pp. 225–238. [Online]. Available: http://www.springer.com/978-1-4419-8013-7
19. gCube Team. (2013) gcube wiki. [Online]. Available: https://gcube.wiki.gcube-system.org/gcube
20. T. Blanke, L. Candela, M. Hedges, M. Priddy, and F. Simeoni, “Deploying general-purpose virtual research environments for humanities research,” *Philosophical Transactions of the Royal Society A*, vol. 368, pp. 3813–3828, 2010.
21. H. Avancini, L. Candela, and U. Straccia, “Recommenders in a Personalized, Collaborative Digital Library Environment,” *Journal of Intelligent Information Systems*, vol. 28, no. 3, pp. 253– 283, June 2007.

A Study on the Improvement of Legal System for Collecting and Registering National R&D Reports

Kiseok Choi
Korea Institute of Science & Technology Information, KISTI, Korea

Abstract
A R&D report, which is the result of national research and development project of each government ministry, has a characteristic as a public knowledge-based asset. So it is used with various systems for disclosing and expanding its outcome to the public. In fact, however, systematically collecting and utilizing its related information in KISTI, Korea Institute of Science & Technology Information(National R&D Reports Registration and Administration Agency), is insufficient. Formal regulation for submitting and registering national R&D reports to KISTI, which is wholly responsible for circulating national R&D information, has been arranged well so far. However, specific and concrete regulation system is not enough to make sure of their actual submission and registration. Therefore, this study examines and proposes a specific improvement plan for making good progress of collecting and registering national R&D reports at country levels. Also, it examines improvement methods for standardizing national R&D report style and applying for it, in order to increase work efficiency in submitting and registering the information of national R&D reports.

Introduction

Nowadays, a R&D report, which is the result of national research and development project of each government ministry, has a characteristic as a public knowledge-based asset. So it is used with various systems for disclosing and expanding its outcome to the public. In fact, however, systematically collecting and utilizing its related information in KISTI, Korea Institute of Science & Technology Information(National R&D Reports Registration and Administration Agency), is insufficient. National R&D reports, which have been collected and built up as a database in KISTI, are 7,322 in 2007 and 6,581 in 2008. And DB building rate of the whole national R&D reports is about 14%. Especially, in case of closed R&D reports, it is limited not to disclose its outcome in public. But even if its several matters, such as its closed period, selection process of target subject, the reason of its closing, etc. are removed, the process of public disclose and its expansion is insufficient. According to it, although the closing rate of national R&D reports is over 60% in 2007 and 2008, the actual output, such as collecting, disclosing, or utilizing its relevant information to the national science and technology information agency etc., is very inadequate. Also, after national R&D reports which are collected to KISTI are data-based, the using process of them has a possibility of copyright infringement. Because existing legislation doesn't have specific rules for regulation. That is, administering institutes have the copyright of national R&D reports, but matters of collecting and utilizing relevant information are settled by each research agreement. Therefore, when KISTI collects, manages, or serves relevant information, copyright problems would arise without obvious agreement from an administering institute. On a background like this, this study analyzes what the problems of existing relevant legal system are. And it also inspects and proposes its improvement plan.

Content and scope of the study

Improvement of Legal System for Collecting and Registering National R&D Reports

Formal regulation for submitting and registering national R&D reports to KISTI, which is wholly responsible for circulating national R&D information, has been arranged well so far. However, specific and concrete regulation system is not enough to make sure of their actual submission and registration. Therefore, this study examines and proposes a specific improvement plan for making good progress of collecting and registering national R&D reports at country levels. Also, it examines improvement methods for standardizing national R&D report style and applying for it, in order to increase work efficiency in submitting and registering the information of national R&D reports.

Improvement of Legal System for National R&D Reports Information Closure

According to the existing legislations, national R&D reports could not to be disclosed or registered, if they need external security or the interested parties require not to disclose to the public within the certain period of time. Therefore, they have some effect on collection and circulation of R&D reports information. This study examines the present condition of closure system of national R&D reports, and presents a system improvement plan for collecting and utilizing national R&D reports at national level.

Solutions to the Copyright Infringement of Circulating National R&D Reports Information

National R&D reports are not only important knowledge assets, which should be collected, managed, preserved and circulated systematically at national level, but also researchers' intellectual works, which should be protected by copyright. Therefore, when KISTI builds up an information system for collecting, managing and circulating relevant information, copyright infringement problems, such as illegal copying or transmitting, could arise. But the existing legislations don't have specific regulations. And the interpretation of them is uncertain. With the possibility of copyright infringement from now on, this study examines and researches solutions to the copyright infringement problems, which are involved in collecting, managing and circulating information of national R&D reports.

Result of the Study

Improvement Plan for Regulations for National R&D Reports

This study shows that formally, the existing legal system, related to the comprehensive collection, management, and circulation of national R&D reports information, is generally appropriate in organizing and utilizing them. However, as to specific strategies of a relevant system or improving system efficiency, the following system improvement should be in need by each section. The following improvement plan is needed, in operating the circulation and registration system of national R&D reports. That is, in order to collect and submit national R&D reports to KISTI systematically, it is desirable to maintain the existing R&D reports circulation system, apart from the research product registration system. In order to make research reports registration system practical and to grasp its exact state periodically, it is necessary that periodic reporting procedure of national registration of R&D reports to government department in charge should be settled on a relevant law. In order to manage and circulate national R&D reports information effectively, it is desirable to make rules, which are involved in the standard form of research reports and every ministry and agency have to use obligatory. The Improvement Plan for Maximizing Disclosure and Utilization of Closed National R&D Reports is as follows. It is essential to maintain the existing security and closure system of national R&D reports, to prevent national research information from foreign stealing. But the following improvement plan is required for the access to the relevant information smoothly and using it. It is desirable to set the closure procedure and methods of national R&D reports in detail. And it is also desirable to state clearly the time limit of closure by each cause. It is advisable to set clearly disclosure procedure and methods of relevant information after the time limit, in order to promote relevant information circulation. And it is also advisable to prepare the periodical reporting procedure of operation status of security and closure system, in order to make its effectiveness high.

Possible Solutions for Copyright on the R&D Reports Information circulation

Generally National research report has characteristics as public goods. In other words, national R&D achievements are technologies developed and supported by national budget or public funds. So, it should be managed to meet public welfare because its characteristic is basically public goods. Also, the purpose of national research and development is to contribute to the development of the national economy through research and industrial use of national R&D achievements. Therefore, in case when we cannot take advantage smoothly of the research and development achievements, government should intervene in that situation. The national research and development achievements are developed for the national public purposes, so it can be used to guarantee the extent of public usage. On the other hand, in the case of national research and development achievements strategically important to homeland security, it is necessary to establish institutional arrangements to prevent leakage or abuse to be disclosure of them.

In the case of national research and development projects, I think it is the problem that the existing provisions of a copyright laws applying to national research report as it is in a point that national research and development projects have Public interest because its funding is Public R & D budget. In other words, who would be owned the study's copyright from taxpayers finance is need to be treated differently from general case. It May be desirable to treat R&D reports as mission work under certain conditions for the smooth distribution and utilization of Report of the National Research. However, the mission work in order to be recognized by the business will be required to meet the requirements of the mission work is. However, looking at the process of creating the actual state of research reports there is a limit cannot be denied to treat research reports as mission works in the strict sense. Therefore, it is thought to be desirable that the writings of the national research reports as a special type of work are to be treated as a writings governed by copyright law and other relevant laws, as well as by the co-management regulations.

Consequently, in order not to infringe copyright, such as national R&D reports processing, DB construction, on-line information service, etc., the following improvement and supplementation plans are in need. When a kind of national institute, such as a national intelligence agency and etc., collects national R&D reports, constructs DB, or services them for the public purposes, they are regarded as taking an approval, such as a right to reproduce, circulate, and transmit, from the responsible parties. That is, a kind of legitimate consent system should be prepared in the relevant legislation. And when making a national R&D agreement, relevant matters, such as collection and utilization of relevant information, should be set clearly.

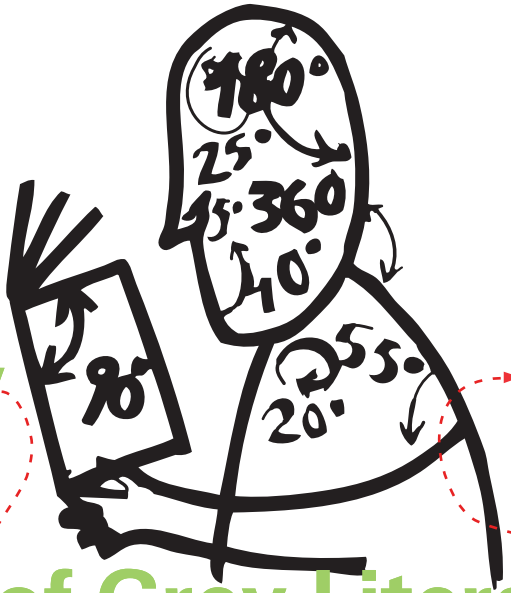
Application Schemes

It could be used as a basic material of the improvement of the existing legal system, so as to build up an effective information collection and systematic information provision system. Some of the results are already reflected in the process of system improvement, through an amendment of the rules of national R&D project on August 11, 2010. It is necessary that others also should be reflected when later amending of relevant legislation. It could be used as practical guidelines on building an information system for collecting and managing national R&D reports and information circulating service by KISTI. And it also could be used as useful materials when other researchers and policy makers fulfill their duties.

NTK

50°6'14.083"N, 14°23'26.365"E
Národní technická knihovna
National Technical Library

NUŠL



National Repository in the Czech Republic of Grey Literature

Features

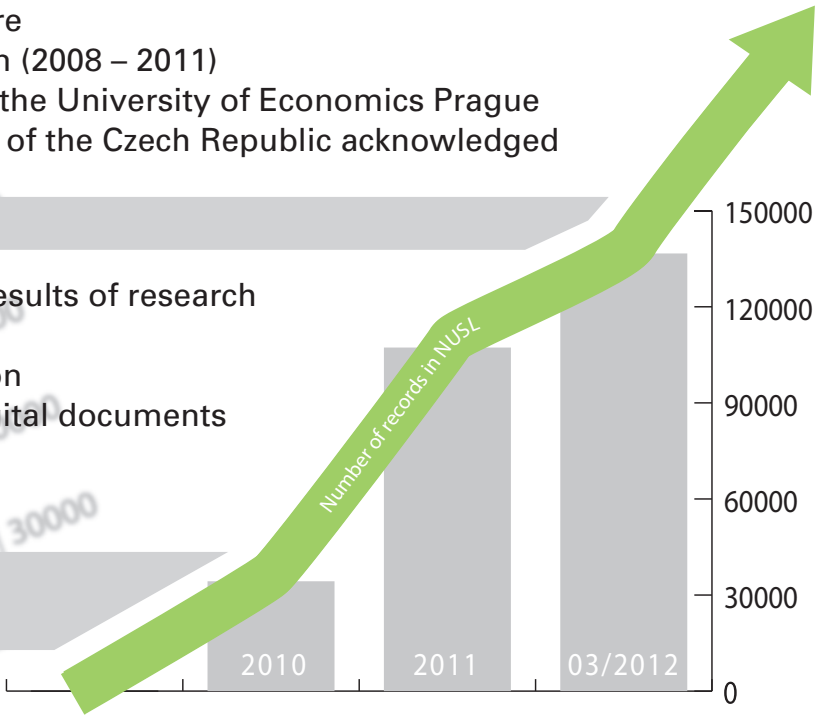
Website: www.nusl.cz
Provider: National Technical Library
Records: over 200 000 records
Partners: over 90 organizations
Source area: Academy of Science, Public Research Institutions, Universities, Libraries etc.
International Cooperation: OpenGrey, DRIVER, ROAR, OpenDOAR
Collection provenance: Czech Republic

Based on

Project: The Digital Library for Grey Literature
– Functional Model and Pilot Implementation (2008 – 2011)
Participants: the National Technical Library, the University of Economics Prague
Financial support: by the Ministry of Culture of the Czech Republic acknowledged

Goals

- Central access to grey literature and the results of research and development in the CR
- Support of science, research and education
- Systematic collection of metadata and digital documents
- Long-term archiving and preservation
- Cooperation with foreign repository



Support of expert discussion about Grey Literature

Annual Workshops:
http://nrgl.techlib.cz/index.php/Workshop
Informative Web pages: http://nrgl.techlib.cz
Publication: Grey Literature Repositories
http://nrgl.techlib.cz/index.php/Book

www.nusl.cz

Survey of enhanced publications in the Czech Republic

Petra Pejšová and Hana Vyčítalová
National Library of Technology, NTK, Czech Republic

Abstract
The poster presents survey about the state of enhanced publications in the Czech Republic. The goal of this survey was to find out situation of enhanced publications at research institutions in the Czech Republic. What research data arise? How are they stored and archived? To whom they made available? Are there the real enhanced publications, i.e. research publications linked directly to research data? 113 Czech research institutions were contacted with online questionnaire during 11th September to 7th October 2013. The poster will show the results from 65 Czech research institutions which filled in the questionnaire.

Keywords: Enhanced publications, research data, research publications, research institutions

According to the definition from the DRIVER-II project, enhanced publications are publications that are enriched with three categories of information – research data, extra materials and post-publication data. (DRIVER, 2013) Research data include for instance measuring records, experiment results etc. Extra materials means models, algorithms, images, metadata, etc. Post-publication data include commentaries, ranking and other types of information that are produced only subsequently.
This constitutes a next level of scholarly publishing, the objective of which is no longer just a publication as a text, but a work which should at the same time contain such features as underlying data of models, algorithms, etc. Adding these basic details from research to publications facilitates verification, reproduction and reuse of research results. The strength of enhanced publications rests in description of the relationships between underlying data and research output in a meaningful way, both in a readable and machine-processable form. (SURF, 2013) Reference possibilities enable researchers to link any type of object to another object. It is thus easy to link an article to lectures, a review or an interview which may be found on the Internet at various places. This way, the relationships between individual types of information can be described in a structured manner and at a single location. This is what we call enhanced publications.
The state of enhanced publications has not been mapped in the Czech Republic yet. Therefore, the National Library of Technology[1] has carried out a survey this year. The aim of the survey was to find out what research data are produced and archived by research institutions. Furthermore, we were also interested in the manner of data storing and archiving and, most importantly – whether data are linked with research publications and we can therefore talk about enhanced publications in the Czech Republic.
What was our hypothesis?

1. Research data exist in the Czech research institutions.
2. Research data are stored on workstations of individual research workers or on shared disk.
3. Data are archived, but the process is not specifically defined.
4. Data are available for other research workers, but not online.
5. Free licenses like Creative Commons are used very rarely.

The survey was targeted on public research institutions, private research institutions and other institutions engaged in research. We addressed the representatives of these institutions that were expected to have a comprehensive knowledge of the research activities of the institution concerned; we chose persons holding the posts of science officers, science secretaries or deputy directors for research or science. In cases when it was not possible to determine who holds such post in the institution or no such posts have been set up by the institution, representatives from the institution's management were selected.
The survey was performed through online questionnaire, consisting of nineteen questions. This was a structured questionnaire; most of the questions were closed with one or more answers possible. Some questions followed in the previous ones, depending on the answer chosen for the preceding question. One of online applications on the web was used to disseminate the questionnaire.
Selected respondents received an informative email with a request to fill in the questionnaire. The email contained information on the notion of enhanced publications, the aim of the survey as well as the use of the data obtained. The letter also contained a request asking the addressee to pass the information and link to the questionnaire on to some of their colleagues in case they cannot complete the survey themselves.

