https://doi.org/10.26069/greynet-2026-000.546-tgj
ISSN 1574-1796

An International Journal on Grey Literature

Summer 2014 – TGJ Volume 10, Number 2

'Research Communities and Data Sharing'

www.textrelease.com GreyNet Grey Literature Network Service www.greynet.org

The Grey Journal

An International Journal on Grey Literature

COLOPHON

CIP
The Grey Journal (TGJ): An international journal on grey literature / D.J. Farace (Journal Editor); J. Frantzen (Technical Editor) ; GreyNet, Grey Literature Network Service. - Amsterdam: TextRelease, Volume 10, Number 2, Summer 2014.
CVTISR, EBSCO, FEDLINK-Library of Congress, Inist-CNRS, ISTI-CNR, KISTI, NIS-IAEA, NTK, and NYAM are Corporate Authors and Associate Members of GreyNet International. This serial publication appears three times a year - in spring, summer, and autumn. Each issue is thematic and deals with one or more related topics in the field of grey literature. The Grey Journal appears both in print and electronic formats.
ISSN 1574-1796 (Print)
ISSN 1574-180X (E-Print/PDF)
ISSN 1574-9320 (CD-Rom)

Subscription Rates:
€240 institutional, including postage & handling

Contact Address:
Reprint Service, Back Issues, Document Delivery, Advertising, Inserts, Subscriptions:

TextRelease
Javastraat 194-HS
1095 CP Amsterdam
Netherlands
T/F +31 (0) 20 331.2420
info@textrelease.com

http://www.textrelease.com/glpublications.html

About TGJ

The Grey Journal is a flagship journal for the international grey literature community. It crosses continents, disciplines, and sectors both public and private.
The Grey Journal not only deals with the topic of grey literature but is itself a document type classified as grey literature. It is akin to other grey serial publications, such as conference proceedings, reports, working papers, etc.

The Grey Journal is geared to Colleges and Schools of Library and Information Studies, as well as, information professionals, who produce, publish, process, manage, disseminate, and use grey literature e.g. researchers, editors, librarians, documentalists, archivists, journalists, intermediaries, etc.

About GreyNet

The Grey Literature Network Services was established in order to facilitate dialog, research, and communication between persons and organizations in the field of grey literature. GreyNet further seeks to identify and distribute information on and about grey literature in networked environments. Its main activities include the International Conference Series on Grey Literature, the creation and maintenance of web-based resources, a moderated Listserv, and The Grey Journal. GreyNet is also engaged in the development of distance learning courses for graduate and post-graduate students, as well as workshops and seminars for practitioners.

Full-Text License Agreement

In 2004, TextRelease entered into an electronic licensing relationship with EBSCO Publishing, the world's most prolific aggregator of full text journals, magazines and other sources. The full text of articles in The Grey Journal (TGJ) can be found in *Library, Information Science & Technology Abstracts* (LISTA) full-text database.

Contents
'Research Communities and Data Sharing'

https://doi.org/10.26069/greynet-2026-000.546-tgj

Pisa Declaration on Policy Development for Grey Literature Resources

A wealth of knowledge and information is produced by organizations, governments and industry, covering a wide range of subject areas and professional fields, not controlled by commercial publishing. These publications, data and other materials known as grey literature, are an essential resource in scholarly communication, research, and policy making for business, industry, professional practice, and civil society. Grey literature is recognized as a key source of evidence, argument, innovation, and understanding in many disciplines including science, engineering, health, social sciences, education, the arts and humanities.

Grey literature document types in print or electronic formats include among others: research and technical reports, briefings and reviews, evaluations, working papers, conference papers, theses, and multimedia content, representing an important and valuable part of research and information.

In order to realize the benefits of research and information for scholarship, government, civil society, education and the economy, We, the signatories to this declaration, call for increased recognition of grey literature's role and value by governments, academics and all stakeholders, particularly its importance for open access to research, open science, innovation, evidence-based policy, and knowledge transfer.

To achieve the full benefits of grey literature for local, national and global communities we call for and encourage the following:

Organizational

1. Greater commitment to open access by governments and organizations.
2. Greater cooperation and coordination among organizations engaged in the production, use, collection and management of grey literature.
3. The use of persistent identifiers and open metadata standards for grey literature.

Research/Educational

4. New forms of recognition and reward for quality grey literature materials by governments, universities and other institutions.
5. Improved standards in the production and bibliographic control of grey literature.
6. Development and implementation of interoperable standards in the management of grey literature.
7. Development of good practice guides for the production, dissemination, and evaluation of grey literature.

Legal

8. Changes to legal deposit and copyright law that enhance the capacities of libraries, collecting services and educational institutions and programs to collect and provide access to grey literature, particularly non-commercial public interest materials.
9. Addressing legal obstacles to the dissemination of grey literature.
10. Further strides in licensing grey content for both commercial and non-commercial purposes.

Financial/Sustainable

11. Identifying available funding for research involving grey literature.
12. Increased support for collection development and long term preservation of grey literature.
13. Increased investment in infrastructure and new technologies for accessing and using print and digital grey literature.

Technical

14. Strategies to tackle link rot and enhance the stability and accessibility of online content.
15. Systems for linking data and other non-textual content to their grey literature publications together with interoperability standards for sharing grey literature.

GreyNet Resource Policy Committee
http://greyguide.isti.cnr.it

https://doi.org/10.26069/greynet-2026-000.546-tgj

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

Massimiliano Assante, Leonardo Candela, Donatella Castelli,
Francesco Mangiacrapa, Pasquale Pagano (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.

* First published in the GL15 Conference Proceedings, March 2014.

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.

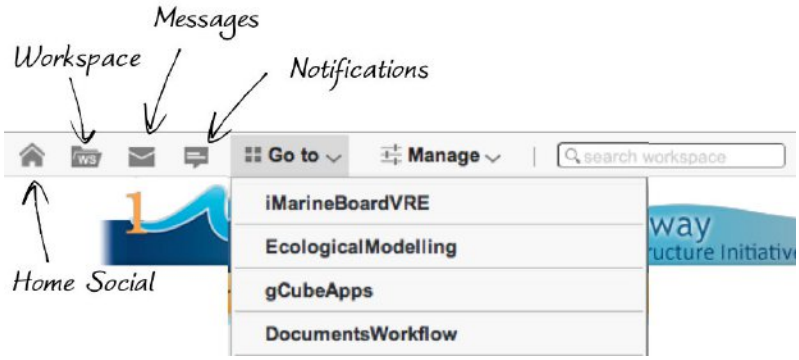


Figure 1. Dockbar

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).

