LIBER Case Study:

Factors for Enabling Sharing and Reuse of Research Data – Library and Archive Services at the University of Vienna

AUTHORS: Paolo Budroni, Vienna University Library, Paolo.Budroni@univie.ac.at; Barbara Sanchez Solis, Vienna University Library, Barbara.Sanchez.Solis@univie.ac.at

KEYWORDS: generic, international, collaborative

1 What was the starting point?

In November 2013, the NOAD (OpenAIRE National Open Access Desk) Austria initiated a study called "Factors for Enabling Sharing and Reuse of Research Data". The reason being is that throughout the projects periods (OpenAIRE and OpenAIREplus), it became evident that most academics, researchers and even repository managers are not aware of the new challenge they face with regard to the perpetual use of data and open access, both for already published articles and, even more so, for research data. This study should help to deal with these new tasks and find solutions, not only for researchers but also for repository managers and institutions involved in this topic.

The paper focuses on the many agents, strategies and techniques who have to work together to keep digital information accessible and usable over the long term and sheds light on three important actors in the whole process: data creators, repositories and downstream users. In order for research data to be shared, they must be properly curated and archived. Data creators are fundamentally involved in the processes of making data interoperable on a technical, semantic and legal level. The contents of a repository are highly heterogeneous: there are different content types and formats, licences and terms of use. Repositories thus combine people, workflows and technologies which together facilitate making data usable and reusable. Downstream users expect different levels of access and as few restrictions as possible. They want maximum freedom to use the latest tools and services in order to make best use of the information available there. In order to enable a smooth process, it is necessary to identify the different roles and responsibilities of these agents and include the views of other stakeholders. Collaborative science is, after all, based on a collaborative data infrastructure. The paper also includes findings from personal interviews conducted with professionals from selected institutes at Austrian universities with the purpose of understanding the researchers' current field of research and context as well as gaining a closer look into the current practices of research processes and research data management in higher education.

2 What kind of research data is targeted?

The target is to consider all kinds of research data without disciplinary focus in order to emphasise their multi-faceted character. They may be classified into:

- Observational: data captured in real time, usually unique and irreplaceable. For example, sensor data, survey data, sample data, neurological images
- Experimental: data from lab equipment, often reproducible, but can be expensive. For example, gene sequences, chromatograms, toroid magnetic field data

June 2014

- Simulation: data generated from test models where model and metadata may be more important than output data from the model. For example, economic or climate models
- Derived or compiled: resulting from processing or combining "raw" data, often reproducible but expensive. For example, text and data mining, compiled database, 3D models
- Reference or canonical: a (static or organic) conglomeration or collection of smaller (peer reviewed) datasets, most probably published and curated. For example, gene sequence databanks, chemical structures, or spatial data portals
- Generated data (from the perspective of data providers) might further be classified into text-based, numerical or multimedia.

Examples of the formats in which data are generated, many of these proprietary: plain text files, MS Word, Portable Document Format (PDF), Rich Text Format (RTF), CSV, Hyper-Text Markup Language (HTML), Extensible Markup Language (XML), SQL database, MS Excel, SPSS, Stata, SAS, Office Open XML (OOXML) and OpenDocument Format (ODF), JPEG, TIFF, PNG, GIF, SVG, Dicom, MPG, DivX, Quicktime, Flash Video, Bitmap, WAV, AIFF, FLAC, Geospatial vector and raster data, ESRI, CAD, GIS data.

Also included are data for research-based teaching issues ("Learning Objects").

Data is collected, provided or generated by students and researchers/scientists.

Quality measures in place: Here the understanding of the processes and their interactions (see above, "Description of the study") is important:

(a) Quality of data from the content provider perspective (efficient RDM, considering format, documentation and metadata, legal and ethical implications, archiving)

(b) Quality of data from the repository perspective (monitoring and validation, generating persistent identifiers, metadata, controlling metadata, version control, processing preservation actions, providing tools for representation and visualisation)