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

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

Table 1 Number of Responses

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

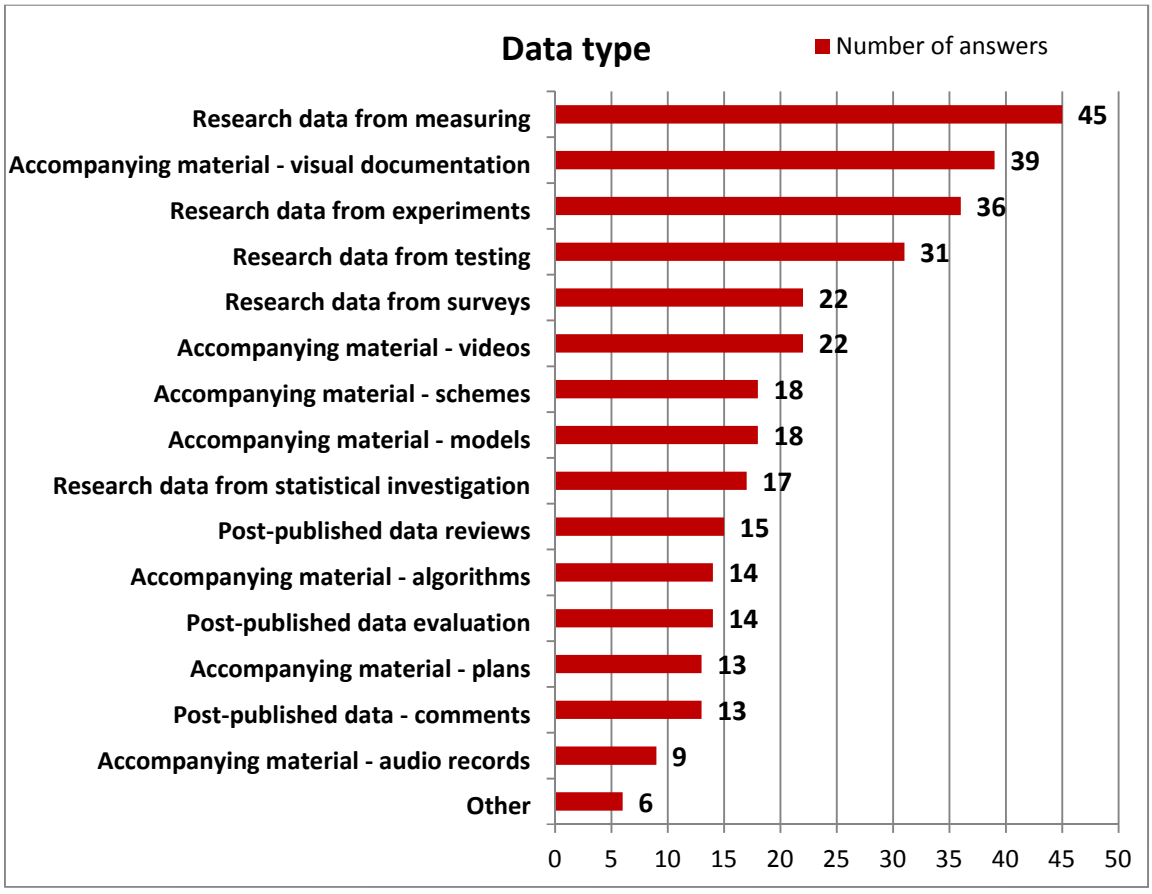


Figure 1 Types of Research Data

Data stored and archived in PDF, DOC and XLS formats were most frequent. Other data formats were not included so frequently. XML, CSV or image formats (JPEG, TIFF etc.) belonged to the more common formats too. Less common were SAV, CIF formats and various formats used in cartography.
We were also interested in the total volume of data. We asked the respondents at least for a rough estimate; the answer was not obligatory. Nonetheless, it seems that this question was difficult to answer as only a small number of respondents provided an answer to it. The total volume of research data is probably also related to the field in which the institution is engaged. The usual answer was in the order of tens or hundreds of gigabytes; several answers reached the order of terabytes.
Another significant area to be investigated was the question where and how research data are stored and archived. The respondents were again offered several options and could choose one or more of them or add their own one. The answers confirmed our hypothesis that most often data are not stored centrally in a shared repository, but that they remain on workstations (computers) of individual researchers. This answer was selected 52x. Other frequent manners of storage were a shared directory on a disc (29 answers) and central data repository of the institution – this answer was selected by 18 respondents. Some of the respondents indicated e.g. two or three of the provided options, which means that the conception of data storage is clearly not entirely resolved and unified in many of the institutions. However, most of the institutions are not going to make any changes for the time being, which was the focus of the following questionnaire question.

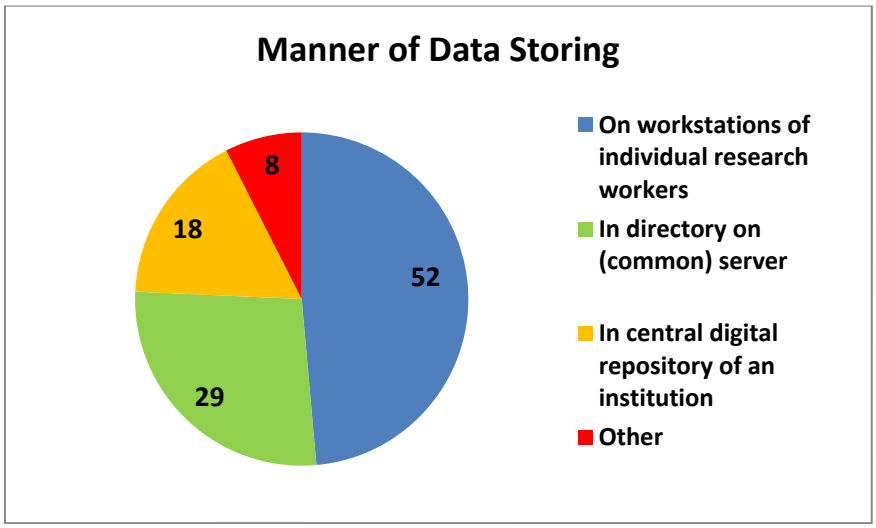


Figure 2 Manner of Data Storing

Also the question as to whether the institutions archive their data for a period of more than ten years rendered interesting findings. 54% of respondents answered yes, 31 % did not know, but 15% (10 respondents) answered that they do not archive data for such a long time, which is quite a large number. Yet, the majority of the respondents at the same time indicated that they reuse data in further research.
Data are taken care of mostly by researchers themselves; this was the answer of 53 respondents. But respondents could choose more than one option for this question as well. Taking care of the data is further often the responsibility of authorized persons in the library, archive, IT department or of science secretaries or officers or managers.
The core of the survey consisted in the question as to whether we can actually talk about enhanced publications in the Czech Republic, i.e. whether research publications are linked with research data. The questionnaire outlined several answers as to how such a linkage can function in practice. Apart from the ideal method of linkage – joint storage of the publication in electronic form together with the relevant research data in a digital repository – several other alternative options were suggested. It was a relatively big surprise that 58% of respondents stated that they link research publications and data in one way or another. 5 representatives of institutions that do not link publications and data at the moment indicated that they would like to change the situation.

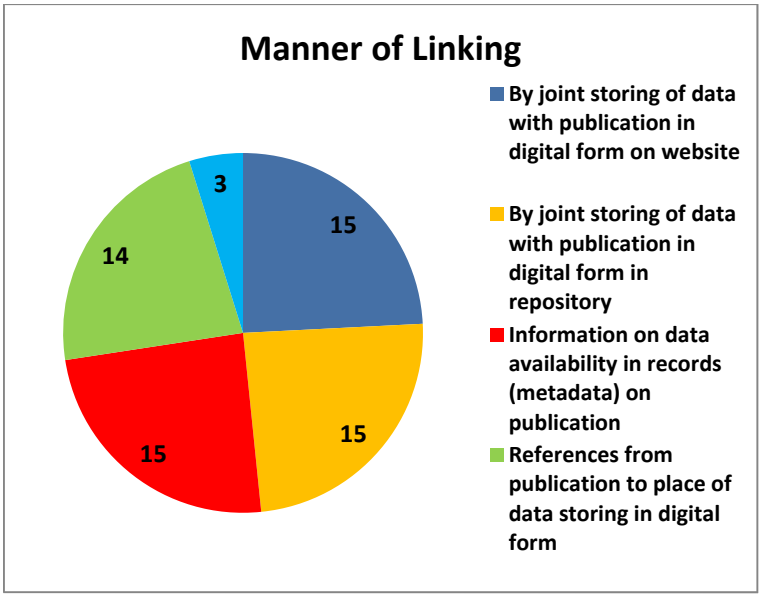


Figure 3 Manner of Linking

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

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

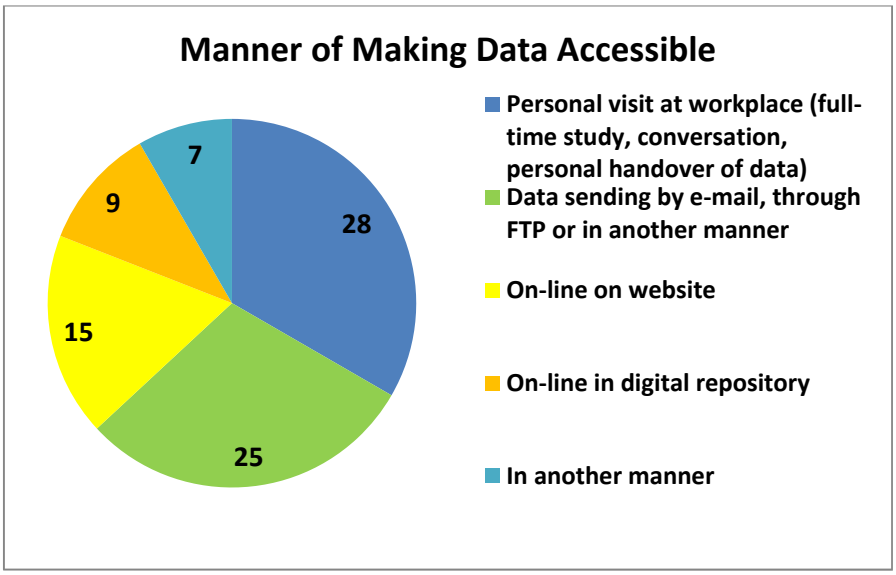


Figure 4 Manner of Making Data Accessible

The last area on which the survey focused was familiarity with and use of public licenses (specifically Creative Commons[2]). Under public licenses the author may provide their work or data to the public under specific conditions. A user may use or distribute the work under the same conditions as determined by the author by means of the selected type of license. Only six percent (4 respondents) stated that they use licenses to designate and thus also make research data available. Two other respondents confirmed that they would like to start to use public licenses.

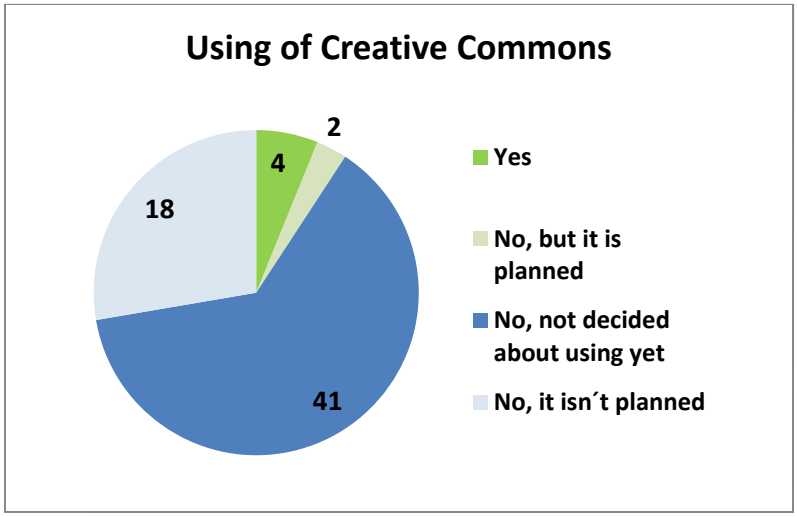


Figure 5 Using of Creative Commons

Conclusion

It follows from the information obtained that a more widespread existence of enhanced publications in the Czech context is not excluded. At the moment, actual enhanced publications, i.e. research publications linked with research data available ideally online, are at the very beginning. The manner of storing and archiving of research data at Czech research institutions would need improvement. It would be advisable to have a clear concept with a central repository for the entire institution and also greater technical support to researchers.

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

References

Enhanced Publications. In: *DRIVER – Digital Repository Infrastructure Vision for European Research* [online]. 2013, 18 February 2009 [cit. 2014-01-02]. Available from: http://www.driver-repository.eu/Enhanced-Publications.html

Enhanced publications. In: *SURF: Collaborative organisation for ICT in Dutch higher education and research* [online]. 2013, 26-11-2013 [cit. 2014-01-02]. Available from: http://www.surf.nl/en/themes/research/research-data-management/enhanced-publications/index.html

[1] http://www.techlib.cz/en/
[2] http://creativecommons.org/

Industrial Philology: Problems and techniques of data and archives preservation for future generations

Eva Sassolini, Manuela Sassi, Sebastiana Cucurullo, Alessandra Cinini, Stefano Sbrulli
Istituto di Linguistica Computazionale, "Antonio Zampolli" CNR, Italy

Management policies of texts collections in electronic format

The main objective of digital archiving of texts is their re-use and preservation. The concept that guides these initiatives is linked to structural and organizational needs which heavily influence the definition of the *format specifications* that describe the organisation of the archives at various levels and consists of a more or less complex document. A format specification provides the details needed to build a file from a text, establishes the admitted encodings and software applications that can decode the file and make its content accessible.

These structural specifications can have an extremely variable size and they depend on the complexity of the format. Although some format specifications are, for the most part, independent of the specific software (for example, ASCII and Unicode[1] codes), many of them are related to the historical period in which the texts were acquired and also by dated software technologies. The file format specification should evolve hand in hand with the related software, and the fate of one is in fact often linked to that of the other. It is therefore appropriate to face the issue of obsolescence of software together with the obsolescence of file formats and of storage medium.

Obsolescence: file formats and software

A file format may become obsolete for several reasons:

- the latest versions of software do not support the previous files;
- the format itself is superseded by a new one, or becomes more complex;
- the format is not so widely adopted, or the scientific community does not support the creation of compatible software;
- the format is not anymore compatible with the current computers;
- decline of the software that supports the format.

Digital formats are a challenge for texts conservation. In early decades of computing, few were aware of the threat posed by the obsolescence of file formats, for long-term digital preservation. A systematic effort to collect software documentation or all specifications necessary for the conservation of textual files was completely missing. Without proper documentation, the task of interpreting the contents of an old file becomes very tiring. It is only recently that we began to catalogue them, document, and understand their relationships and variations.

While most of the software is updated regularly, files are often not updated to meet the new format requirements, which may make unreadable to the latest version of the software, and if older versions are no longer available – of do not run on the recent computer or in the current version of the operating system – the data is "lost". Furthermore, because of the complexity and nature of some file formats, it can be extremely complex to know if a converted file in another format has retained all its features.

Conservation measures

The public specifications produced by international standards bodies are extremely safe from the point of view of their long-term accessibility. Usually the standards need to respond to a large community of users, not linked to individual economic interests. The large number of people who participated in the drafting of these standards also promotes a general recognition at the time of use.

Compatibility with available standards however is usually not a priority for data producers, because either it is costly, or there are commercial pressures to make older formats become quickly obsolete.

On the other hand, standard formats are not necessarily the best choice for all situations and in any case they need to be carefully applied to specific applications and domains. Certainly, the offer great advantages for long-term preservation and archiving. To reduce the risk of obsolescence, a standard must be shared and flexible, while minimizing its possible variations in order to minimize software adaptation. It is a standardization process which should primarily concern the formats most at risk, such as those created by obsolete or outdated software versions.

[1] Unicode Consortium: www.unicode.org/

The ILC textual heritage: history

In the early 60s of the last century, almost everywhere in the world computer-based technologies in the field of philology, lexicology, lexicography and the humanities in general were developed. Italy and Pisa in particular, were at the forefront in this field, first with the creation of a language section in the first and largest electronic computing Italian Centre, CNUCE[2], and then within the National Research Council , with the establishment of the Institute of Computational Linguistics (ILC). It has always acted as a reference point for the national and international scientific community for the study and implementation of procedures for automatic text analysis and lexical data. Since the first attempts to use the computer to analyze linguistic material, ILC has gathered an aggregate wealth of knowledge, tools and materials thanks also to the collaboration of scholars from various disciplines (linguists, lexicographers, philologists, historians, philosophers, jurists etc..). A large background of skills, standard processing and coding, processing procedures, and a big archive of textual material now constitutes the ILC wealth, which nevertheless runs the risk of being lost because of obsolescence.

The first texts in electronic format

Electronic processing has always been articulated in the three basic steps: input (input or acquisition of data within the procedure), processing (analysis of the material and specific processing depending on the intended purpose) and output (output, on suitable media, of the results of the previous stages). The output of any processing phase might be considered as a final result in its own right, even if in a specific project it is just an intermediate analysis subject to subsequent processing phases.

A fundamental parameter for the whole process is the type of storage medium used to store the material at the different processing stages. In the 70s, the choice of storage media was dictated primarily by financial reasons: at ILC magnetic tapes were extensively used as they allowed to store large amount of data at a relatively low cost. Their only drawback consisted in the forced sequential access to data, which meant that for reading (or writing) a piece of data recorded on that medium it was necessary to read (or rewrite) in sequence all the preceding data. In fact, this technology entailed objective limitations in the possibility to implement particularly sophisticated and efficient algorithms for data access.

Technological evolution

The sudden and widespread development of information technology, both hardware and software; the creation of increasingly more powerful machines at lower costs; the creation of more user-friendly work environments (Operating Systems) with high performance; all this resulted in a widespread distribution of computing resources. The landscape in which computing resources were concentrated in a few geographically-located computing centres, had drastically changed thanks to such development and technological evolution. In the new landscape of "distributed computing", where disconnected individual PCs and workstations substituted the mainframe-based architecture, data and software production continued with many different approaches influenced by the various environments in which they developed. We have thus lost the ability to use a single procedure for texts acquisition; shared initiatives, which at the time seemed promising, have proliferated but sometimes have proven to be unsuitable or insufficient to represent all the information contained in the text. The institutions, that previously ran the whole procedure of text processing, from the setting of the project until its complete execution, changed their role and how ILC seek to contribute to the definition of standard processing and coding to create textual materials in standardized formats and make them available to the community.

Recovery and standardization of textual materials

There are various levels of recovery, but they all have in common the requirement of the final destination: a standardized and universally recognized format, that allows the exchange and storage.

The first and main level is obviously the magnetic readability media and this already raises strong concerns: in order to safely preserve the tapes they were stored in absolute protection from heat, moisture, and other harmful agents. However, the hardware units needed for the reading such supports were not maintained, which makes such tapes no longer readable within the Computing Centre. The reasons for this situation are to be found, as we have said, in the fragmentation introduced by distributed computing and economic issues.

To better understand the data we have, we must understand the reasons which have produced them. For many years the "input" phases have been possible only through the preparation of punched cards that already now seem to belong to the pre-history of computer science; later came units of data entry,

[2] The University of Pisa founded C.N.U.C.E. (National Centre for University Computing) in 1965 as the Computing Centre of the University and quickly established a partnership with the IBM Scientific Centre of Pisa. In 1974 C.N.U.C.E. became a Institute of CNR (National Research Council).

able to record on magnetic media or to operate in direct connection with the electronic computer.
The "output" phases normally consisted in the creation of two types of results:

- storing the results both as a rescue of these same results, and as an intermediate input to further processing;
- printing the results on fanfold paper to get the final output of the process, drafts to be used for checking the correctness of the work performed, or to have a working medium that can be enriched with new data or used to classify the data already stored.