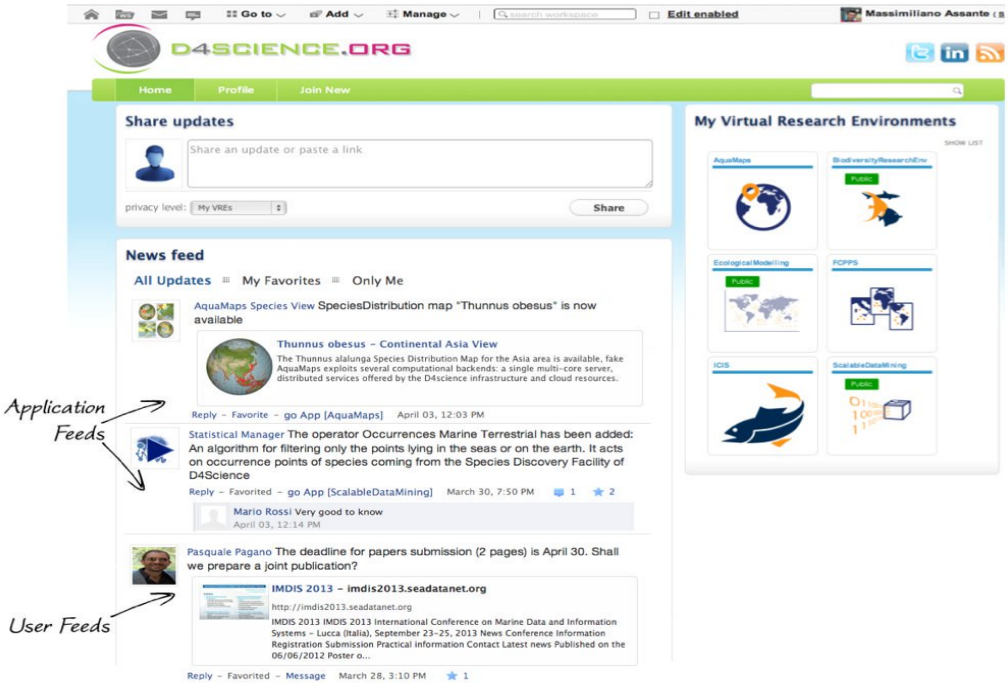


Figure 2. Home Social Facilities

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.

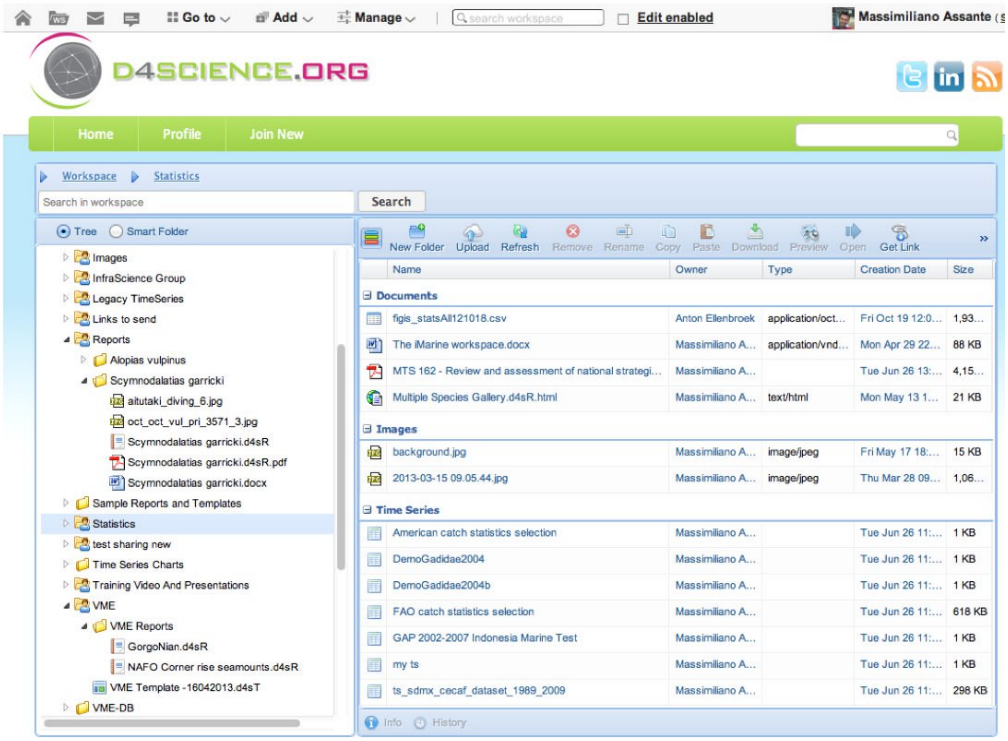


Figure 3. Workspace

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.

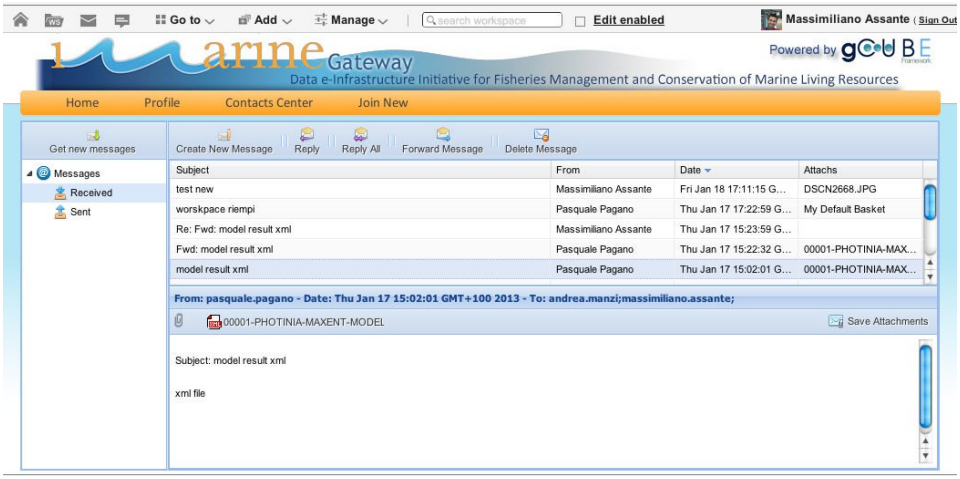


Figure 4. Messages

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.