3 What is the organisational framework?

Over the years, Vienna University Library has acquired expert knowledge and experience in terms of technical innovation of academic library services. The library has built an institutional repository for the University of Vienna with the aim of harvesting and preserving e-publications and documents with web applications. The infrastructure has been devised so that interoperability with other repositories is ensured. Vienna University Library offers the University of Vienna's academic community digital asset management, self-archiving and collaborative archiving solutions with a view to long-term data preservation. Library services include training librarians, academic and administrative staff in managing, retrieving and distributing electronic materials from institutional and distributed repositories. With Phaidra (Permanent Hosting, Archiving and Indexing of Digital Resources and Assets), Vienna University Library has not only developed a repository that interoperates with the University of Vienna CRIS system, but one that is the core of the University's institutional repository. This digital asset management system has also been developed with a particular focus on open access policies. There are initiatives to develop data management plans, metadata schemes for research data and involvement of learning data within the Phaidra system. Phaidra international is not only implemented at several local institutions, but is also in use internationally at universities in Serbia, Montenegro and Italy, for example. It also is offered as a content aggregation system to Austrian scientists involved in EU-funded projects. The University of Vienna is a member of COAR (Confederation of Open Access Repositories).

Policies

The policy underlying Phaidra covers:

- Policy of Open and Free Access: Developed based on open source software, but access to the system is available without restrictions to all members of the University of Vienna – teachers and students. The active use of Phaidra, that is, the storing and linking of objects – without bureaucratic obstacles (red tape) – is open to users with a mailbox account, typically employees of the University of Vienna, as well as externals who obtain permission, and students with a u:net account. Searching and viewing the contents – if desired – is possible worldwide without logging in. All members are permitted to invite "guests" (guest accounts).
- Policy to Support University-wide Open Access: In the context of the Open Access Policy, not only full texts in PDF format should be made accessible, but also "multi-resource content type" objects should be created. Therefore, it is not enough to only create PDFs, since research also requires models for multimedia objects, primary data, mathematical formulas, and so on.
- The second addition to the Open Access requirement of the Budapest Declaration: Free access to the very instrument of publication itself. The content provider in research (owner of the digital objects) should be able to publish digital objects and assign usage licenses to them at any time and without bureaucratic restrictions.
- Policy of Persistent Citation of Data Stored in Phaidra: "Once in Phaidra – always in Phaidra"
- Policy for a Sophisticated Access Concept: Members of the University of Vienna can create digital objects that are accessible worldwide and can lock these objects at any time without much effort (i.e., independently and without assistance of other persons or administrators), and they can make these available to a limited circle of users. The content provider should be able to control access.
- Policy for a Sophisticated Access Concept: Members of the University of Vienna can create digital objects that are accessible worldwide and can lock these objects at any time without much effort (i.e., independently and without assistance of other persons or administrators), and they can make these available to a limited circle of users. The content provider should be able to control access.
- Policy on the Clarification of Legal Issues: Clarification of legal issues and the availability of licensing models and conditions of use provide security and foster trust. Phaidra makes it possible to secure the legal status of the stored objects.
- Policy for the extendable and contemporary content model: Objects can be stored in Phaidra with different content types and then be put into relation with one another.
- Policy for the Permanent and Reliable Storing of Data: The support of the formats recommended in Phaidra is carried out at the Central Informatics Service at the University of Vienna, where the flexible open source software Fedora is supported. All metadata is converted into XML format, which supports the permanent storing functions of Phaidra and promotes interoperability with other systems.

Structured Metadata: The basis of the metadata sets includes Dublin Core and Learning Object Metadata (LOM). The system automatically generates an XML-based metadata record and can be viewed by each user. EXIF viewer entries are generated automatically.

Languages – Multilingual User Interfaces: Phaidra uses UTF-8 (a variable-width encoding that can represent every character in the Unicode character set) throughout the entire system. This assures its use in all languages. The user interface is currently multilingual, including English, German, Italian, and Serbian.

Support and training: Online support is integrated throughout the system and can be specifically accessed whenever needed (just move your cursor over the relevant item for four seconds and then press the Help button). For technical questions, the university-wide help desk of the University of Vienna is available.

Reasons for storing/archiving data include:

- Enhanced opportunities for multi-disciplinary research – leading to new discoveries
- Avoidance of the duplication of research efforts (sometimes irreplaceable and/or expensive)
- Long-term preservation may be required by research funders
- Open Access purposes
- Self-archiving purposes
- Sharing of data with other repositories
- Interlink data to projects/publications
- Provide richer context for data value

Incentives provided to the researcher are, in particular, visibility, citation (through persistent identifier), reputation, and the long-term preservation of data.