The main texts problems

In dealing with the management of a text, the main difficulty concerns the management of sets of special characters such as diacritics, or the management of languages with non-Latin alphabets, or texts that have multiple levels of annotation (e.g. comments, notes in the margin, various types of footnotes, structured text or dramatic text, etc.).
The development of new encoding standards has pushed us to rethink the models of representation of the data. In most cases the approach used in the past was to adopt the "ANSI format", with all issues related to the sharing of sets of positions between the tables of the ISO 8859 family . In the process of revision of the entire methodology of conservation of the text, an assessment was made of different formats and of related conversion costs. The analyzes made have brought to the adoption of the Unicode encoding, which assigns a unique code to every character, regardless of the software platform and also of the language. Unicode is the official implementation of the international standard ISO / IEC 10646 and allows us to make your application platform-independent, and adequate for different countries and languages.

First Step

Over the past 50 years, technology has evolved in a compulsive and ungoverned way, this attitude is observed in conservation strategies that, in the same way, are not organized (parallel example is global warming, where only the energy lobby governs, in perpetual conflict of interest).
We define our strategy "protocol in hindsight". In effect, this long period of crisis has led to react precariously, pursuing more by urgency than by programming. The scarcity of human and economic resources, the questionable competence of the latest ruling class, are the causes that have allowed almost to reach a point of no return in the recovery of the materials. It is urgent to act now, before the retirements and fear for the future causes the disappearance of skills to re-interpret codes and methods made obsolete by technological development.
The first prototype was dictated by the need to recover the texts in Greek for a research project. In that case we set the table of equivalence between the set of character transliterations into the Latin alphabet (originally translated to EBCDIC to ASCII set and Unicode) and polyphonic Greek, taking into account the multi-encodings necessary, in the seventies, to express the presence of up to three accents in a single character. Having dealt with and positively resolved the more difficult test, we are now dedicated to the study of a protocol for the remaining texts, which do not have these problems, but are however in need of expertise to interpret the historical memory testimonials remained, including tabulated and technical descriptions (track-record, coding, procedures and intermediate stages of processing, etc.).

Text acquisition strategy

If we retrace the steps of texts acquisition which saw the ILC among the pioneers in the industry, it is easy to understand that there is no single conversion mapping but that it is necessary to make an assessment that considers the different types of material and their specific recovery paths.
Actually, it has been possible to make only an estimate around the textual heritage. However, this is sufficient to set up a common procedure and useful to evaluate the costs of the whole operation. Depending on the type of material (ranging from texts on magnetic tapes to machine readable and editable digital texts) we have prepared different phases of recovery.
The acquisition strategy adopted is linked with the software that has always been used for the treatment of ILC texts: DBT [3,4]. In fact this tool is still available and usable, we tried to maintain compatibility with the various stages up to where it was possible.
The single tasks or phases are therefore intermediate text formats, supported by the DBT software, needed to get to the final mapping onto a single TEI-XML file. For example, to convert a text file with obsolete character encodings requires at least three phases: then a first mapping involves the conversion to an intermediate format, typically an ANSI encoding, a second format is produced by the recognition, management and remapping of all annotations inserted in the text; and finally, the last phase involves the construction of a parser that can read these annotations and convert them into appropriate TEI-XML tags.

Table 1 below summarizes the information that are typically associated to of the main categories of recoverable texts.

Source text	Perc.	Transition phases (TP) required	Meta data
Text in magnetic tapes	10%	Many type TP	study and research in the archives ILC
Text divided into separate resources	5%	TP>3	recovered from paper-based data
Text in obsolete file	10%	TP>2	recovered from paper-based data
Digital text with obsolete character encoding	10%	2<TP<3	recovered from: - paper-based data - the digital format
Digital text	65%	One TP	recovered from the digital format

Table 1:text acquisition strategy

A more complex case is represented by the lemmatized texts, where the annotations are at the level of word and then more extensive. Even for the annotation of lemmatized texts there was a massive use of the DBT software. In this specific case, a different version of DBT were developed in order to be adapted to as the text lemmatization as it evolved over time. Often improvements were dictated by the particular requirements of the texts; sometimes they were due to technological adaptation. In the acquisition protocol for this type of text, this level of analysis is added to the others together with the evaluation of the type of software tool that was used at the time.

Source text	Transition phases (TP) required	specific annotations type encoding	Meta data
Text in magnetic tape	Many type TP	?	work long and difficult
Text divided into separate resources	TP>3	DBT type encoding	recovered from paper-based data
Text in obsolete file	TP>2	Obsolete type encoding	recovered from paper-based data
Text digital with obsolete character encoding	2<TP<3	Specific type encoding	recovered from: - paper-based data - the digital format
Digital text	One TP	ILC text encoding	recovered from the digital format

Table 2: annotated text acquisition strategy

Conclusion

The technological process is the sum of all the forces that contribute to it, both private and public, and cannot be controlled in all its aspects and consequences. The preservation of that data produced with outdated technologies should be taken handled especially by the public institution, as this is part of the historical heritage; therefore we have to do our part to make this happen so that these resources can be reused. This will be possible only through a joint effort of the institutions involved, at the regional, national and international level. ILC is currently establishing co-operation agreements, such as the one with the “Accademia della Crusca”, and other similar arrangements, with the goal of gathering data

resources for maintenance, preservation and re-use by third parties.
Finally, we take this opportunity to call for collaboration with groups that use or possess are still working magnetic tape units such as those shown in the figure below.

Table 3: magnetic tape

References

1. A. Cinini, S. Cucurullo, P. Picchi, M. Sassi, S. Sbrulli, E. Sassolini (2013). I testi antichi: un patrimonio culturale da conservare e riutilizzare. In: 27a DIDAMATICA 2013, Tecnologie e Metodi per la Didattica del Futuro, Pisa. (Pisa, 7-8-9 may 2013). Proceedings, AICA, Pisa, pp. 867 - 870.
2. E. Picchi, M. L. Ceccotti, S. Cucurullo, M. Sassi, E. Sassolini (2004). Linguistic Miner. An Italian Linguistic Knowledge System. In: LREC Fourth International Conference on Language Resources and Evaluation (Lisboa-Portugal, 26-27-28 may 2004). Proceedings, M.T. Lino, M.F. Xavier, F. Ferreira, R. Costa, R. Silvia (eds.), pp. 1811 – 1814.
3. E. Picchi (2003). Pisystem: sistemi integrati per l'analisi testuale, in Linguistica Computazionale, Vol. XVIII-IX, I.L.C. and Computational Linguistics, special issue, A. Zampolli, N. Calzolari, L. Cignoni, (Eds.), I.E.P.I., Pisa-Roma, (2003) pp. 597-627.
4. E. Picchi (2003). Esperienze nel settore dell'analisi di corpora testuali: software e strumenti linguistici, Informatica e Scienze Umane, in LEO (Lessico Intellettuale Europeo), a cura di Marco Veneziani, S. Olschki editore Maggio 2003, pp.129-155

Data Management & Curation Services: Exploring Stakeholders Opinions

Plato L. Smith II
Florida State University, FSU, United States

Abstract
The purpose of the research study was to explore stakeholders' opinions on select data management and curation services issues that currently affect all disciplines. A data management and curation services 10-questions survey questionnaire was developed and administered to select data management and curation promoters (funders), stakeholders (institutions), and users (evaluators) from November 5, 2012 to December 5, 2012. The survey was approved by the Florida State University Institutional Review Board (IRB) and assigned the HSC No. 2012.9198 on November 2, 2012. The survey was started by 64 participants, completed by 53, and garnered an 83% completion rate. The survey's findings from the data management and curation key concepts category lead to the development of the data management and curation (DMC) framework (see Figure 1).

Background & Significance

The development of this pilot study stemmed from the data management and curation issues discovered in literature and in practice. Some of the data management and curation issues that contributed to the development and the purpose for this pilot study include:

- Definitional confusion of data management and curation key concepts
- Competing data management and curation models/frameworks
- Underdeveloped theory of digital preservation and theory of digital curation
- Under-utilization of relevant standards, best practices, and guidelines where appropriate
- A need to articulate data management and curation across disciplines

Within the scope of this research study data management planning is defined as the planning of policies for the management of data types, formats, metadata, standards, integrity, privacy, protection, confidentiality, security, intellectual property rights, dissemination, reuse/re-distribution, derivatives, archives, preservation, and access (NSF, 2011). The management of data includes analog [physical], digitized [made electronic] & born digital [no physical surrogate] data. NSF's data management plan requirements have incentivized the development of a multitude of programs, projects, and initiatives aimed at promoting and providing data management planning knowledge, skills, and abilities for NSF data management plan requirements compliancy. Without the specification, clarification, & definition of key concepts; assessment of current data management practices, experiences, & methods; interrelationships of key concepts; and utilization of multiple methodological approaches, data management will be problematic, fragmented, and ineffective. The accomplishment of effective data management is contingent on funders, stakeholders, and users' investment and support in *Infrastructure, Cultural Change, Economic Sustainability, Data Management Guidelines, and Ethics and Internet Protocol* (Blatecky, 2012, p. 5) across organizations, institutions, & domains.

For the purposes of this pilot study, data management and curation (DMC) is defined as a research data management concept that includes: (1) data management planning, (2) data curation, (3) digital curation, and (4) digital preservation key concepts. Together theses independent yet interrelated key concepts comprise the data management and curation (DMC) framework (see Figure 1). These key concepts are focused on the lifecycle management of data. The DMC practices include four major data lifecycle management processes that:

1. Fulfill departmental, institutional, organizational policies & data management requirements; (data management planning)
2. Provide data creation (primary, secondary, tertiary data), data publication, minimal data description; (data curation)
3. Facilitate added value (metadata), management & storage of archived data over data lifecycle; (digital curation)
4. Integrate a series of technical & strategic actions and consultations to ensure continual data authenticity (digital preservation) (Lord & Macdonald, 2003)

Another purpose of this pilot study was to address several major findings from the literature review. Major findings from the literature review include:

- The four key concepts of data management are not clearly distinguished, defined, and clarified across institutions, organizations, & disciplines
- Data management plans (DMP) are required for NSF funding since 2011
- The White House OSTP issued public access notice memorandum to all government department heads in 2013
- Data management plans (DMP) are now required for NEH Office of Digital Humanities starting in 2013
- Researchers/scientists need DMC & DMP education & resources
- DMC activities and practices vary within and across multiple disciplines
- Multiple disciplines face massive data management and storage issues

Research Design & Methods

This research used a survey questionnaire quantitative research method. The questionnaire included 10 questions and covered several categories. The several categories include questions on (1) culture, (2) funders and stakeholders, (3) organization, (4) data management and curation key concepts, (5) theoretical frameworks/perspectives, (6) data management plans, (7) data seal of approval assessment guidelines, (8) primary organization designation, and (9) primary individual role.

Population

The sample population for this research study included representatives from the government funding agencies, iSchools, higher education institutions, commercial, publishing, and non-profit industries. Survey participants were selected from formal and informal professional networking contacts that included international & national conferences proceedings, academic faculty, government funding agencies, and professional list serves.

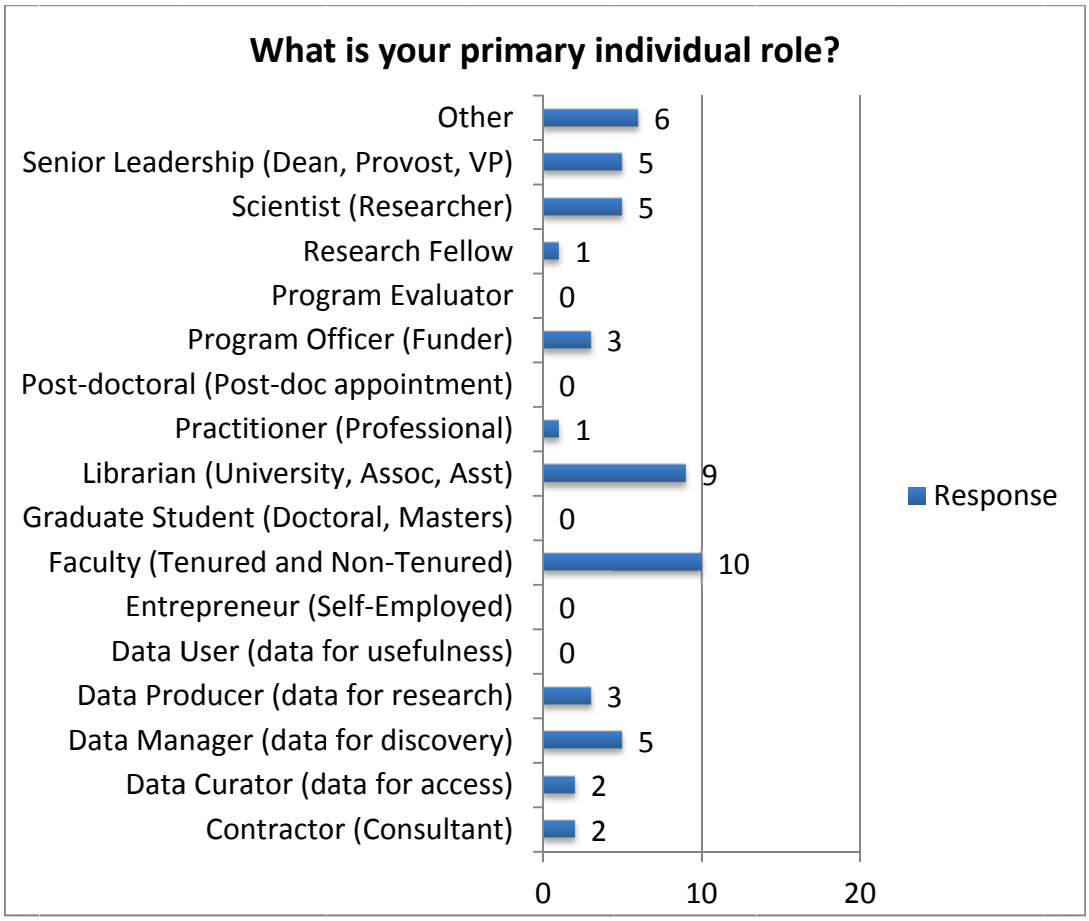


Table 1: Survey Participants' Roles

Findings

The data management and curation (DMC) framework (see Figure 1) developed from the findings from the category on data management and curation key concepts is featured first followed by findings from the other categories.

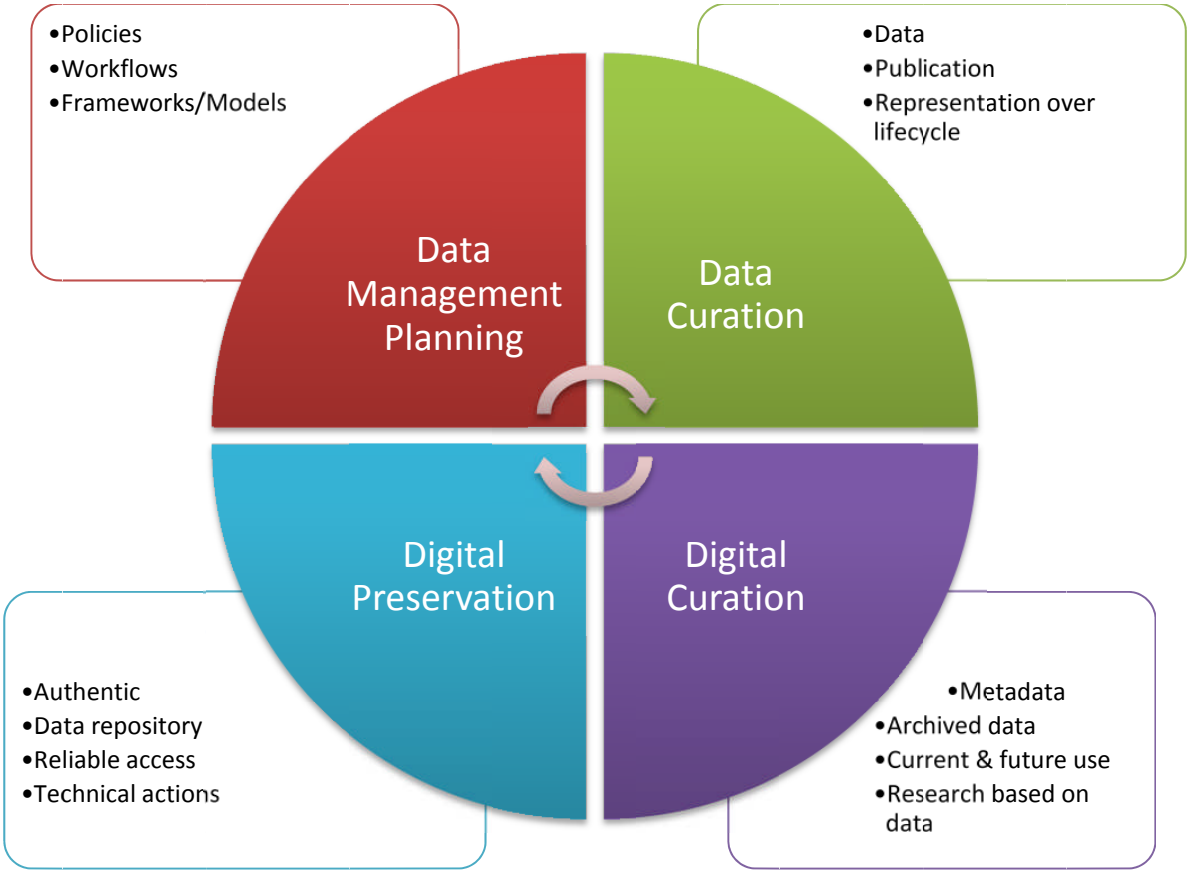


Figure 1: Data Management and Curation (DMC) Framework

Each of the key concepts in the DMC Framework can be mapped to existing data management models and frameworks. The (1) data management planning key concept can be mapped to the National Science Foundation (NSF) data management plan requirement and the DCC Curation Lifecycle Model (DCC, 2007/2014); (2) data curation key concept can be mapped to the level one data curation traditional academic flow of information (Lord, 2003); (3) digital curation key concept can be mapped to the level two data curation information flow with data preservation (Lord, 2003); and (4) digital preservation key concept can be mapped to the level three data curation information flow with data archiving (Lord, 2003). The goal of the DMC Framework is to convey the identification, clarification, and interrelation of the key concepts for theory development (Merton, 1968). The key concepts represent definitions that varied across institutions, organizations, and disciplines with the advancement of tools, methods, and technology.