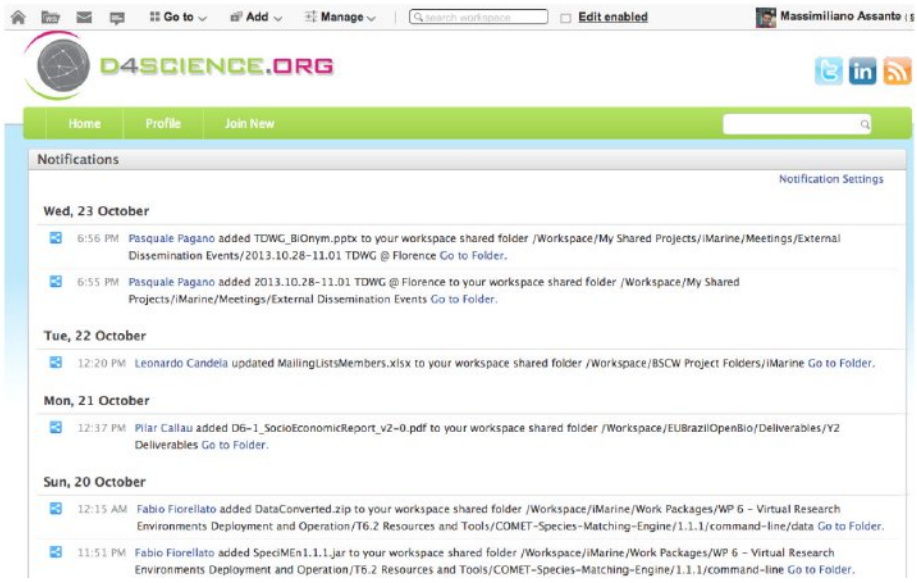


Figure 5. Notifications

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.

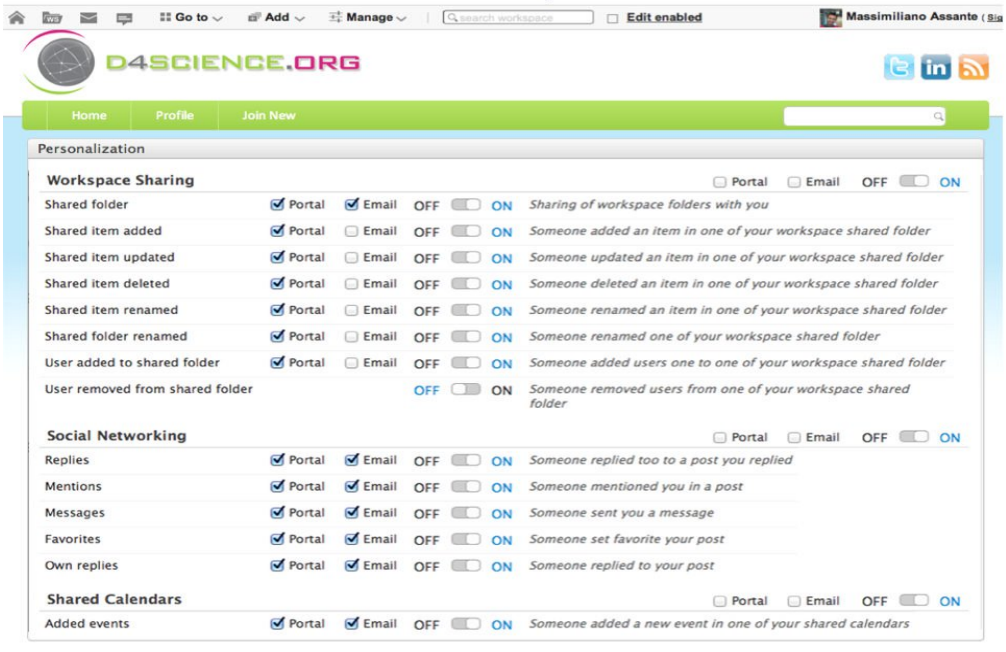


Figure 6. Personalisations

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.

https://doi.org/10.26069/greynet-2026-000.546-tgj

Korea Institute of Science and Technology Information (KISTI)

English version - http://en.kisti.re.kr/

*** Vision**

World-class information research institute creating values for customers

*** Main functions**

*** Management and service of Korean R&D reports**

KISTI exclusively manages, preserves, and serves Korean R&D reports for citizens and government officials. It provides Korean R&D reports and their information with National science & Technology Information Service (NTIS) and National Discovery for Science Leaders (NDSL).

***Contact information**

KISTI email address: hcpark@kisti.re.kr

Headquarters: Tel : +82-42-869-1004, 1234 Fax: +82-42-869-0969

https://doi.org/10.26069/greynet-2026-000.546-tgj

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

Maciej Ostaszewski (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.

* First published in the GL15 Conference Proceedings, March 2014.

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 platform[i]
- **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

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

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

Jela Steinerová and Andrea Hrčková (Slovak Republic)

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.

* First published in the GL15 Conference Proceedings, March 2014.

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).

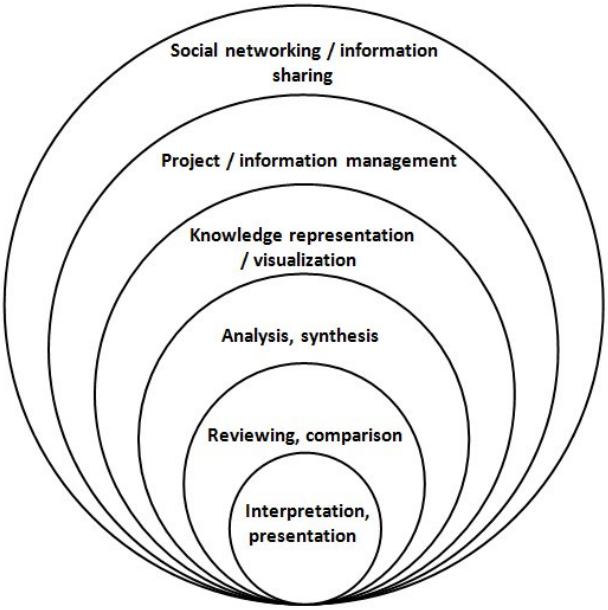


Fig. 1 The model of information support of information interactions

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

https://doi.org/10.26069/greynet-2026-000.546-tgj

https://doi.org/10.26069/greynet-2026-000.546-tgj

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

Sergey Parinov, Mikhail Kogalovsky, Victor Lyapunov, (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.

* First published in the GL15 Conference Proceedings, March 2014.

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).

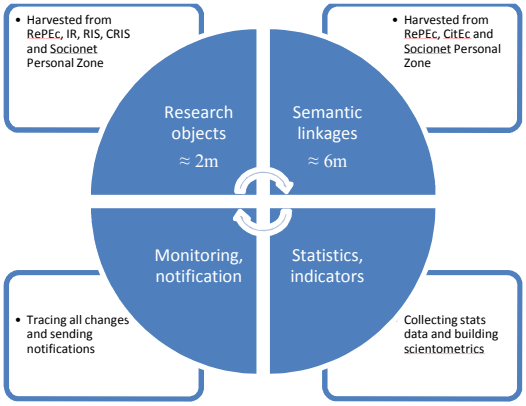


Figure 1. Socionet main subsystems

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).

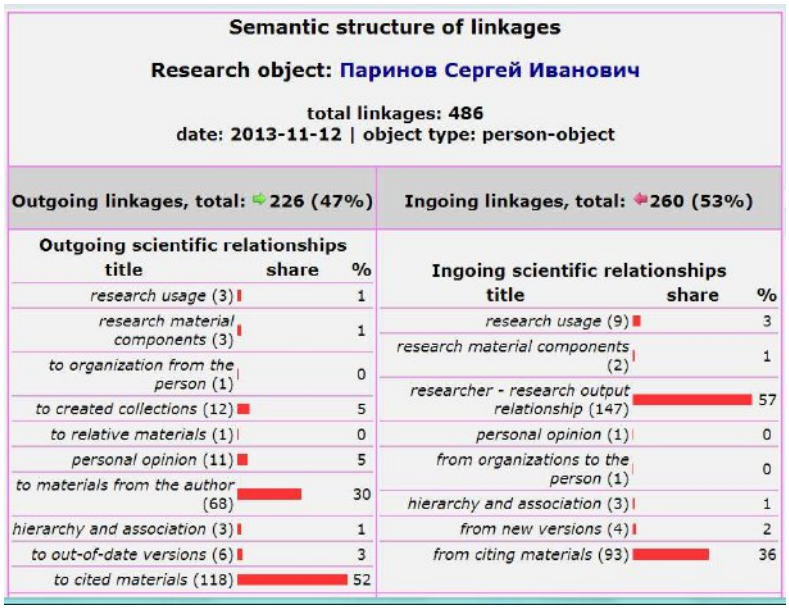


Figure 2. An example of scientific relationship classes distribution for a person

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.

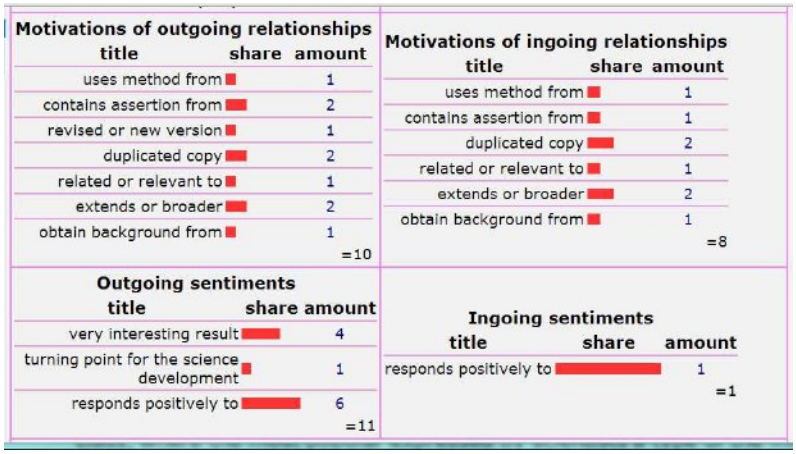


Figure 3. An example of motivation distributions of a person

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

https://doi.org/10.26069/greynet-2026-000.546-tgj

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

Danica Zendulkova and Juraj Noge (Slovak Republic)

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

* First published in the GL15 Conference Proceedings, March 2014.

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.

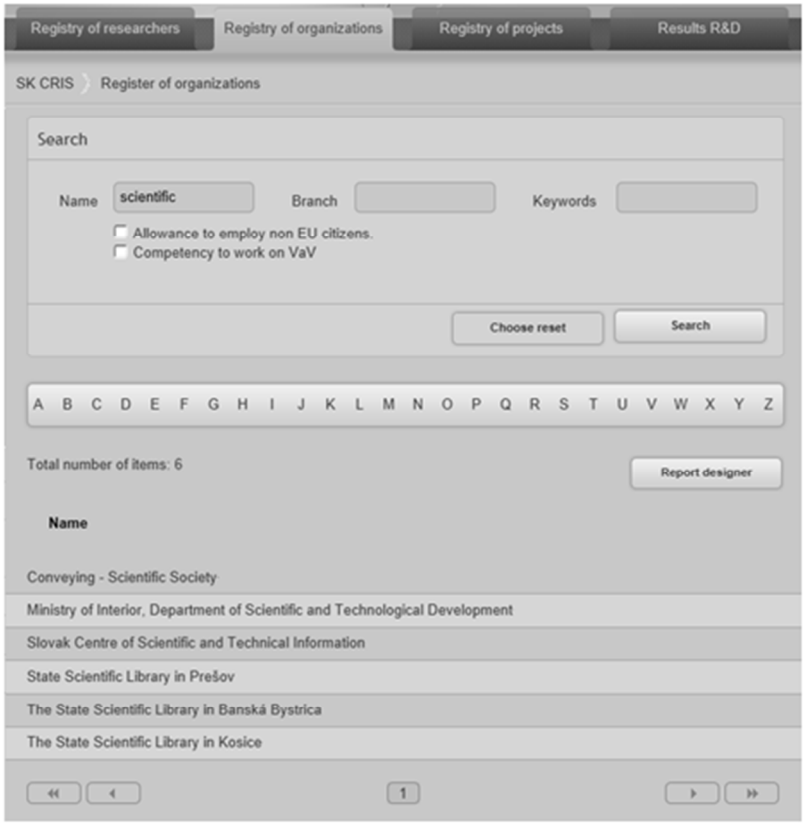


Figure 2 Registry of Organizations, user interface

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 .

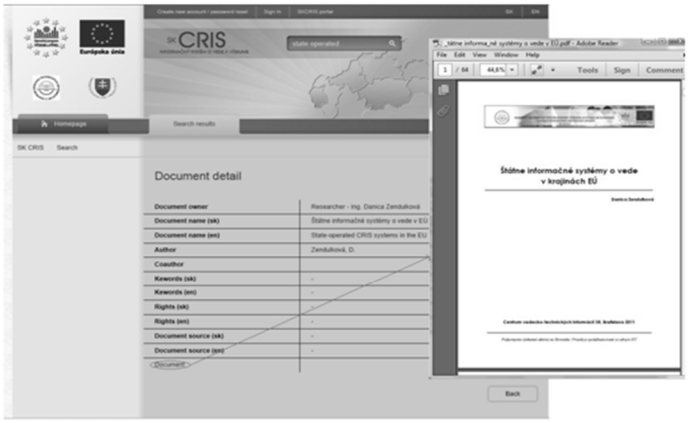


Figure 3 R&D Results: Document metadata and fulltext

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 information 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	Bílský 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.