Below are findings to the following categories: (1) culture, (2) funders and stakeholders, (3) organization, (4) data management and curation key concepts, (5) theoretical frameworks/perspectives, (6) data management plans, (7) data seal of approval assessment guidelines, and (8) primary organization designation.

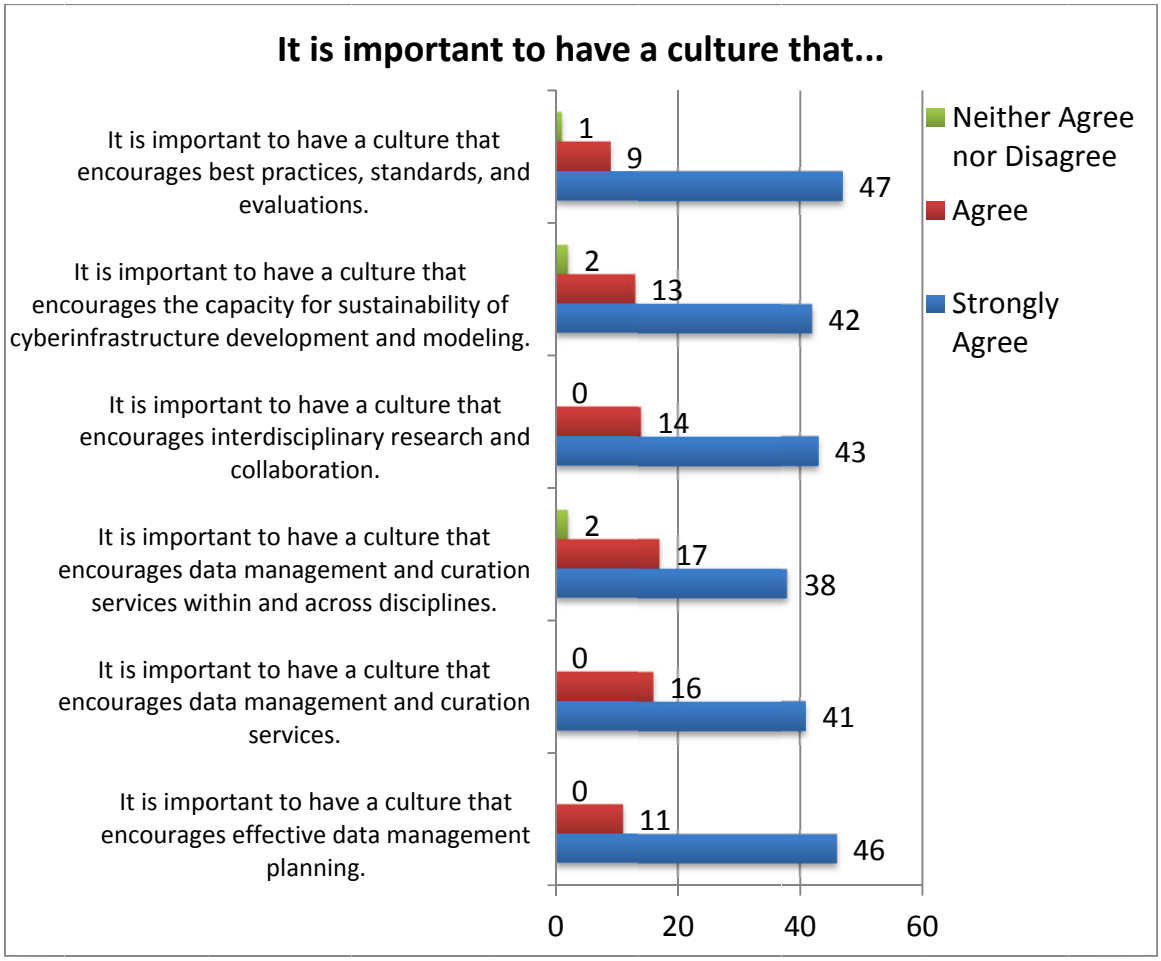


Table 2: Organizational Culture

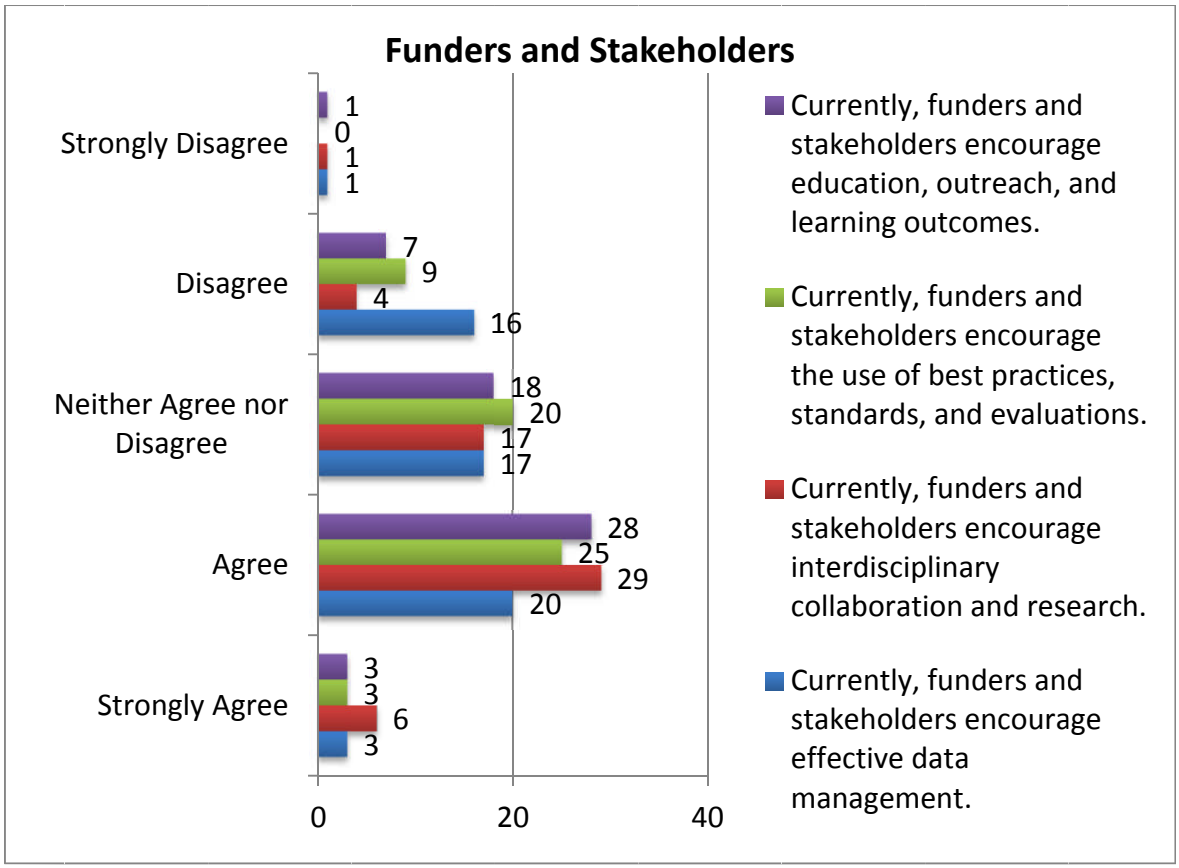


Table 3: Funders and Stakeholders

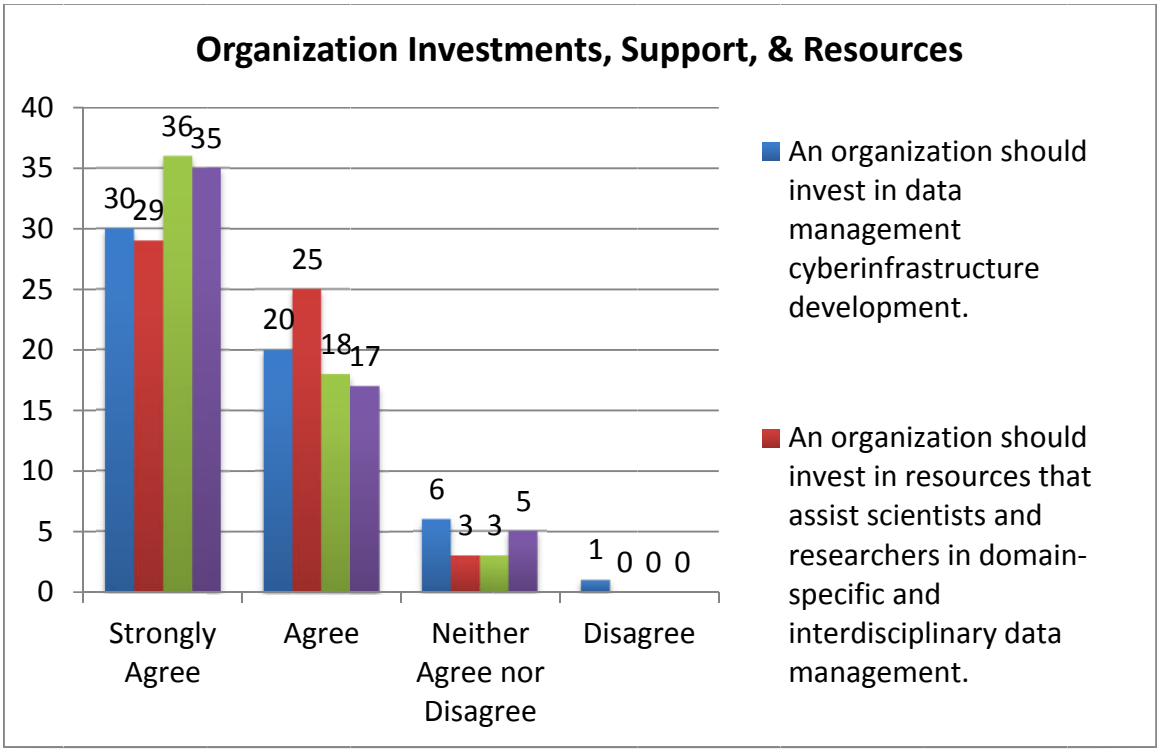


Table 4 – Organization Investments, Resources, & Support

Category 4 - Data management and Curation key concepts

- 55% disagree that data curation is the same as digital curation
- 61% disagree digital curation is the same as digital preservation
- 80% agree data curation, digital curation, and digital preservation are independent yet interrelated concepts
- 77% agree data management and curation services include data curation, digital curation, and digital preservation
- 59% agree there is a need to develop data curation theory from similarities, differences, and interrelationships from multiple competing models or frameworks
- 70% agree there is a need to develop interdisciplinary undergraduate data management and curation services programs

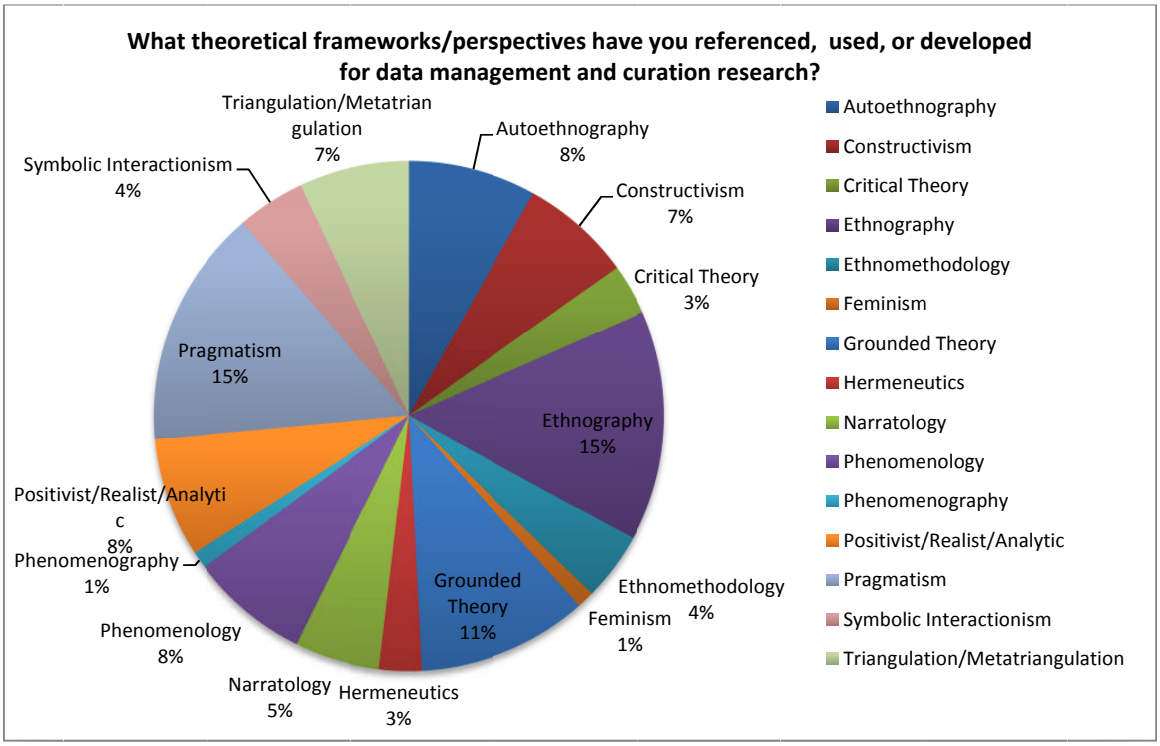


Table 5 - Theoretical Frameworks/Perspectives

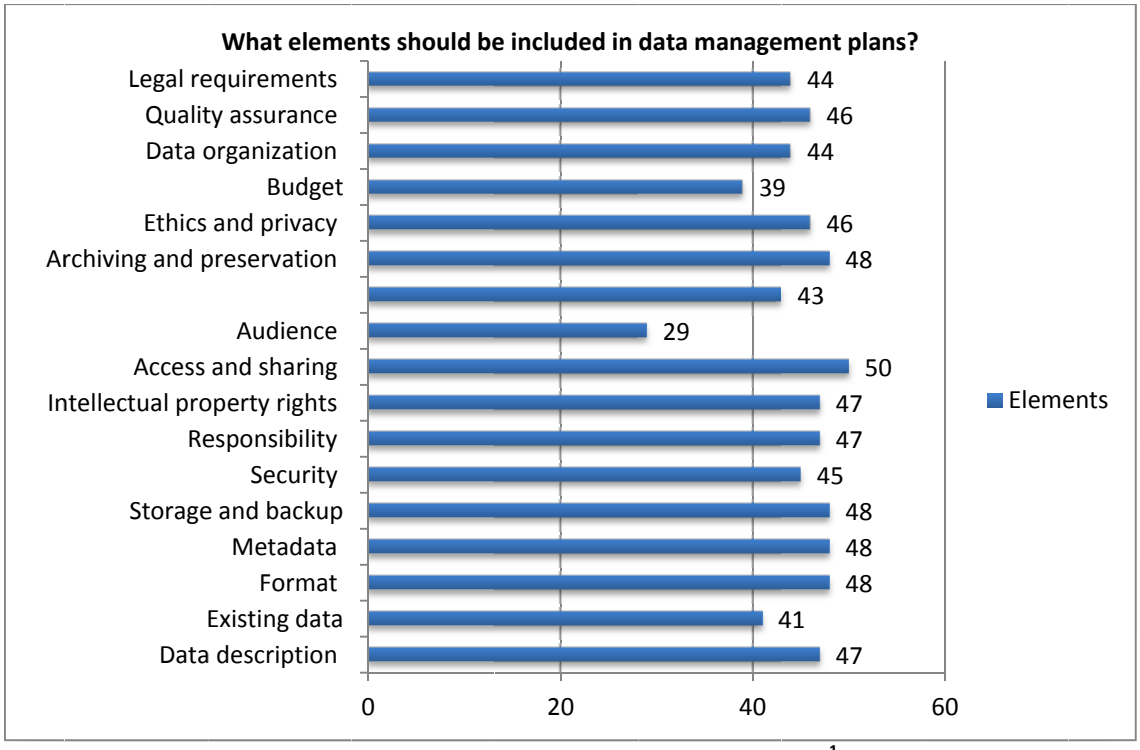


Table 6 – Elements of a Data Management Plan[1]