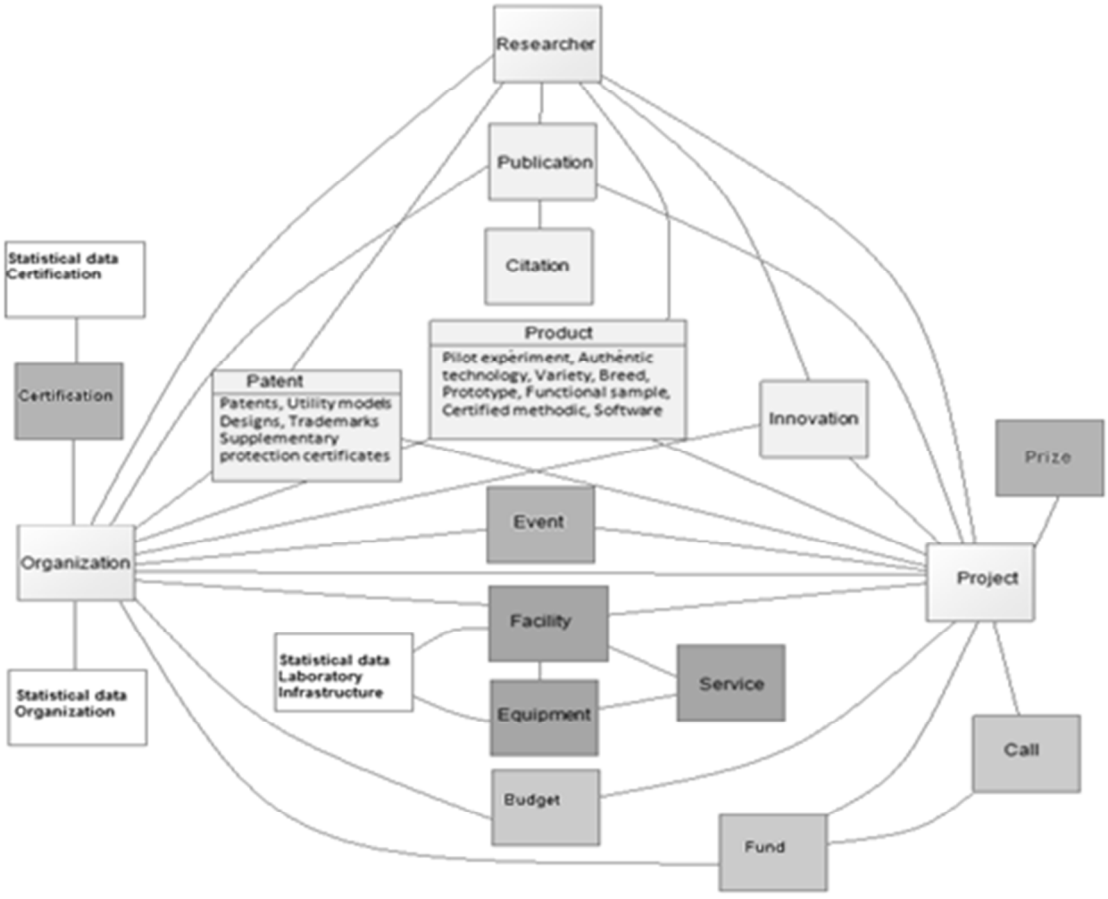


Figure 5 SKCRIS data model

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.

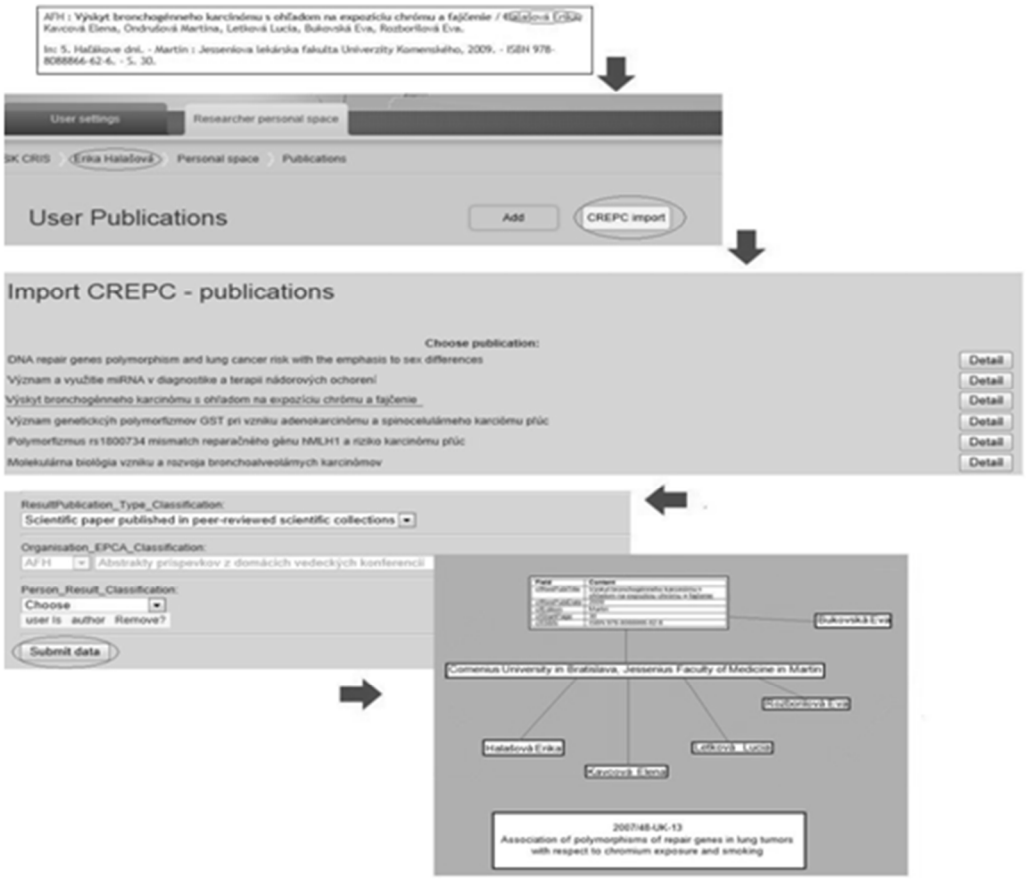


Figure 6 Connection between CRPA and SK CRIS

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 >

https://doi.org/10.26069/greynet-2026-000.546-tgj

Auditing Grey in a CRIS Environment*

Keith G. Jeffery (United Kingdom) and Anne Asserson (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.

* First published in the GL15 Conference Proceedings, March 2014.

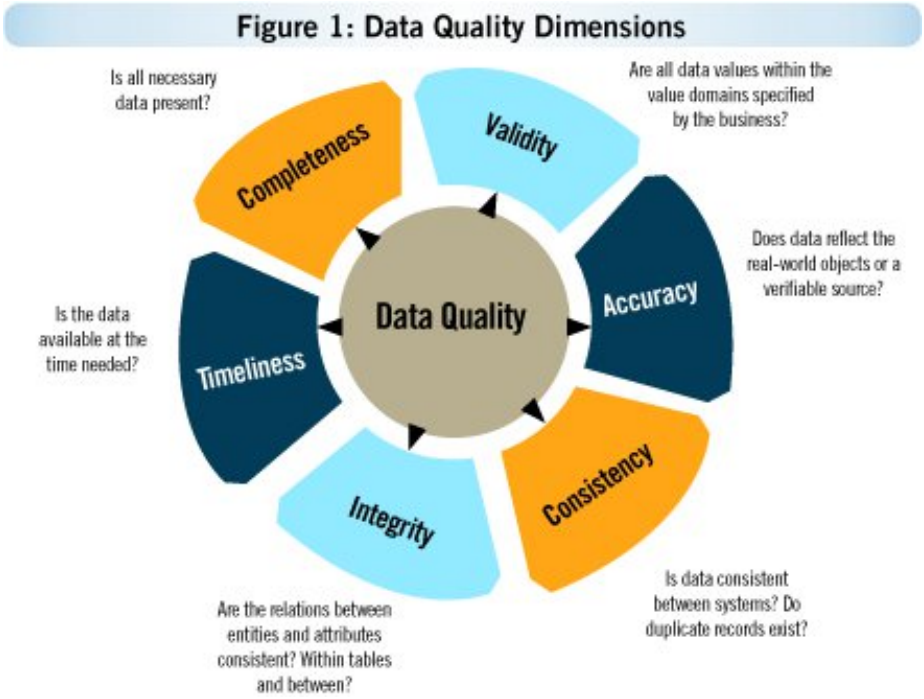
2. Reliable information

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 1.



(With acknowledgements to FINETIK)

Figure 1: Dimensions of Data Quality

First and foremost data quality (using data in the wide sense of any kind of information presented using any medium[i]) 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.

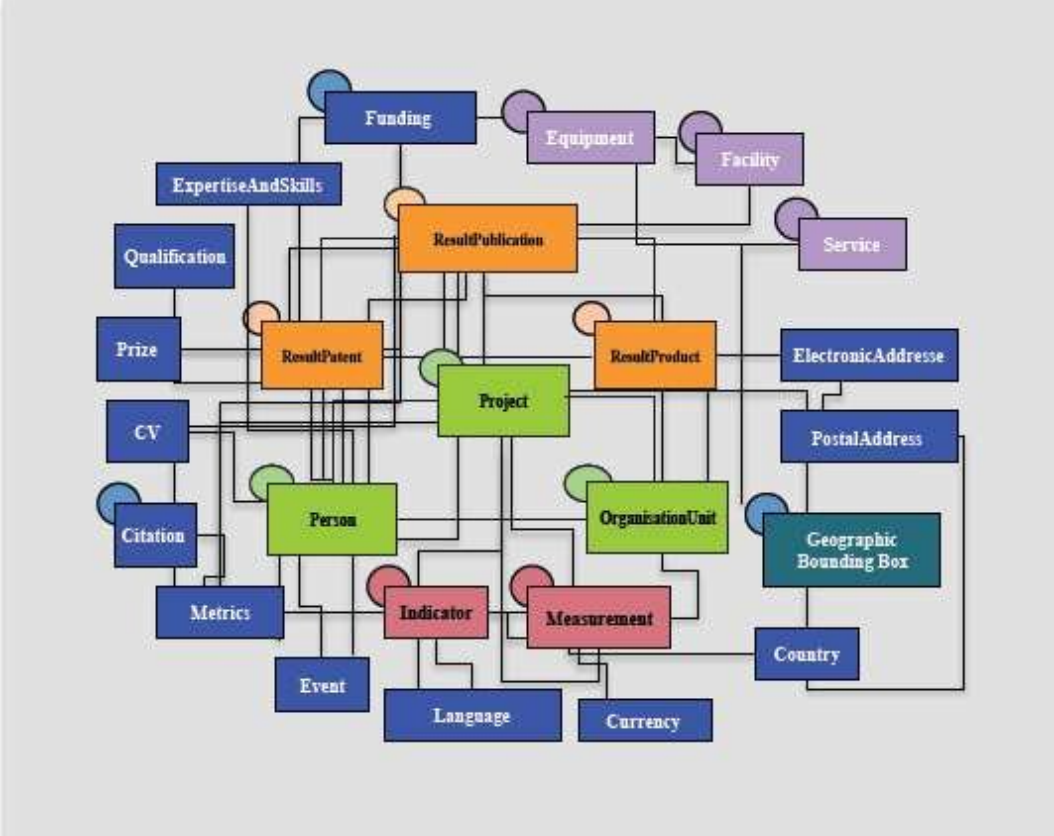


Figure 2: The CERIF Datamodel

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[ii]. 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 3): 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 3: 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)

[i] Formally information is structured data in context
[ii] http://www.w3.org/2013/05/lcsv-charter.html

https://doi.org/10.26069/greynet-2026-000.546-tgj

Sixteenth International Conference on Grey Literature

Grey Literature Lobby: Engines and Requesters for Change

The Library of Congress, Washington DC, USA

Federal Library Information Network, December 8-9, 2014

Conference Program – Day 1

OPENING SESSION 9:00-10:30

Welcome Address *Blane Dessy, Library of Congress, United States*

Opening Address *Mary Alice Baish, U.S. Government Printing Office, United States* ***(to be confirmed)***

Keynote Address *Amanda Lawrence, Institute for Social Research, Swinburne University of Technology, Australia*

Day Moderator, Amanda Wilson, U.S. Department of Transportation, United States

SESSION 1 – PUBLIC AWARENESS OF GREY LITERATURE 11:00-12:30

The Intersect of Cultural Studies with other focus areas in the HDIAC Database
June Crowe and Crystal Sherline, Information International Associates Inc., United States

Mapping Italian Grey Communities: What Is There Beyond The Academy?
Silvia Giannini and Stefania Biagioni, Istituto di Scienza e Tecnologie dell'Informazione;
Sara Goggi and Gabriella Pardelli, Istituto di Linguistica Computazionale; CNR, Italy

Shades of Grey: Integrating Metadata for Discovery in a Mixed-Content Single-Subject Library
Genevieve Podleski, Federal Reserve Bank of St. Louis, United States