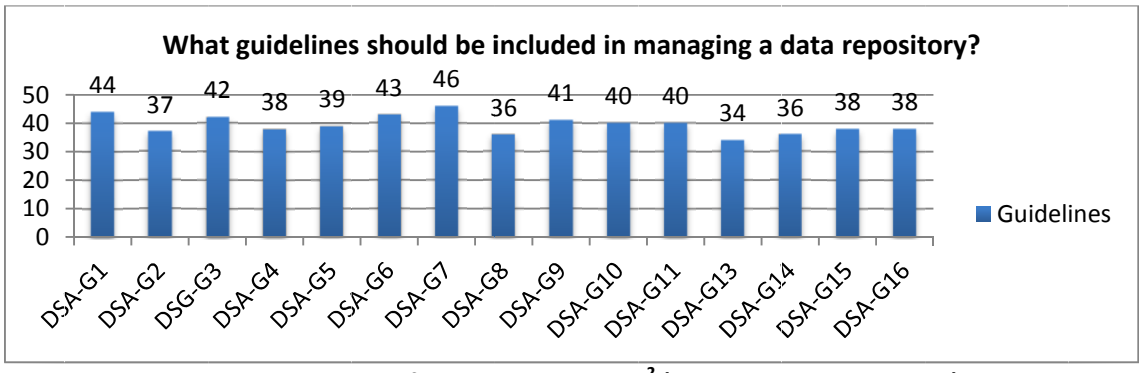


Table 7 - Data Seal of Approval Guidelines[2] **(DSA-G12 omitted in error)**

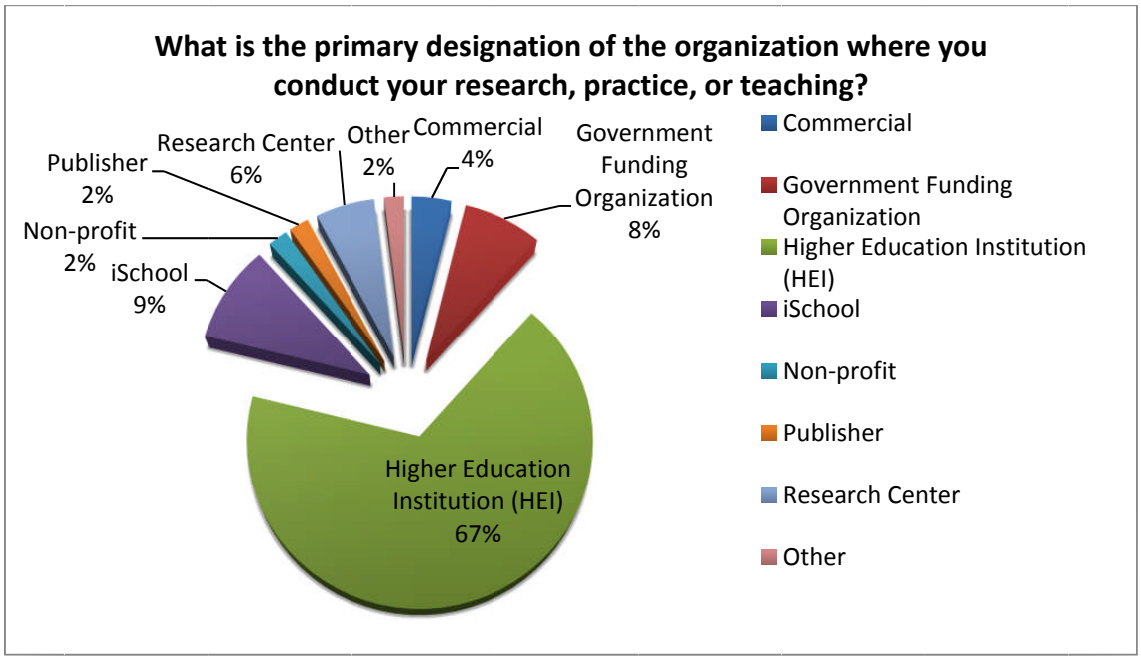


Table 8 – Primary Designation of Organization

The primary discipline or domain expertise of the survey participants include: grey literature (1), sociology (1), large scale (big data) data management (1), data management and curation (6), economics (1), oceanography (1), biomedical informatics (1), Library and Information Science - LIS (11), environmental sciences (1), social science (1), digital humanities (1), information management (1), computer science (2), engineering (1), biology/biodiversity (1), scientific & technical information (1), biological systematics (1).

Originality & Value

This research study combined theoretical frameworks/perspectives (Patton, 2002), elements of a data management plan (ICPSR, 2014), and data seal of approval assessment guidelines (DSA, 2009) into a survey questionnaire for the introduction, education, and articulation of the need for DMC in research & theory development.

Implications

The research, practical, and social implications from this study include:

1. **Research** - stimulate organized research data management awareness and raise data management cognition;
2. **Practical** - improve departmental and institutional level research data management accountability, advance development of data management policies, and support funding agencies data management plan requirement;
3. **Social** - contribute to the wider research data management & scholarly communities.

Limitations

Survey participants that completed the survey did not answer all questions. The responses to questions varied and were inconsistent throughout the survey. This pilot study may lack generalizability to other disciplines. The study includes population selection bias and the responses may not be accurate. The sample size and number of participating disciplines need to be increased for further research.

Conclusions

There is a need to continue data management and curation research and theory development education and articulation within and across multiple disciplinary domains. The DMC Framework can be used for education, research, teaching, and learning: http://platosmith.com/research/research_datamgmtcurate.

References

Blatecky, A. (2012). Open Remarks by Project Sponsors. In The future of scientific knowledge discovery in open networked environments: summary of a workshop, p. 3-5. National Research Council of the National Academies. The National Academies Press: Washington, DC.

DCC. (2007/2014). The DCC Curation Lifecycle Model.

Lord, P. (2003). As cited in Lord & Macdonald.

Lord, P. & Macdonald, A. (2003). E-Science Curation Report: data curation for e-Science in the UK: an audit to establish requirements for future curation and provision.

Merton, R. (1968). Social theory and social structure. The Free Press.

NSF. (2011). Data Management Plan Requirements. National Science Foundation. Retrieved January 21, 2014 from http://www.nsf.gov/eng/general/dmp.jsp.

[1] ICPSR. (2006/2014) Elements of a data management plan. Source: http://www.icpsr.umich.edu/icpsrweb/content/datamanagement/dmp/elements.html.

[2] DSA. (2009/2014-2015). Data Seal of Approval Guidelines. Source: http://www.datasealofapproval.org/en/information/guidelines/.

PopVaT – Popularisation of Science and Technology in Slovakia

Patrícia Stanová, Slovak Centre of Scientific and Technical Information, CVTI SR, Slovakia

Science and Technology are in Slovakia still in an unfavourable position. Competent workers try to change this situation, however, their efforts are not very unified to be able to reach a more significant success. Science and technology must be popularised in many directions. Young people should be motivated to study scientific and technological studies where we have significant lack of students, the humanist and economic studies are much more popular today. We wish to persuade wide public that public resources means paid to science and technology are not lost but they are a logical investment that would return in economic and material growth in the country. We also wish to motivate the scientific community to popularise their works and we wish to persuade them that all their great findings and innovations that are implemented in Slovakia are to be presented to public.

For instance in the Czech Republic there is one priority axis within their Operation Programme Research and Development for Innovation for popularisation of science and technology. They understand very well how important it is to present the results of research and development implemented in their country to the public. In Slovakia popularisation has no priority axis within the Operation Programmes. However, CVTI SR from May 2013 implements the national project aimed to popularisation of science and technology: **PopVaT – Promotion of Science and Technology in Slovakia,** implemented within the Operation Programme Research and Development. The Project will last till October 2015 and its specific objective is: *Increased information of public, including young people, in importance of science and technology and the scientific community on importance of popularisation of sciences.* The Project will be implemented in all eight regions in Slovakia and it is the fifth national project implemented by Slovak Centre of Scientific and Technical Information (CVTI SR).

Project Objectives

The objective of the Project is to increase the understanding of science and technology via selected popularisation activities to three main target groups: wide public, young people and scientific community. Each of the target groups will be addressed by specific popularisation instruments that will be implemented according to experience of NCP VaT at CVTI SR, and other organisations dealing with science popularisation. Popularisation activities would observe different efficiency of media campaigns and popularisation instruments according to particular target groups and population according to regions and will be optimised upon results of analyses and efficiency research.

Target Groups

The objective of the activity in the case of the target group ***Wide Public*** is to popularise science and technology via using efficient promotion methods and instruments and also via area-wide massive media and promotion campaigns in electronic media (radio, TV, the Internet) and also in print media (press, billboards). These activities would help the wide public including industrial sphere to perceive science and technology as a significant part of our life and to understand also necessary financial support to science by the Government as a returnable investment that at the end of the day would have a positive impact and improve economic results in our country and increase life standard of our population.

The objective of the Project for the target group ***Young People*** is to positively influence opinions of young generation on science and technology in innovative, entertaining and stimulative forms and actively influence them in choosing their future jobs in some of the areas in science and technology. Primary school pupils of 8th and 9th grades and secondary schools and grammar schools students will form the target group, for them the question of job selection is the most important. The Project will support motivation of the students and also development of long term cooperation between the secondary schools and the universities and research – development institutions, including the Slovak Academy of Sciences. This would help sustainability of the Project and development of mutual cooperation in area of popularisation of science and technology within particular regions and also at the national level in Slovakia. It is also important to create suitable conditions to search for talented pupils in science and technology, to motivate them to participate in research projects. To motivate universities and research institutions to create suitable conditions for accepting talented pupils to practice mobilities and for implementation of scientific projects.

The objective of the Project for the target group ***Scientific Community*** is their support in communication and promotion of their results and achievements in scientific activities to wide public. Support to scientists and scientific institutions will be implemented mainly with the aim to secure active popularisation of acquired results in research and development. Scientific community will be reasoned to understand the significance and their moral obligation to present results of their scientific – research activities to the public via several popularisation instruments. At this target group we calculate mainly with active work with scientific workers to popularise their scientific activities, publishing scientific works, organising expert events, etc.

Main Activities of PopVat Projet are:

- ***Research – Popularisation Multimedia Programmes***.
  These events will be aimed to pupils of final grades at primary schools, and secondary schools students, and their main goal will be to positively influence the young people in process of choosing their job in area of science and technology.
- ***Audiovisual Works and Other Media Products***.
  The objective of this activity is to create series of audiovisual works aimed to popularisation of science and technology towards wide public.
- ***Internet Promotion***.
  A new Internet webpage will be created aimed to popularisation of science and technology, and various online applications and games aimed to popularisation of science and technology.
- ***Social Networks Communication***.
  WebPages and accounts will be created at social networks such as for instance Facebook, Twitter, LinkedIn, Azet, Youtube and others, and they will secure communication with the target group including active promotion of science and technology.
- ***Popularisation – Expert Events***.
  Within this activity various lectures, seminars, conferences, workshops, a film festival, and the National Award in Science and Technology will be implemented, as well as other activities aimed to popularisation of science and technology among scientific community and also among wide public and young people.
- ***Scientific – Popular Publications***.
  The objective of this activity is to support creation of popularisation – scientific publications and periodic media.
- ***Scientific Center.***
  An interesting “Hands On“ Museum known in the world as the *Scientific Center* will be established. The objective of this Center is to attractively and in and interactive way present to public and mainly to young people how various physical phenomena function, how production process of various products functions (car, television, etc.). The whole activity will be implemented with the aim to motivate young people to study natural scientific and technical studies and to present to the public the hidden process of everyday consumption objects and importance of research and development and innovation for economic growth of the country.

Conclusions

Various institutions deal with popularisation of science and technology such as the NCP VaT at CVTI SR which implement many activities, however, the limitations mainly financial ones form obstacles for implementing complex popularisation activities. The PopVaT Project implemented by CVTI SR, may change much, as it thanks to financial means from the structural funds may help to create a system for popularisation of science and technology in Slovakia. The PopVaT Project will help the scientific community and the science and technology as such to create the area for their presentation in current society that wishes sensations.

„Gold Value“ Offer to Science and Research in Slovakia – is presented by the NISPEZ project and related activities

Mária Žitňanská, Michal Sliacky, Marián Vaňo, and Eva Králiková
Slovak Centre of Scientific and Technical Information, CVTI SR, Slovakia

„DATA IS THE NEW GOLD“, Neelie Kroes, Vice-President of the European Commission responsible for the Digital Agenda

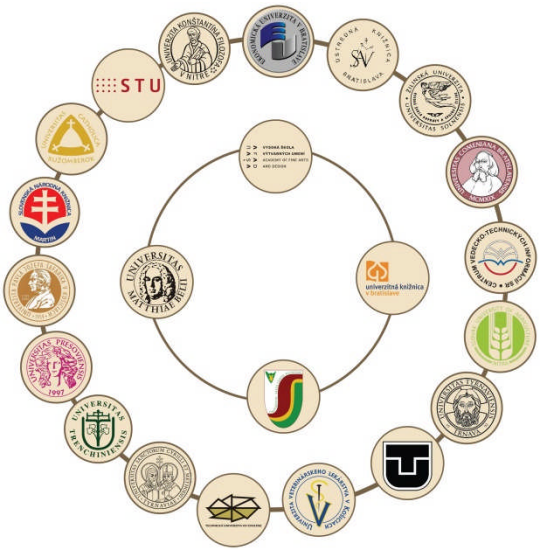
Abstract
The national project activities are aimed at direct support to research and development (R&D) in Slovakia at national level and in long-term horizon. To provide access to a wide offer of world electronic information resources (EIR) is the most important activity. Access to EIR creates a part of a model of completely centralized providing of access to EIR as well as other activities serving to their effective usage. Creating a system for centralized access, search, use and administration of EIR for R & D is the second important project activity.
The third and the same important part of the offer to science and research in Slovakia is represented by building of SciDAP – a central bibliographic database and portal access to the Slovak EIR for research and development. The Slovak scientific and professional journals including open access journals and grey literature documents are concerned. The SciDAP represents a tool to process documents which would be kept for a long time in the institutional repository of SC STI digital documents.
These three activities create a part of the NISPEZ (National Information System to Promote Research and Development in Slovakia – Access to Electronic Information Resources and the Slovak Centre of Scientific and Technical Information is the NISPEZ project solver.

Keywords: NISPEZ, electronic information resources (EIR), for research and development (R&D), centralized access, Scientia.sk – search portal for science and research, SciDAP – Central database and portal access to the Slovak EIR for R&D, open access, grey literature

Introduction
The Slovak Centre of Scientific and Technical Information (SC STI) is a national information centre and a specialized scientific library of the Slovak Republic focused on technical and selected branches of natural, economic sciences and humanities, with key position in information support to science, research and academic sector in Slovakia.
The SC STI is a project solver of several national projects, co-financed from the EU resources, solved under the framework of Research and Development (R&D) Operational programme. Some of the proper goals contribute to fulfilment of one of the strategic objectives of the SC STI – i.e. **complex information support to research and development (R&D) in Slovakia**.
The paper is focused on introduction and outline of one of the implemented projects, directly related to information support to R&D in Slovakia **by means of providing access to appropriate scientific content in a digital version to the wide public**. **The National Information System to Promote Research and Development in Slovakia – *Access to Electronic Information Resources***, known to academic sphere and in all country under the acronym **NISPEZ**, has four specific goals. The paper introduces three of them.

Specific goal 1: Centralized Provision of Access to EIR
During the project implementation, the SC STI followed experience gained by academic libraries and the Slovak Academy of Sciences (SAS) which have provided centralized, consortial purchase of accesses to ***Electronic Information Resources*** (EIR) since the year 2000. In the end of 2008, when the project starting activities were launched, it should be mentioned that **the NISPEZ project has shifted the given issue solution to the systematic level and at it same time contributed to marked extension of EIR portfolio**.

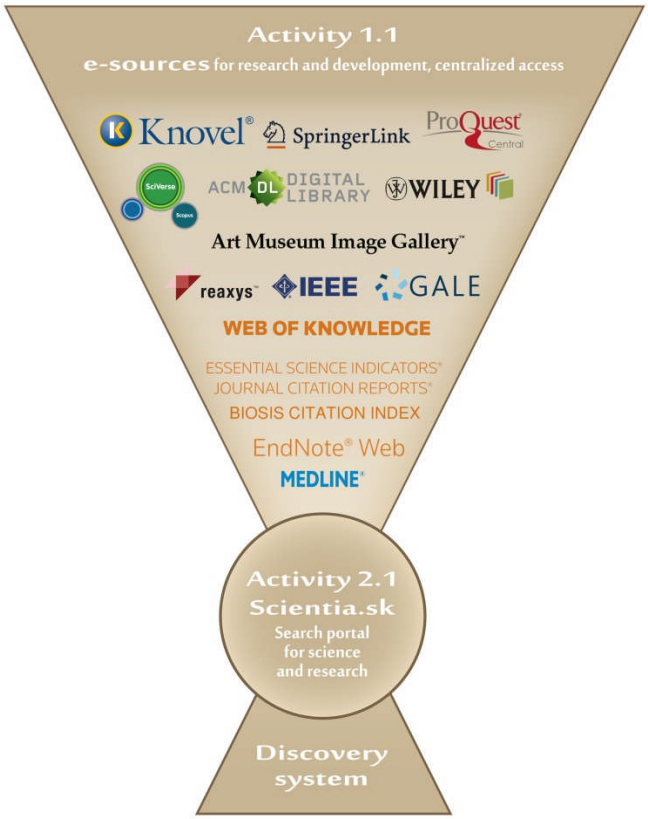
EIR portfolio
With their content, the databases accessed in the framework of the NISPEZ project cover scope of individual scientific branches and are focused mainly on full text information accessed according to the profile of the proper college, university or scientific library to the users of institutions participating in the project (see Fig. 1).



(Fig. 1 – 25 participating institutions)

Bibliographic, citation and scientometric databases of Web of Knowledge and SCOPUS platforms, majority of participating institutions has access to, create the important part of the portfolio. Performance of individual university departments, higher education institutions and universities can be evaluated as well as latest scientific knowledge and development trends can be examined depend necessarily on these databases. Both above mentioned databases are multidisciplinary in nature, mapping the entire field of scientific disciplines, thus they are used by nearly all mentioned institutions. Both databases were chosen by the European Research Council (ERC) as a tool to evaluate performance of researchers and scientists and subsequent allocation of financial funding they represent the first choice among databases to be accessed.

In selecting databases to the EIR NISPEZ portfolio, much attention was given to the needs of the proper participating institutions. It can therefore be stated that the current portfolio (see Fig. 2) of accessed EIR poses an appropriate basis for providing information support to R&D in Slovakia. In case of more special requirements on more profiled EIR it is, however, necessary on the side of universities, other higher education institutions and research centres to cover funding for their procurement by government budget and/or other resources.



(Fig. 2 – EIR portfolio + scientia.sk)

Centralized solution and its contribution

In addition to centralized provision of access to EIR other contributions of complete solution can be identified as follows:

Centralized providing of accesses to EIR helps academic libraries to eliminate oversized agenda related to acquisition of financial funding for purchase of licences on access to EIR.

Centralized evaluation of statistical indicators of EIR usage resulted in unified methodology for their observation which created possibilities for mutual comparison of institutions. There are also backgrounds for more targeted publicity of EIR, organisation of events, seminars resulting in increased information literacy and awareness of EIR users. The mutual comparison of institutions can thus deeper targeting of such events.

Based on the response, it could be pointed out that the evaluation report results together with inquiries could be considered as not only highly important feedback and appropriate tool for evaluate EIR usage effectiveness but for seeking of optimised model for designing future strategies in provision of access to EIR to serve the needs of R&D in Slovakia as well.

Publicity and strengthening of the role of libraries in Slovakia not only in supporting R&D in Slovakia and in society should be emphasized.

Responsibility for the adequate EIR usage would always also on individual institutions which are the primary users of these resources. It is necessary to perform maximum effort to use the EIR effectively and make them real support to research, exploration and university education in Slovakia. Colleges, universities and research centres should stimulate their researchers to consider the EIR a necessity and use them systematically. Libraries can also perform maximum effort in this field, the real need should however follow from R&D community itself in the case high quality research should be reached.

Specific goal 2: National search portal for science and research – scientia.sk

The main goal of the national search portal for science and research – scientia.sk is to enable users to search information simultaneously in heterogeneous remote EIR focussed on science and research. The EIR portfolio consists of both licenced and open access resources. At the same time the issue of effective access of authorized users to fulltext documents is solved.

The portal is based on the products of the ExLibris company: **Metalib – federated search engine, SFX linking server and bX recommendation service.** Its routine operation started in October 2010.

The federated search engine consolidates the retrieved results, eliminates duplicates ranks results by relevance, groups retrieved information info topic clusters as well as by metadata facets, and finally presents them in a unified format. In addition it offers the users further functions and tools.

The main task of the linking server is to dynamically generate links to fulltext documents to which the user institution has the access rights. In addition to the above mentioned function it provides other supplementary services to retrieve documents. The solution includes also the electronic journal portal with content tailored to the individual needs of the project participating institution.

In the framework of the bX recommendation service to the retrieved e-article are presented other relevant e-articles, which has been interesting for other scholars. For the purpose of this bX service data related to article usage are collected from hundreds implementations of the SFX linking server. If recommendations for particular article are available, these are displayed in the SFX service menu. During 2012 the bXHot articles service has been activated offerring 10 most frequently used/read scientific articles in the selected branch of sciences, obtained by means of statistical analyses of behaviour of millions of users for the last month.

The portal´s user community comes primarily from 25 institutions participating in the NISPEZ project. The EIR portfolio creates both licenced (provided in the framework of the project and by individual participants) and open access EIR. At present, there are 128 activated EIR in the portal (a list of licensed e- resources and their accessibility can be found on the NISPEZ project website). Access to the licensed e-resources is regulated by filtering the IP adresses of user computers from which access was made.

Specific goal 3: Central Database of Slovak EIR for R&D

The task of building the **and routine operation SciDAP – Central database of the Slovak Electronic Information Resources for Research and Development,** its optimisation and routine operation was also fulfilled in the NISPEZ project implementation. At the beginning, it was necessary to collect and bibliographically process article records from 72 Slovak scientific journals taken from bibliographic and citation databases in Web of Science (WoS) and Current Content Connect (CCC) platforms. Then other Slovak scientific and professional journals processed in world databases are joined to the SciDAP. It will

be purposeful however to concentrate on Open Access (OA) journals and on the documents from the category of grey literature.
The SciDAP database is designed as a bibliographical tool to create descriptive metadata enabling **bibliographic processing** of journals and other types and kinds of documents (up to analytical description level), as well as searching and browsing. The following materials and entities are concerned:

- grey literature (proceedings, reports, yearbooks, presentations),
- authors, institutions, Slovak EIR publishers and the most important events in the field of science and research,
- observation of journals according to different criteria, e.g. according to publisher: journals published by the Slovak Academy of Sciences (SAS), colleges and universities, research and commercial organizations, according to verification by official resources: WoS, SCOPUS, OpenAIRE – European Initiative for Open Access Policy, DOAJ – directory of open access journals, DASPER – planned database of verified Slovak scientific and professional journals aimed at publication activity evaluation).

The SciDAP should enable also creating of bibliographic record outputs for users in compliance with ISO 690 standard.
Comprehensive solution of the grey literature issue should be given highest attention. In addition to continuous mapping, selection and processing also long-term preservation in the prepared repository at the SC STI in compliance with the appropriate legislation. Accessing of grey literature database by means of planned discovery system will be of concern.

National Point of Reference for the area of Open Access in Slovakia
Since 2013 the SC STI fulfils the role of the national point of reference for the area of Open Access to and preservation of scientific information – OA. Based on the appeal of the European Commission the Ministry of Education, Science, Research and Sport empowered the SC STI to perform this function.
Tasks and competences following from this activity shift the SC STI to the level of national authority for OA area in Slovakia. In addition to the above mentioned tasks the SC STI considers as obligation to regularly follow and map OA documents of Slovak origin generated by the Slovak R&D community, to process them in detail, preserve and of course access them to the users (SciDAP).
The final accessing of the full texts as well as bibliographic data to the users should be in compliance with copyright and supported by licences for all document types (R&D journals, grey literature, OA journals, etc.).

Conclusions
The paper has revealed three specific goals of the NISPEZ project having one common objective: to implement information support to R&D in Slovakia by means of providing access to appropriate scientific content in electronic version. Once again, i.e. a centralized solution providing accesses to outstanding scientific EIR portfolio all around the world, serving as an effective tool to facilitate the users orientation and search in these resources (scientia.sk search portal) or building of a database of the Slovak scientific EIR (SciDAP database). The NISPEZ project will terminate in 2014. Its contribution and bequest is of high importance for the Slovak science. The common feature is seeking for strategies and solutions of initiated goals and tasks.

In this context, leading issues and challenges concerning **issues of financial means to cover licences for access to EIR, particularly in long-term horizons, and also issues related to adequate management of administrative agenda,** should be mentioned. After nearly four year implementation of the NISPEZ project it can be stated, that the SC STI as its main solver – gained a lot of experiences with solution of centralized provision of access to EIR at national level which can result in challenging initiatives and inspirations in the given area.

Upgrade to discovery system can be undoubtedly concerned as the challenge for scientia.sk search portal for science and research, providing information gold value to users more simply, user friendly and quickly, provided by the fact that EIR portfolio is at least at the current state level.

Cooperation of the SciDAP bibliographic database with the planned SC STI repository can be viewed as added value to information support to R&D community in the strategic horizon. In this connection, SciDAP represents **a tool for bibliographic processing of the selected document type up** to description level and repository represents a tool for permanent surviving of full texts of these documents, and finally discovery system for searching and accessing information/documents to scientists and researchers.

References

Recommended Literature

BIROVÁ, Daniela – ŽITŇANSKÁ, Mária. 2009. Národný informačný systém podpory výskumu a vývoja na Slovensku – prístup k elektronickým informačným zdrojom – *INFOS 2009 : zborník príspevkov z 35. medzinárodného informatického sympózia INFOS 2009. Pamäťové inštitúcie v digitálnom prostredí. Stará Lesná, 27. – 30. apríla 2009.* 1. vyd. – Bratislava : Spolok slovenských knihovníkov, 2009. – ISBN 978-80-969674-3-8. – s. 9 – 19.

BIROVÁ, Daniela – ŽITŇANSKÁ, Mária. 2009. Národný informačný systém podpory výskumu a vývoja na Slovensku – prístup k elektronickým informačným zdrojom. – *Bulletin Slovenskej asociácie knižníc*. – ISSN 1335-7905. – Roč. 17, č. 1 (2009), s. 10 – 15.

SLIACKY, Michal. 2010. Efektívnosť využívania e-zdrojov NISPEZ – dotazníkový prieskum. – *ITLib. Informačné technológie a knižnice* [online]. ISSN 1336-0779. – Roč. 14, č. 4 (2010), s. 57 – 58. ISSN 1336-0779. Available online at: http://www.cvtisr.sk/itlib/itlib103/sliacky.htm.

ŠEFČÍK, Michal – VAŇO, Marián. 2010. Nový vyhľadávací portál pre vedu a výskum. – *ITLib. Informačné technológie a knižnice* [online]. ISSN 1336-0779. – Roč. 14, č. 4 (2010), s. 57 – 58. [cit. 2011-01-31]. Available online at: http://www.cvtisr.sk/itlib/itlib104/sefcik_vano.htm.

TURŇA, Ján – ŽITŇANSKÁ, Mária – CVIK, Oleg – VAŇO, Marián – KRÁLIKOVÁ, Eva – BIROVÁ, Daniela. 2009. Informačná podpora výskumu a vývoja na Slovensku. – *INFORUM 2009, 15. ročník konferencie o profesionálnych informačných zdrojoch. Praha, VŠE, 27. – 29. mája 2009*.
Available online at: http://www.inforum.cz/archiv/inforum2009/sbornik/78/.

ŽITŇANSKÁ, Mária: Národný informačný systém podpory výskumu a vývoja na Slovensku – prístup k elektronickým informačným zdrojom. – *ITLib. Informačné technológie a knižnice*. – ISSN. 1335-793X. - Roč. 13, č. 1 (2009), s. 20 – 29. Available online at:
http://www.cvtisr.sk/itlib/itlib091/zitnanska.htm.

ŽITŇANSKÁ, Mária – SLIACKY, Michal – VAŇO, Marián. 2011. Centralizovaný nákup EIZ na Slovensku – máme šancu pokračovať? – INFORUM 2011, 17. ročník konferencie o profesionálnych informačných zdrojoch. *Praha, VŠE,* 24. – 26. mája 2011. Available online at: http://www.inforum.cz/cs/sbornik.

ŽITŇANSKÁ, Mária – SLIACKY, Michal. 2012. Centralizovaný prístup k EIZ, jeho optimálne nastavenie pre efektívne využívanie ako nástroja pre strategické rozhodovanie alebo štvrtý rok projektu NISPEZ. – *INFORUM 2012 : 18. ročník konference o profesionálních informačních zdrojích, Praha, VŠE, 22. – 24. května 2012 [online]*. Praha : Albertina icome Praha, 2012. Available online at: http://www.inforum.cz/cs/sbornik. ISSN 1801–2213.

ŽITŇANSKÁ, Mária – SLIACKY, Michal – VAŇO, Marián – KRÁLIKOVÁ, Eva. 2013. NISPEZ a jeho pozícia v informačnej podpore výskumu a vývoja na Slovensku – *INFOS 2013 : zborník príspevkov z 37. medzinárodné informatické sympózium o postavení a úlohách pamäťových inštitúcií v oblasti rozvoja kultúry, vedy, techniky a vzdelávania. Stará Lesná, 8. – 11. apríla 2013.* 1. vyd. – Bratislava : Spolok slovenských knihovníkov, 2013. Available online at:
http://www.infolib.sk/sk/spolok-slovenskych-knihovnikov/infos-2013/infos-2013.html

ŽITŇANSKÁ, Mária, KRÁLIKOVÁ, Eva, SLIACKY, Michal. 2013. SciDAP : Centrálna databáza slovenských elektronických informačných zdrojov pre výskum a vývoj, jej prínos a perspektívy rozvoja pre podporu VaV na Slovensku In: INFORUM 2013 : sborník z 19. ročníka konferece o profesionálních informačních zdrojích. Praha, VŠE, 21. – 22. května 2013 [online]. – Praha : AIP, 2013. – nestr. Available online at: http://www.inforum.cz/sbornik/2013/34.

We promote R&D activities in Slovakia / The project is co-financed by the EU resources
Podporujeme výskumné aktivity na Slovensku / Projekt je spolufinancovaný zo zdrojov EÚ

LIBRARY OF CONGRESS
FEDLINK

Pre-Conference Announcement

Sixteenth International Conference on Grey Literature

Grey Literature Lobby: Engines and Requesters for Change

● Think Tanks ● Advocacy Groups ● Research Teams ● NGOs ● IGOs ● Foundations ● Coalitions
● Lobbyists ● Special Interest Groups ● Task forces ● Watch Dog Groups ● Associations ● Etc.

Federal Library Information Network | FEDGrey Working Group

GL16 Program and Conference Bureau

TextRelease

Javastraat 194-HS, 1095 CP Amsterdam, The Netherlands
www.textrelease.com ▪ conference@textrelease.com
Tel/Fax +31-20-331.2420

FEDLINK

an organization
of federal agencies
working together
to achieve optimum use of
resources and facilities
of federal libraries
and information centers
by promoting
common services,
coordinating and sharing
available resources, and
providing continuing
professional education.

LIBRARY OF CONGRESS
FEDLINK

101 Independence Ave, SE ~ Washington, DC 20540-4935
FEDLINK Main Number (202) 707-4800
FEDLINK Hotline (202) 707-4900

List of Participating Organisations

Alberta Health Services, Cancer Care	Canada
Australian Council for Educational Research, ACER	Australia
Banque de données en santé publique	France
Biblioteca Centrale "G. Marconi"; CNR	Italy
Central Economics and Mathematics Institute, CEMI	Russia
Centre National de Recherché Scientifique, CNRS	France
Centre of Information Technologies and Systems, CITIS	Russia
Comenius University	Slovakia
Dschang University Library	Cameroon
EBSCO Publishing	United States
l'Ecole des hautes études en santé publique, EHESP	France
Emerald	United Kingdom
European Organization for Nuclear Research, CERN	Switzerland
Federal Library Information Network, FEDLINK	United States
Florida State University, FSU	United States
Grey Literature Network Service, GreyNet International	Netherlands
Health Information Network Calgary, HINC	Canada
InfoEthics UK	United Kingdom
InfoKey	Slovakia
Information Processing Center, OPI	Poland
Institut de l'Information Scientifique et Technique, INIST	France
Institute of Computational Linguistics, ILC	Italy
Institute of Economics and Industrial Engineering, IEIE	Russia
Institute of Information Science and Technologies, ISTI	Italy
Institute of Research on Population and Social Policies, IRPPS	Italy
International Atomic Energy Agency, IAEA	Austria
Japan Atomic Energy Agency, JAEA	Japan
Japan Science and Technology Agency, JST	Japan
Keith G Jeffery Consultants	United Kingdom
Korea Institute of Science & Technology Information, KISTI	Korea
Library of Congress; African and Middle Eastern Division	United States
Market Economy Institute	Russia
Naples University, Physics Department	Italy
National Documentation Center, EKT	Greece
National Hellenic Research Foundation	Greece
National Research Council, CNR	Italy
National Library of Technology, NTK	Czech Republic
New York Academy of Medicine, NYAM	United States
Nuclear Information Section, NIS-IAEA	Austria
Oxford Business Intelligence	United Kingdom
Pratt Institute, School of Information and Library Science	United States
PricewaterhouseCoopers, PwC	Netherlands
Russian Academy of Sciences, RAS	Russia
Slovak Centre of Scientific and Technical Information, CVTISR	Slovakia
Suweco	Czech Republic
SVOP Ltd	Slovakia
Swinburne University of Technology, SWIN	Australia
TextRelease, Program and Conference Bureau	Netherlands
TIND Technologies	Norway
Thomson Reuters	Slovakia
Université Charles de Gaulle Lille 3	France
University of Bergen, UiB	Norway
University of Calgary	Canada
University of Iceland, School of Social Sciences	Iceland
Victoria University, VU	Australia

Author Information

Massimiliano **Assante** holds a master degree (M.Sc.) on Information Technologies received from the University of Pisa, where he is undertaking a Ph.D. on Information Engineering. He is research staff at the Istituto di Scienza e Tecnologie dell'Informazione "Alessandro Faedo" (ISTI), an institute of the Italian National Research Council (CNR). He joined ISTI in 2007 and is currently member of the iMarine EU Project and EUBrazilOpenBio Project. In the past he has been member of D4Science II, D4Science, DILIGENT and DRIVER European Projects. His research interests include Data Infrastructures, Next Generation Digital Libraries, Information Systems and NoSQL Data Stores. Email: massimiliano.assante@isti.cnr.it

Anne **Asserson** holds a master from the University of Bergen, UiB. She has been working with Research Documentation, and has participated in substantial parts of CRIS developmental work, locally and nationally since 1992. Anne Asserson has been part of the establishing and implementing of several CRIS both at the UiB and nationally. For several years she was the chairwoman of the Steering Group of the national CRIS system and project secretary of a National system for academic administration. Anne Asserson is presently representing UiB in the national project group of CRIStin. She has also participated in The CORDIS funded European-wide project on " Best Practice" 1996 and was a member of the working group set up 1997 that produced the report CERIF2000 Guidelines (1999) www.cordis.lu/cerif, coordinated by the DGXIII-D4. euroCRIS is now the custodian of the CERIF model www.eurocris.org. Anne Asserson is a member of the euroCRIS board with the responsibility Member Strategy and External Relations. anne.asserson@fa.uib.no