How Grey Literature Informs Policy and Decision Making: The Necessity to Understand the Processes
Bertrum H. MacDonald, James D. Ross, Suzuette S. Soomai, and Peter G. Wells, Dalhousie University, Canada

SESSION 2 – PUBLISHING AND LICENSING GREY LITERATURE 1:30-3:00

Is the Licensing of Grey Literature Using the Full Palette of "Contractual" Colors?
Tomas A. Lipinski, School of Library and Information Science, Kent State University;
Andrea J. Copeland, School of Informatics and Computing Indiana University, United States

Publishing Geodesy, Topography and Cartography Research via Invenio
Jiří Drozda and Veronika Synková, Research Institute of Geodesy, Topography and Cartography;
Petra Pejšová, National Library of Technology, Czech Republic

Free Licences and Creative Commons: A Powerful Tool for Open Access Publishing in Grey Literature
Petra Pejšová, National Library of Technology, Czech Republic
Marcus Vaska, Knowledge Resource Service, University of Calgary, Canada

Degrees of Openness: Grey Literature in Institutional Repositories
Joachim Schöpfel, Charles de Gaulle University Lille 3;
Hélène Prost, CNRS, Associate member GERiiCO, France

INTRODUCTION TO CONFERENCE POSTERS 3:30-4:30

Group Tour ***(to be announced)***

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

https://doi.org/10.26069/greynet-2026-000.546-tgj

Sixteenth International Conference on Grey Literature

Grey Literature Lobby: Engines and Requesters for Change

The Library of Congress, Washington DC, USA

Federal Library Information Network, December 8-9, 2014

Conference Program – Day 2

POSTER SESSION AND SPONSOR SHOWCASE 9:00-11:00

See **Call for Posters**

Day Moderator, Michele Masias, U.S. Department of Justice, United States

SESSION 3 – OPEN ACCESS TO RESEARCH DATA 11:00-12:30

Data: Is it Grey, Maligned or Malignant?
Julia Gelfand and Daniel Tsang, University of California, Irvine, United States

Big data availability and full exploitation: A 'wide screen' overview on potential opportunities and current criticalities
Luisa De Biagi, Massimiliano Saccone, Roberto Puccinelli, National Research Council, Central Library, Italy

Introduction to a "data flagging" system for utilizing Internet information containing a tremendous amount of numerical data: The case of Fukushima Archive at the JAEA Library
Kiyoshi Ikeda, Mayuki Gonda, Shun Nagaya, Misa Hayakawa, Yukinobu Mineo, Katsuhiko Kuni, Minoru Yonezawa, Keizo Itabashi, Japan Atomic Energy Agency, Japan

Moving towards Open Access to research data in health sciences: a changing landscape for Grey Literature
Paola De Castro, Istituto Superiore di Sanità, Italy

GREYFORUM –MANAGING CHANGE IN GREY LITERATURE 1:30-3:00

Driving policy development for grey literature – The Pisa Declaration
Herbert Gruttemeier, Inist-CNRS, National Centre of Scientific Research, France

Federating European Grey Literature through OpenGrey
Christiane Stock, Inist-CNRS, National Centre of Scientific Research, France

GreyGuide, GreyNet's web access portal and lobby for Grey Literature
Dominic Farace and Jerry Frantzen, GreyNet International, Netherlands
Stefania Biagioni and Carlo Carlesi, ISTI-CNR, National Research Council, Italy
Christiane Stock, Inist-CNRS, National Centre of Scientific Research, France

Publishing and Communication: The GreyNet Community Management Committee
Joachim Schöpfel, Charles de Gaulle University Lille 3, France

Grey Lit Tutorials
Rachel Smiley, Sarah Davis, and Debbie Rabina;
Pratt Institute, School of Information and Library Science, United States

CLOSING SESSION 3:30-4:15

Group Tour ***(to be announced)***

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

https://doi.org/10.26069/greynet-2026-000.546-tgj

Sixteenth International Conference on Grey Literature

Grey Literature Lobby: Engines and Requesters for Change

The Library of Congress, Washington DC, USA

Federal Library Information Network, December 8-9, 2014

Call for Posters

● Think Tanks ● Advocacy Groups ● Research Teams ● NGOs ● IGOs ● Foundations ● Coalitions ● Lobbyists ● Special Interest Groups ● Task forces ● Associations, Etc.

Guidelines for Abstracts

Title of Poster:	**Topic/Theme:**
Author Name(s):	**Telephone:**
Organization(s):	**Fax:**
Address/P.O. Box:	**Email:**
Postal Code - City - Country:	**URL:**

Participants who seek to present a poster during GL16 are invited to submit an English abstract between 300-350 words. The abstract should describe the project and/or related information products and services. The abstract should likewise include a title, name(s) of the presenter(s) and their full address information. Abstracts are an important source of information available prior to the conference that is accessible to both delegates and conference participants.

GL16 Conference Themes
☐ **Public Awareness of Grey Literature**
☐ **Publishing and Licensing Grey Literature**
☐ **Open Access to Research Data**
☐ **Managing Change in Grey Literature**
☐ **Other Poster Topics/Themes:**

Due Date and Format for Submission

Abstracts in MS Word should be emailed to conference@textrelease.com on or before **October 15th 2014**. The author will receive verification upon receipt. This will be accompanied by further guidelines for submission of PowerPoint slides, Conference Registration, etc. ***GL16 will accommodate up to 20 posters, timely registration is a guarantee of your placement on the program.***

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

https://doi.org/10.26069/greynet-2026-000.546-tgj

Sixteenth International Conference on Grey Literature

Grey Literature Lobby: Engines and Requesters for Change

The Library of Congress, Washington DC, USA

Federal Library Information Network, December 8-9, 2014

Conference Registration

● **Think Tanks** ● **Advocacy Groups** ● **Research Teams** ● **NGOs** ● **IGOs** ● **Foundations** ● **Coalitions** ● **Lobbyists** ● **Special Interest Groups** ● **Task forces** ● **Associations, Etc.**

The Conference Registration Fee includes attendance during the Sessions, a copy of the GL16 Program Book, the full-text of the Conference Papers and PowerPoint Slides, as well as the conference pouch and participant badge. Also included are the lunches and refreshments during the breaks, and the post-conference tour.