Stefania **Biagioni** graduated in Italian Language and Literature at the University of Pisa and specialized in data processing. She is currently a member of the research staff at the Istituto di Scienza e Tecnologie dell'Informazione "Alessandro Faedo" (ISTI), an institute of the Italian National Research Council (CNR) located in Pisa. She is head librarian and member of the ISTI Networked Multimedia Information Systems Laboratory (NMIS). She has been the responsible of ERCIM Technical Reference Digital Library (ETRDL) and currently of the PUMA (PUblication MAnagement) & MetaPub, a service oriented and user focused infrastructure for institutional and thematic Open Access repositories looking at the DRIVER vision, http://puma.isti.cnr.it. She has coauthored a number of publications dealing with digital libraries. Her activities include integration of grey literature into library collections and web access to the library's digital resources, including electronic journals and databases. She is a member of GreyNet since 2005. Email: stefania.biagioni@isti.cnr.it

Lubomir **Bilsky** started his professional career in 2001 at the Business and Innovation Centre Bratislava, where he was responsible for international projects implementation. He was an active member of the Innovation Relay Centres (IRC) Network dealing with the support of transnational research and technological co-operation development. In 2008, he became the head of Innovation section within the Enterprise Europe Network Slovakia consortium, an EC initiative supporting business, innovation and research co-operation of research organisations and SMEs. At CVTI SR, he has been responsible for preparation and implementation of several national and international projects focused on support of scientific community in Slovakia, including active promotion of science and technology in society. He also co-ordinates activities related to building the national infrastructure for technology transfer support in Slovakia. Email: lubomir.bilsky@cvtisr.sk

Leonardo **Candela** is a researcher at Networked Multimedia Information Systems (NeMIS) Laboratory of the Italian National Research Council - Institute of Information Science and Technologies (CNR-ISTI). Graduated in Computer Science in 2001 at University of Pisa and completed a PhD in Information Engineering in 2006 at University of Pisa. Joined the NeMIS Laboratory in 2001 and was involved in various EU-funded projects including CYCLADES, Open Archives Forum, DILIGENT, DRIVER, DELOS, D4Science, D4Science-II and DL.org. He was a member of the DELOS Reference Model Technical Committee and of the OAI-ORE Liaison Group. He is currently involved in the iMarine and EUBrazilOpenBio projects. His research interests include Digital Library [Management] Systems and Architectures, Digital Libraries Models, Data Infrastructures. Email: leonardo.candela@isti.cnr.it

Donatella **Castelli** joined the CNR-ISTI in 1987. Since then she has participated actively in several EU and nationally funded projects on Digital Libraries and Research Infrastructures. Among them: “ERCIM Digital Library-DELOS”, DELOS ECHO, SCHOLNET (as scientific co-ordinator), CYCLADES, Open Archives Forum, DLib Competence Center, DILIGENT (as scientific co-ordinator) and DRIVER. Currently she is the Scientific Co-ordinator of the D4Science and D4Science-II projects. She is also involved in the BELIEF-II, DRIVER-II, OpenAIRE, DL.org and GRDI2020 FP7 projects. Her research interests include digital libraries content modelling, interoperability and architectures.donatella.castelli@isti.cnr.it

Marta **Dušková** studied library and information science at Comenius University in Bratislava (Slovakia). Since July 2010 she works in the Slovak Centre of Scientific and Technical Information in Bratislava (Slovakia) in Publication Evaluation Department. She deals with grey literature and The Central Registry of Publication Activity. She coordinates activities associated with obtaining and making grey literature available and cooperates with the processing and verification data publications included in The Central Registry of Publication Activity. From 2012 she is studying PhD study at Comenius University in Bratislava (Slovakia) with the theme: Communication of knowledge through of grey literature. Email: marta.duskova@cvtisr.sk

Dominic **Farace** is Head of GreyNet International and Director of TextRelease, an independent information bureau specializing in grey literature and networked information. He is a native Louisianan and holds degrees in sociology from Creighton University (BA) and the University of New Orleans (MA). His doctoral dissertation in social sciences is from the University of Utrecht, The Netherlands, where he has lived and worked since 1976. After six years heading the Department of Documentary Information at the Royal Netherlands Academy of Arts and Sciences (SWIDOC/KNAW), Farace founded GreyNet, Grey Literature Network Service in 1992. He has since been responsible for the International Conference Series on Grey Literature (1993-2013). In this capacity, he also serves as Program and Conference Director as well as managing editor of the conference proceedings. Farace is editor of The Grey Journal (TGJ) and provides workshops and training in the field of grey literature. Email: dominic.farace@textrelease.com

Jerry **Frantzen** graduated in 1999 from the Amsterdam University of Applied Sciences/Hogeschool van Amsterdam (AUAS/HvA) in Library and Information Science. Frantzen is the technical editor of The Grey Journal (TGJ). And, since 1996, he is affiliated with GreyNet, Grey Literature Network Service, as a freelance technical consultant. info@greynet.org

Silvia **Giannini** graduated and specialized in library sciences. Since 1987 she has been working in Pisa at the Institute for the Science and Technologies of Information “A. Faedo” of the Italian National Council of Research (ISTI-CNR) as a librarian. She is a member of the ISTI Networked Multimedia Information Systems Laboratory (NMIS). She is responsible of the library automation software “Libero” in use at the CNR Research Area in Pisa and coordinates the bibliographic and managing activities of the ISTI library team. She cooperates in the design and development of the PUMA (PUblication MAnagement) & MetaPub, an infrastructure software for institutional and thematic Open Access repositories of published and grey literature produced by CNR. silvia.giannini@isti.cnr.it

Mária **Harachová** studied English and Slovak languages and literature at the Comenius University in Bratislava (Slovakia). During her career practice attended number of courses for information specialists in the field of industrial property, patent information sources and patent searches. Since 1986 she has been working at the Slovak Centre of Scietific and Technical Information in various positions: an editor in chief of technical newsletters, an information specialist in the EU information sources and patent information, head of the department aimed at technology transfer information support. At present she is also involved in implementation of national projects: National Information System Promoting Research and Development in Slovakia – Access to Electronic Information Resources and National Infrastructure for Technology Transfer in Slovakia. Email: maria.harachova@cvtisr.sk

Andrea **Hrčková** has been an internal PhD. student at the Department of Library and Information Science, Comenius University in Bratislava since 2010. She is writing dissertation thesis „Discussion Groups as an Information Resource“ and co-created two university textbooks. She also

Author Information

completed an internship as an Information and Knowledge Management Advisor in NATO (2011-2012). Andrea is currently a member in the research project Tradice, where she investigates the information behavior of different PhD. students. Email: hrckova.andrea@gmail.com

Kiyoshi **Ikeda** works as librarian at the Central Library of Japan Atomic Energy Agency (JAEA). He also currently works as the chief of International Nuclear Information System (INIS) and Nuclear Information Section of JAEA. Since 2011, he is in charge of selecting and classifying for INIS database input data. He joined JAERI (former JAEA) in 2004, and had been working for acquisition of library materials (domestic/foreign books, journals and technical reports). Between 2007 and 2010, he had been working for editing work at "JAEA Reports" and "JAEA R&D Review" which are grey literatures. He holds a degree in library and information science (M.A.) from the Graduate School of University of Library and Information Science (in Japan). Email: ikeda.kiyoshi@jaea.go.jp

Jana **Ilavská** holds a master degree (Mgr.) from Comenius University in Bratislava, where she is undertaking a PhD. on Librarianship and Information Science. Her research interests are oriented on bibliometrics, scientometrics, infometrics and related topics including repositories. Since 2001 she has been working in Academic Library of Comenius University as system administrator and librarian. She is experienced in database systems, metadata and knowledge organization. She participates on several university or national projects, including building of Central Registry of Publication Activities of Slovak universities. Email: ilavska@rec.uniba.sk

Keith **Jeffery** is an information consultant. He is the former Director of International Relations at STFC (Science and Technology Facilities Council) based at Rutherford Appleton Laboratory. Keith previously had strategic and operational responsibility for ICT with 360,000 users, 1100 servers and 140 staff. Keith holds 3 honorary visiting professorships, is a Fellow of the Geological Society of London and the British Computer Society, is a Chartered Engineer and Chartered IT Professional and an Honorary Fellow of the Irish Computer Society. Keith is currently President of ERCIM and President of euroCRIS, and serves on international expert groups, conference boards and assessment panels. He had advised government on security and green computing. He chaired the EC Expert Groups on GRIDs and on cloud computing. Email: keith.jeffery@keithgjefferyconsultants.co.uk

Stefanía **Júlíusdóttir** is an Assistant Professor in the Department of Library and Information Sciences, Faculty of Human and Social Sciences at the University of Iceland. Her main fields of teaching and research are systematization of knowledge and information (i.e. cataloguing, classification and indexing), retrieval of knowledge and information, and the development of the subject field of collections of recorded knowledge and information (i.e. in libraries, records centres and archives). Moreover, her research field includes the effects of e-publishing on public access to publications (particularly grey literature), as stipulated in Icelandic legal deposit acts. The topic of her PhD thesis on which she is presently working, focuses on the effects of developments in the research areas named above on career opportunities and working conditions of staff in libraries, records centres and archives in Iceland. Her professional experiences include Director of the Library of the Icelandic Building Research Institute, Librarian in the National Department of the National Library of Iceland working on the National bibliography, Library Director (of public and school libraries) at the Ministry of Culture and Education, Library Director at the Directorate for Health in Iceland. Education 1974 BA-in Librarianship and Biology at the University of Iceland, MS-degree in Librarianship from Columbia University in the City of New York, PhD studies at Åbo Akademi in Finland and at the University of Iceland. Email: stj12@hi.is

Mikhail R. **Kogalovsky** is leading researcher at Market Economy Institute of Russian Academy of Sciences in Moscow, member of editorial boards of a number of scientific journals, co-chair and PC member of scientific conferences, secretary of Moscow ACM SIGMOD Chapter. Mikhail graduated from the Saratov State University, Mathematics and Mechanics Department, and got his Ph.D. (CS) from the Central Economical and Mathematical Institute of Russian Academy of Sciences. He is author more 200 articles and books, translator and scientific editor of translations to Russian of wide-known books by Jeffry Ullman, Cris Date, Alan Simon, CODASYL and ANSI/X3/SPARC reports. His research concerns scientific digital libraries and scientometrics, data management, conceptual and ontological modeling, Web technologies and XML standards, information resources integration. Mikhail has been associate professor (half day) of Moscow State University during 25 years. Email: kogalov@gmail.com

Marcela **Kopecká** studied Library and Information Science at Comenius University in Bratislava, Slovakia and she works as an information specialist in the Academic Library at the same university. She is a student of PhD degree at Comenius University in Bratislava. In her research she concentrates on e-publishing and the usage of institutional repositories in academic communication. She participates in research projects on information behaviour of PhD students. Other areas of interest include Open Access, repository development, metadata of open archives and open repositories, data interoperability and digital libraries.
Email: kopecka@rec.uniba.sk

Xanthoula **Kostaras** is the manager of the Guideline Utilization Resource Unit (GURU), a knowledge management program within CancerControl Alberta in Alberta, Canada. GURU supports provincial tumour teams in the development, implementation, and evaluation of clinical practice guidelines. She received an MSc in Community Health Sciences from the University of Calgary in 2008, with a specialization in epidemiology. Her research interests relate to strategies to improve guideline implementation and adherence, as well as the use of knowledge transfer strategies to support cancer patients to receive care that is based on best evidence and best practices.
Email: bxanthoula.kostaras@albertahealthservices.ca

Eva **Králiková** graduated from Faculty of Philosophy, Comenius University in Bratislava – Department of Library Science and Scientifical Informations. She has been working in the Central Library of Slovak Academy of Sciences in Bratislava for 10 years dealing with electronic information resources, reference and research services. Currently she is working in the Slovak Centre of Scientific and Technical Information in the Electronic Information Resource Department.
E-mail: eva.kralikova@cvtisr.sk

Miroslav **Kubiš** works in Slovak Centre of Scientific and Technical Information since 2012 as a Head of Department of Technology transfer and project manager of national project: The national infrastructure for technology transfer support in Slovakia – NITT SK.
Email: miroslav.kubis@cvtisr.sk

Ľubomír **Kucka** graduated from Library and Information Science at the Faculty of Philosophy of the Comenius University in Bratislava. In 1983 he obtained a Doctor of Philosophy degree. He started his career at the Slovak Technical Library in the Patent Literature Department. In the period of 1993 - 1996 he acted as a Director of Patent Information Department in the Industrial Property Office of the SR. From 1999 to 2007 he was at the position of the director of the Slovak Centre of Scientific and Technical Information in Bratislava. He attended trainings organised by the EPO and is an active lecturer in several courses in the field of industrial property information. At present he is a head of PATLIB centre. His professional interest is focused on industrial property information, search in databases and information services.
Email: lubomir.kucka@cvtisr.sk

Seon-Hee **Lee** is Senior Researcher at the Korea Institute of Science and Technology Information (KISTI). She has a master's degree in Library and Information Studies (MLIS) from the University of California, Los Angeles (1996) and a Master of Art degree in Philosophy from Ewha Women's University (1988). Her research interests include: collection development, grey literature, e-journals, information services, and collaborative digital reference services (CDRS). Email: wisdom@kisti.re.kr

Victor **Lyapunov** is a member of the Socionet work group about 15 years. He is the main software developer and server administrator of the Socionet system. Victor's permanent employment is the Institute of Computational Mathematics and Mathematical Geophysics, Siberian Branch of RAS (Novosibirsk, Russia) and also he is a member of Socionet software team at the Institute of Economics and Industrial Engineering, Siberian Branch of RAS (Novosibirsk). Email: vic@ieie.nsc.ru

Author Information

Emily **MacLeod** is one of the Tumour Group Facilitators for the Guideline Utilization Resource Unit (GURU). Her background of health sciences support her role as a registered nurse focused on clinical practice guideline implementation. Emily collaborates with knowledge management specialists and members of provincial tumour teams to create communication strategies, support knowledge transfer and facilitate uptake of evidence based oncology guidelines. E-mail: Emily.MacLeod@albertahealthservices.ca

Elysa **Meek** is one of the Tumour Group Facilitators for the Guideline Utilization Resource Unit (GURU). She is a bachelor-prepared registered nurse who works with the provincial tumour teams and focuses her efforts on strategies to support guideline implementation and evaluation. She is passionate about creating a solid implementation infrastructure regarding communication and team collaboration and concentrates her efforts on building the supports needed to facilitate the knowledge transfer of our evidence-informed guidelines into practice. E-mail: Elysa.Meek@albertahealthservices.ca

Juraj **Noge** After graduating in Computer Science from the Comenius University in Bratislava, Juraj Noge built his practical experience as an IT specialist over the course of 30 years and in various software houses. For the past four years, he has been in charge of IT development at the Slovak Centre of Scientific and Technical Information (SCSTI), where his duties include implementing, managing and developing systems such as the Central Register of Theses and Dissertations (CRTD), the Anti-plagiarism System (APS), the Slovak Current Research Information System (SK CRIS) as well as various library systems. At the same time, Juraj acts as the co-ordinator and expert guarantor of activities within projects financed through the state-funded National Information System for Supporting Research and Development in Slovakia – Access to Electronic Information Resources (NISPEZ), Infrastructure for Research and Development – Data Centre for Research and Development (DCVaV/DCR&D) and National Infrastructure for Supporting Technology Transfer in Slovakia (NITT SK). Email: juraj.noge@cvtisr.sk

Miriam **Ondrišová** studied computer science at the Slovak University of Technology in Bratislava. She is a staff member of the Department of Library and Information Science of Faculty of Philosophy, Comenius University, Bratislava. As a teacher she reads lectures devoted to Bibliometrics, Database and Information Systems and Information technology. In research she works in a field of bibliometrics, information literacy and institutional repositories. miriam.ondrisova@fphil.uniba.sk

Maciej **Ostaszewski** is currently involved into sociological research on the knowledge and attitude of the academic community in Poland towards various aspects of Open Science. He has led some research projects in the sociology of science addressing problems of the institutional/geographic mobility of Polish scientists. His main research interests refer to new open models of scholarly communication, knowledge sharing and social aspects of Internet technology developments. Email: mostaszewski@opi.org.pl

Pasquale **Pagano** is a Senior Researcher at the Networked Multimedia Information Systems Laboratory of the "Istituto di Scienza e Tecnologie della Informazione A. Faedo" (ISTI) of the Italian National Research Council (CNR). He received M.Sc. in Information Systems Technologies from the Dept of Computer Science of the University of Pisa (1998), and Ph.D. degree in Information Engineering from the Dept of Information Engineering: Electronics, Information Theory, Telecommunications of the same university (2006). The aim of his research is the study and experimentation of models, methodologies and techniques for the design and development of distributed virtual research environments (VREs) which require the handling of heterogeneous resources provided by Grid and Cloud based e-Infrastructures. Has a strong background on distributed architectures. Participated to the design of the most relevant distributed systems and e-Infrastructure enabling middleware developed by ISTI-CNR. Currently the Technical Director of the Data e-Infrastructure Initiative for Fisheries Management and Conservation of Marine Living Resources (iMarine) and member of VENUS-C European project. Also involved in the GRDI2020 expert working group, serves EUBrazilOpenBio initiative as consultant. In the past, he has been involved in the D4Science-II, D4Science, Diligent, DRIVER, DRIVER II, BELIEF, BELIEF II, Scholnet, Cyclades, and ARCA EU projects. pasquale.pagano@isti.cnr.it