Payment after October 1st 2014 add a $75 surcharge

☐ **$595** Participant Fee ☐ **$495** Author Fee ☐ **$275** Day Pass ☐ **$100** Student Fee

I work for a **U.S. Government Agency** and am entitled to a 20% reduction on the fee ☐

I am a *GreyNet* Member and am entitled to a 20% reduction on the conference fee ☐

Registrant's Name: _____

Organization: _____

Address/P.O. Box: _____

Postal Code/City/Country: _____

Tel/Fax/Email: _____

Check one of the boxes below for your method of payment:

❑ Direct bank transfer to TextRelease, Account 3135.85.342 - Rabobank, Amsterdam, Netherlands
BIC: RABONL2U **IBAN:** NL70 RABO 0313 5853 42, with reference to GL16 and Registrant's Name

❑ MasterCard ❑ Visa Card ❑ American Express

Card No. _____ Expiration Date: _____

If the name on the credit card is not that of the registrant, print the name that appears on the card, below

_____

Signature _____ CVC II code _____ (*Last 3 digits on signature side of card)*

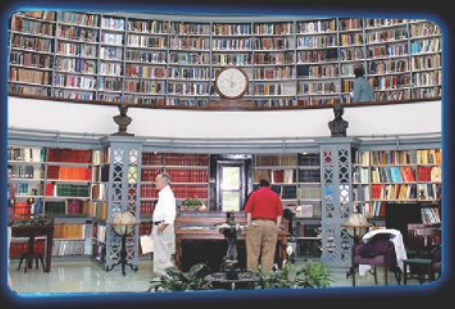
Place _____ Date _____

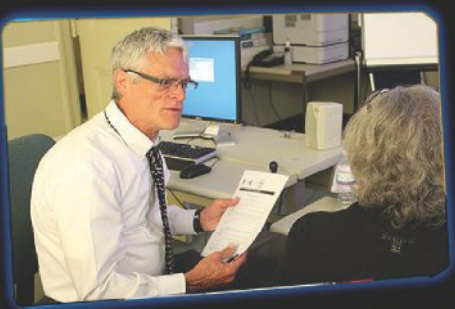
Note: Credit Card transactions can be authorized by Phone, Fax, or Postal services. Email is not authorized.

GL16 Program and Conference Bureau
TextRelease

Javastraat 194-HS, 1095 CP **A**msterdam, **T**he **N**etherlands
www.textrelease.com · **conference@textrelease.com**
Tel/Fax +31-20-331.2420

https://doi.org/10.26069/greynet-2026-000.546-tgj





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

Assante, Massimiliano 65
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

Asserson, Anne 103
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. 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. 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 cordis.lu/cerif, coordinated by the DGXIII-D4. euroCRIS is now the custodian of the CERIF model www.eurocris.org. Asserson is a member of the euroCRIS board with the responsibility Member Strategy and External Relations. anne.asserson@fa.uib.no

Candela, Leonardo 65
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

Castelli, Donatella 65
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

Hrčková, Andrea 79
Andrea has been an internal PhD. student at the Department of Library and Information Science, Comenius University since 2010. She is writing dissertation thesis „Discussion Groups as an Information Resource“ and co-created two university textbooks. She also 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. hrckova.andrea@gmail.com

Jeffery, Keith G. 103
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. Advised government on security and green computing. He chaired the EC Expert Groups on GRIDs and on cloud computing. keith.jeffery@keithgjefferyconsultants.co.uk

Kogalovsky, Mikhail 87
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 of Moscow State University during 25 years. kogalov@gmail.com

Lyapunov, Victor 87
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. Email: vic@ieie.nsc.ru

Noge, Juraj 95
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 coordinator 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 R&D – 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

Ostaszewski, Maciej 73
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

Pagano, Pasquale 65
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.
Email: pasquale.pagano@isti.cnr.it

Parinov, Sergey 87
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

Steinerová, Jela 79
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

Zendulkova, Danica 95
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

https://doi.org/10.26069/greynet-2026-000.546-tgj
TGJ Volume 10, Number 2 2014

An International Journal on Grey Literature

'RESEARCH COMMUNITIES AND DATA SHARING'

SUMMER 2014 – TGJ VOLUME 10, NUMBER 2

https://doi.org/10.26069/greynet-2026-000.546-tgj

Notes for Contributors

Non-Exclusive Rights Agreement

- I/We (the Author/s) hereby provide TextRelease (the Publisher) non-exclusive rights in print, digital, and electronic formats of the manuscript. In so doing,
- I/We allow TextRelease to act on my/our behalf to publish and distribute said work in whole or part provided all republications bear notice of its initial publication.
- I/We hereby state that this manuscript, including any tables, diagrams, or photographs does not infringe existing copyright agreements; and, thus indemnifies TextRelease against any such breach.
- I/We confer these rights without monetary compensation and with the understanding that TextRelease acts on behalf of the author/s.

Submission Requirements

Manuscripts should not exceed 15 double-spaced typed pages. The size of the page can be either A-4 or 8½x11 inches. Allow 4cm or 1½ inch from the top of each page. Provide the title, author(s) and affiliation(s) followed by your abstract, suggested keywords, and a brief biographical note.
A printout or PDF of the full text of your manuscript should be forwarded to the office of TextRelease. A corresponding MS Word file should either accompany the printed copy or be sent as an attachment by email. Both text and graphics are required in black and white.

REFERENCE GUIDELINES

General

i. All manuscripts should contain references
ii. Standardization should be maintained among the references provided
iii. The more complete and accurate a reference, the more guarantee of an article's content and subsequent review.

Specific

iv. Endnotes are preferred and should be numbered
v. Hyperlinks need the accompanying name of resource and date; a simple URL is not acceptable
vi. If the citation is to a corporate author, the acronym takes precedence
vii. If the document type is known, it should be stated at the close of a citation.
viii. If a citation is revised and refers to an edited and/or abridged work, the original source should also be mentioned.

Examples

Youngen, G.W. (1998), Citation patterns to traditional and electronic preprints in the published literature. - In: College & Research Libraries, 59 (5) Sep 1998, pp. 448-456. - ISSN 0010-0870

Crowe, J., G. Hodge, and D. Redmond (2010), Grey Literature Repositories: Tools for NGOs involved in public health activities in developing countries. – In: Grey Literature in Library and Information Studies, Chapter 13, pp. 199-214. – ISBN 978-3-598-11793-0

DCMI, Dublin Core Metadata Initiative Home Page http://purl.oclc.org/metadata/dublin_core/

Review Process

The Journal Editor first reviews each manuscript submitted. If the content is suited for publication and the submission requirements and guidelines complete, then the manuscript is sent to one or more Associate Editors for further review and comment. If the manuscript was previously published and there is no copyright infringement, then the Journal Editor could direct the manuscript straight away to the Technical Editor.

Journal Publication and Article Deposit

Once the journal article has completed the review process, it is scheduled for publication in The Grey Journal. If the Author indicated on the signed Rights Agreement that a preprint of the article be made available in GreyNet's Archive, then browsing and document delivery are immediately provided. Otherwise, this functionality is only available after the article's formal publication in the journal.