Gabriella **Pardelli** was born at Pisa, graduated in Arts in 1980 at the Pisa University, submitting a thesis on the History of Science. Since 1984, researcher at the National Research Council, Institute of Computational Linguistics "Antonio Zampolli" ILC, in Pisa. Head of the Library of the ILC Institute since 1990, responsible for the Archives of the ILC Institute since 2005. Her interests and activity range from studies in grey literature and terminology, with particular regard to the Computational Linguistics and its related disciplines, to the creation of documentary resources for digital libraries in the Humanities. She has participated in many national and international projects including the recent projects:- BIBLOS: Historical, Philosophical and Philological Digital Library of the Italian National Research Council, (funded by CNR); - For digital edition of manuscripts of Ferdinand de Saussure (Research Programs of Relevant National Interest, PRIN - funded by the Ministry of Education, University and Research, MIUR). Email: gabriella.pardelli@ilc.cnr.it

Sergey **Parinov** is a deputy director at the Central Economics and Mathematics Institute of the Russian Academy of Sciences (CEMI RAS), Moscow. He got his PhD (1983, CEMI RAS) in socio-economic modelling and a Doctor of Sciences degree (2001, Novosibirsk State University) in Computer Science. Last decade he was involved in open archives and institutional repositories development for Russian Academy of Sciences organizations. Currently he works on building virtual research environments for the Russian community of scientists by implementing CRIS-CERIF approaches and Open Science ideas. He has led the Socionet project (socionet.ru) over 13 years with a wide network of contacts in Russia. Now his Socionet team is designing a new public service for scientists to explicitly express their knowledge, opinions and hypotheses about scientific relationships that can exist between information objects of research data and information space. During the last years he has been one of leaders of the CRIS of Russian Academy of Sciences (RAS) project, in which more than 400 research institutes of RAS are involved. In this project he is responsible for a CRIS concept and a design of modern research management and assessment information system. From 2009 he serves on the Board of euroCRIS with responsibility for the DRIS/Best Practice Task Group. Email: sparinov@gmail.com

Leonid P. **Pavlov** graduated from Moscow Physical-Engineering Institute, Dipl. Eng. in computer systems. He is a Candidate of Sciences in informatics; and since 1976 is employed with the Scientific and Technical Information Centre of Russia (VNTIC) as Deputy Director. Main works in information systems, scientific and technical information, and grey literature. Email: leonpav@mail.ru

Petra **Pejsova** studied information science and librarianship at Charles University. She works as an information specialist in the State technical Library, Czech Republic. Actually she is leading a project Digital Library for Grey Literature – Functional model and pilot. Email: petra.pejsova@techlib.cz

Hélène **Prost** is responsible for studies at the Institute of Scientific and Technical Information (INIST-CNRS). The different studies concern the evaluation of collections, document delivery, usage analysis, grey literature and open access to information. Expertise in statistical tools and knowledge in library information science allowed her to participate in various research projects and writing of several publications. Email: helene.prost@inist.fr

Inderjeet **Sahota** is a Knowledge Management Specialist for the Guideline Utilization Resource Unit (GURU) at Alberta Health Services – Cancer Control. He received his B.Sc. in Kinesiology (2008) and M.Sc. in Biomedical Physiology and Kinesiology (2011) at Simon Fraser University. As part of GURU he supports Provincial Tumour Teams in articulating their best practices related to the diagnosis, staging, treatment and follow-up of cancer. His current work is focused on the development of evidence-informed clinical practice guidelines for site-specific cancers. E-mail: Inderjeet.sahota@albertahealthservices.ca

Manuela **Sassi** graduated in Foreign Languages and Literature at Pisa University, 110/110 cum laude. Since 1974 she has been working in Pisa at the Institute for Computational Linguistics of the National Research Council. Her interests and experiences range from linguistic to textual data processing and in providing linguistic resources on-line. She has been responsible for many national projects and has participated in numerous international projects. Email: manuela.sassi@ilc.cnr.it

Author Information

Eva **Sassolini** graduated in Computer Science, at Pisa University, is CTER (Research Collaborator) at the Institute of Computational Linguistics (ILC) of the National Research Council - Pisa. She is involved currently in several national and international projects. Research Collaborator in "TextPower" (TP) project, (new technology and approach to treatment and exploitation of texts) and before in the project "Corpus Bilingue Italiano-arabo" for linguistic tools and resources for bilingual Italian/Arabic corpora realization. Junior researcher ILC in the project LINGUISTIC MINER: linguistic Knowledge system for the Italian language; working contribution in "INTERA" (Integrated European Language Data Repository Area) project, for multilanguage terms extraction. Junior researcher ILC in the project: "Progetto Iraq: navigazione nei materiali testuali del museo virtuale di Baghdad in maniera contrastava tra le tre lingue previste dal progetto (italiano, arabo e inglese)". Junior researcher ILC in the project 8: "Diffusione della cultura e valorizzazione del patrimonio letterario della lingua italiana e della lingua araba attraverso una diffusione telematica di banche di dati letterarie". Collaboration with IMSS (Istituto e Museo di Storia della Scienza) for the realisation of web applications for the query on galileian texts. Junior researcher in the project: "Corpus Bilingue italiano-arabo": in the framework of the comprehensive "Linguistica Computazionale: ricerche monolingui e multilingui". Email: eva.sassolini@ilc.cnr.it

Dobrica **Savić** is Head of the Nuclear Information Section (NIS) at the International Atomic Energy Agency (IAEA) in Vienna, Austria, which comprises of the International Nuclear Information System (INIS), the IAEA Library, and the IT support group. Mr. Savić holds a MPhil in Library and Information Science from Loughborough University, UK, an MA in International Relations from the University of Belgrade, Serbia, as well as a Graduate Diploma in Public Administration, Concordia University, Montreal, Canada. He has extensive experience in the management and operations of web, library, information and knowledge management, as well as records management and archives services across various United Nations Agencies, including UNV, UNESCO, World Bank, ICAO, and the IAEA. His main interests are long-term sustainability of information services, democratization of scientific and technical information, and the practical application of modern information technologies.
Email: d.savic@iaea.org

Joachim **Schöpfel** is senior lecturer at the Department of Information and Library Sciences at the Charles de Gaulle University of Lille 3 and Researcher at the GERiiCO laboratory. He is interested in scientific information, academic publishing, open access, GL and eScience. He is a member of GreyNet and euroCRIS. He is also the Director of the National Digitization Centre for PhD Theses (ANRT) in Lille, France. Email: joachim.schopfel@univ-lille3.fr

Melissa **Shea-Budgell** is a specialist with the Guideline Utilization Resource Unit (GURU), a knowledge management program within CancerControl Alberta, in Alberta, Canada. Melissa supports the Genitourinary, Endocrine, Cutaneous, and Gynecologic Provincial Tumour Teams in the development and evaluation of clinical practice guidelines. Melissa holds a Master of Science degree in Nutrition and Health Promotion and has been working in the area of health sciences research since 2002. Melissa is interested in research on the development and use of evidence-based quality indicators in cancer care and the delivery of cancer-related information to patients undergoing follow-up for a malignancy.
E-mail: melissa.shea-budgell@albertahealthservices.ca

Michal **Sliacky** graduated from the Comenius University in Bratislava. In 2006 he obtained title "Master of Library and Information Science" and after that he has permanently worked in the field at different positions among various libraries. From 2009 he is employed at the Slovak Centre of Scientific and Technical Information in Bratislava, right now at Department of Electronic Information Resources (EIR). Department of Electronic Information Resource is occupied by acquisition, use and promotion of EIR. These activities are among the main outcomes of the EU project: National Information System Promoting Research and Development in Slovakia – Access to Electronic Information Resources. Email: michal.sliacky@cvtisr.sk

Plato L. **Smith** II, Head of the Digital Library Center Department at Florida State University (FSU) with rank of Associate University Librarian, my career in research libraries includes over 7 years total Association of Research Libraries experience with 5 years of increasing managerial esponsibilities at FSU. As the Digital Library Center Department Head at FSU since June 2005, I manage major aspects of library's digital collections development, institutional repository, digital assets management system, oversee scanning and digitization projects, and serve as FSU's state-wide digital initiatives subcommittee (DISC) representative, Council of State University Libraries DISC Chair, and technical expert and Co-PI on the grant funded MetaArchive project (LOCKSS), a multi-institutional digital preservation partnership funded by the Library of Congress NDIIPP program. psmithii@fsu.edu

Patrícia **Stanová** studied library and information science at Comenius University in Bratislava (Slovakia). Since 2011 she works in National Centre for Popularisation of Science and Technology in Society established by the Slovak Centre of Scientific and Technical Information in Bratislava (Slovakia). She is actively involved in organizing events and activities focused on science and technology popularization in society. She also participates as a team member at national projects: National Infrastructure for Technology Transfer in Slovakia (NITT SK) and Promotion of Science and Technology in Slovakia (PopVaT). From 2012 she is studying PhD study at Comenius University in Bratislava (Slovakia) with the theme: Information behavior of "Google" generation. Email: patricia.stanova@cvtisr.sk

Jela **Steinerová** is professor in Library and Information Science at Comenius University in Bratislava. She deals especially with human information behavior and information ecology. She teaches courses on Theory of Information Science, Information Behavior, Introduction to Information Science, Information Products, Methodology of research for doctoral students. Since 2002 she has been the head of a number of research projects, namely Interaction of man and information environment, Information use in information society and Information ecology, including an international European project of DELOS digital libraries. She was also engaged in international projects as UNESCO concept of information literacy for central Europe, The European curriculum of library and information science (EUCLID), the European Library, the collaboratory in LIS, etc. She is an author of several monographs and textbooks (information behavior, information strategies in the electronic environment), a member of editorial boards of international journals, lectures at international conferences. She participated in several study stays (Oxford, Brussels, Washington) and organized a series of international conferences (e.g. Information use in information society, 2006, Information ecology and libraries, 2011). Her main research reports and published works dealt with information behavior of students, relevance, information ecology, information literacy. She is a member of boards of the European conference on information literacy (ECIL), the international AEW workshop, Polish international ISKO. Lately she lectured at such conferences as AEW (Turku), Inforum (Prague), LIDA (Zadar), ELAG (Helsinki), CoLIS (London) etc. The report on information ecology of the academic information environment in Slovakia presents results of a qualitative study of information managers in universities, a survey of academic repositories, and experiments with concept mapping in digital repositories. At present she is engaged in the research of information behavior and information practices of PhD. students in cooperation with computer science project on cognitive traveling in the web. She is also engaged in international collaboration in information practices in workplaces (ENWI networking project) and interested in information ethics. She organizes methodological workshops for doctoral students in information studies and information systems. Email: steinerova@fphil.uniba.sk

Christiane **Stock** is the Head of the Monographs and Grey Literature service at INIST, in charge of the repositories LARA (reports), mémSIC (master's theses in information sciences) and OpenGrey. Member of the Technical Committee for the SIGLE database from 1993 to 2005, she also set up the national agency for ISRN (International Standard Report Number). She is member of the AFNOR expert group who prepared the recommended metadata scheme for French electronic theses (TEF). Email: christiane.stock@inist.fr

Author Information

Germain **St-Pierre** is a Digital Preservation Technician at the International Atomic Energy Agency (IAEA). Since 1997, he has been responsible for the collection, processing and dissemination in electronic format of the non-conventional literature (NCL) submitted to the International Nuclear Information System (INIS) by Member States. He is currently involved in two important projects, the digitization of the INIS NCL collection on microfiche and, in collaboration with the IAEA Library, the digitization of all books published by the Agency between 1957 and 1999. Email: g.st-pierre@iaea.org

Brae **Surgeoner** is a Knowledge Management Specialist for the Guideline Utilization Resource Unit (GURU) at Alberta Health Services – Cancer Control. She received her MSc in Food Safety and Risk Communication from the University of Guelph in 2006. As part of GURU she supports Provincial Tumour Teams in articulating their best practices related to the diagnosis, staging, treatment and follow-up of cancer. Her current work is focused on the development of evidence-informed clinical practice guidelines for site-specific cancers.
Email: brae.surgeoner@albertahealthservices.ca

Jaroslav **Šušol** works as a professor in library and information studies, and the dean of the Faculty of Philosophy, Comenius university in Bratislava. He received his master degree in library and information science and English at Comenius University in 1986, license spéciale in information science and documentology at Université Libre de Bruxelles in 1992, and PhD in LIS with the thesis on the topic of retrospective conversion of library catalogues in 1993. He took part in several European research projects connected with network communication environment and its various aspects, like DELOS, CECUP, and CELIP. At present he is involved in research in the area of scholarly electronic communication, digital libraries, information and publishing behaviour and information ecology. Email: susol@fphil.uniba.sk

Marián **Vaňo** is an information specialist at the Slovak Centre of Scientific and Technical Information in Bratislava (Slovakia). He has long time and extensive experience in the field of information retrieval from the scientific and research databases. At present he is responsible for the national search portal for the science and research – scientia.sk, which is one of the outputs of the EU project: National Information System Promoting R&D in Slovakia - Access to Electronic Information Resources. Email: marian.vano@cvtisr.sk

Marcus **Vaska** is a librarian with the Knowledge Resource Service (KRS), University of Calgary, Holy Cross Site, providing research and information support at an Alberta Cancer Care research facility. A firm supporter of embedded librarianship, Marcus engages himself in numerous activities, including instruction and research consultation, with research teams at Holy Cross and beyond. Marcus' current interests focus on collaborative educational techniques aimed at creating greater awareness and bringing grey literature to the forefront in the medical community. Email: mmvaska@ucalgary.ca

Laurissa **Watson** is currently a Research Assistant in Evaluation in Population and Public Health at Alberta Health Services. She graduated from the University of Calgary with a Bachelor's in Health Sciences (Honours) and recently completed an MPH from the University of Saskatchewan. Recent projects include research on HPV vaccine uptake in Alberta, an evaluation of a breast and cervical cancer screening social marketing campaign, and an evaluation of workplace health programs in Alberta. E-mail: Laurissa.Watson@albertahealthservices.ca

Danica **Zendulková** has worked in Slovak Centre of Scientific and Technical Information since 1996 in the ICT division. Until 2006 she provided the coordination of national distribution of library software developed by the UNESCO. Since 2002 she administrated the Information System of Research and Development Potential and the web site concerning science and technology. In 2005 she became Head of R&D Portal Unit. In 2007-2008 she worked as Implementation Manager of the project: Central Information Portal of Research, Development and Innovation. Since 2008 her responsibility is to administrate of Slovak state-operated CRIS system that is one part of the Portal. Danica is currently involved in EC funded projects. Primarily, following activity of national project NISPEZ is her job: Preparing a new CERIF-driven SK CRIS. During her career Danica Zendulkova has worked also as lecturer in specialised courses for librarians and has published several articles on wide issues concerning library automatization and CRIS systems. Since 2011 she has been the EuroCRIS board member with responsibility for task group CRIS-IR which aims at furthering the science and technology of the linkage between CRIS and repositories.
Email: danica.zendulkova@cvtisr.sk

Mária **Žitňanská** graduated from Faculty of Philosophy, Comenius University in Bratislava – Department of Library Science and Scientifical Informations. Since 2007 she has been working as a Deputy Director for library and information activities in the Slovak Centre of Scientific and Technical Information. She deals with development and promotion of library and information systems and services, participates on national projects for support science and research in Slovak Republic. She is a member of the Methodical Board of Slovak National Library, the Slovak Library Association Executive Council and the chair of Technical Committee 69 – Terminology. Information, and Documentation.
E-mail: maria.zitnanska@cvtisr.sk

Index to Authors

GreyNet Members less 20%

The Grey Audit : A Field Assessment in Grey Literature

Slovak Centre of Scientific and Technical Information, Bratislava, Slovakia 2-3 December 2013

Publication Order Form

FIFTEENTH INTERNATIONAL CONFERENCE ON GREY LITERATURE

Publication(s):	*No. of Copies*	x	*Amount in Euros*	Subtotal
GL15 CONFERENCE PROCEEDINGS – Printed Edition ISBN 978-90-77484-22-7 ISSN 1386-2316 *Postage and Handling* ***excluded****)		x	109.00 =	€
GL15 CONFERENCE PROCEEDINGS - CD-Rom Edition ISBN 978-90-77484-22-7 ISSN 1386-2316 *Postage and Handling* ***included***		x	109.00 =	€
GL15 Conference Proceedings – Online Edition ISBN 978-90-77484-22-7 ISSN 2211-7199 ***Password Protected Access***		x	109.00 =	€

*POSTAGE AND HANDLING PER PRINTED COPY**)

Holland		x	5.00	€
Europe		x	10.00	€
Other		x	20.00	€
			TOTAL EURO =	**€**

Customer Information

Name:	
Organisation:	
Postal Address:	
City/Code/Country:	
E-mail Address:	

Upon receipt of payment the publication(s) will be forwarded to your shipping address with an invoice marked paid.

❑ Direct transfer to TextRelease, Account No. 3135.85.342, Rabobank Amsterdam
BIC: RABONL2U **IBAN:** NL70 RABO 0313 5853 42, with reference to "GL15 Publication Order"

❑ MasterCard/Eurocard ❑ Visa Card ❑ American Express

Card No. _____ Expiration Date: _____

Print the name that appears on the credit card, here _____

Signature: _____ **CVC II code:** _____ *(Last 3 digits on signature side of card)*

Place: _____ Date: _____

NOTE: CREDIT CARD TRANSACTIONS CAN BE AUTHORIZED BY PHONE, FAX, OR POSTAL SERVICES. EMAIL IS NOT AUTHORIZED.

TextRelease
www.textrelease.com

GL15 Program and Conference Bureau
Javastraat 194-HS, 1095 CP Amsterdam, Netherlands
T/F +31-(0) 20-331.2420 Email: info@textrelease.com