In Phaidra, there are no restrictions in place (see the policy above). Likewise, there is no restriction concerning the formats or the content ingested by the repository. Of course there is a policy concerning the perpetual preservation of digital objects (this topic is one of the crucial issues of the training units where data providers are informed that long-term preservation can only be guaranteed for certain formats).

There are several kinds of licence models in use, including Creative Commons.

Promotions for (open) data sharing: Phaidra organises workshops, Phaidra Days, international conferences organised by the library, training activities, printed materials, service pages in German and English.

4 What kind of support services are provided to researchers?

Open Access activities: the University of Vienna has established an Open Access Office which organises OA activities at the University of Vienna. It aims to support and advise all members of the University and is involved in co-operations with national and international agencies regarding this discussion. The University of Vienna is also a signatory to the Berlin Declaration. An interdisciplinary working group was installed which, since its establishment in 2008, has been busy implementing an appropriate infrastructure for Open Access publishing at the University of Vienna that is closely linked to the advisory services. Participating colleagues bring expertise from different departments of the University Library. An Open Access Board was also founded in 2012 to give strategic direction to OA at the University of Vienna. Members of this Board are from the Rector's office, University Library, Research Service and Carrier Development Quality Assurance and researchers from five different departments. The University of Vienna is involved in SPARC Europe, COAR, SCOAP3, arXiv.org, LIBER, open-access.net, OpenAIRE and OANA (Open Access Network Austria).

A new ministry-funded project, e-Infrastructures Austria, started at the beginning of 2014. This is a nationwide initiative designed to foster the coordinated and ongoing development of Austria-wide infrastructures for digital resources in research. A key objective is to include research communities, libraries and IT services from all national universities and other scientific institutions, and to jointly develop knowledge for the construction of local repositories in order to pool existing expertise and resources. The intended goal is then to make this knowledge available to all of the project institutions involved. Within the three-year project period, all 25 partner institutions are to set up a local document server or be granted access to a shared document server. A parallel objective is the conceptualisation of either a central or several decentralised Austrian repositories for advanced applications and digital objects, including research data, multimedia content and e-learning content. The benefits of this project are as follows:

- Interoperability with existing systems at each institution and interoperability with existing domestic (e.g. Federal Promotion Funds and Austrian Research Promotion Agency) and international (e.g. Europeana, OpenAIREplus and publishers) systems.
- Facilitation of participation in international projects that require the presence of a repository
- Enabling Open Access according to respective policies throughout all institutions
- Preparation for Horizon 2020

Within the Phaidra LAB framework, workshops are organised on a regular basis (e.g. Phaidra LAB Tech or Phaidra LAB Metadata), along with meetings and hearings. Staff involved in the project are invited to these meetings. There is also a close partnership with the Austrian National Funding Agency FWF (they use the same repository, which is managed by our Computer Centre).

FWF's Open Access Policy for funded projects includes a clause for research data: "Whenever legally and ethically possible, all research data and similar materials which are collected and/or analysed using FWF funds have to be made available free of charge on the Internet either immediately after use in research publications or – if not used in publications – two years after the project is finished. Research data should be deposited in a way that it can be cited and re-used without restriction".

5 What kind of infrastructure is provided?

The starting point was the Repository of the University of Vienna (Phaidra), an acronym for Permanent Hosting, Archiving and Indexing of Digital Resources and Assets. Phaidra has been the comprehensive, university-wide Digital Asset Management System for permanent archiving functions of the University of Vienna since 2008. It enables instructors, researchers, administrative and organisational units as well as individual users to store their publications for research and teaching, as well as to document and archive them for longer periods.

Phaidra supports electronic publishing, archiving of images, display of collections, creation of eBooks, saving and streaming of videos, and much more.

All objects have a permanent digital signature, and the objects can be described in multiple languages ranging from self-archiving to Open Access publishing as well as the use of a sophisticated access system and tailored metadata structures.

Further initiatives involved the following projects:

- OpenAIRE for the setup of an Open Access infrastructure for research in Europe
- OpenAIREplus for Open Access to enhanced scientific information: Open Access publication of research data and metadata and relating them to publications/projects
- Participation in COAR

Phaidra is based on Fedora Commons and related technologies (http://phaidraservice.univie.ac.at/das-system-phaidra/facts/technik/). The repository supports a wide range of formats for images, video, audio and text documents.

Interoperability with other repositories/archives is supported through participation in international projects (e.g. Europeana Libraries) or direct contacts (e.g. OAPEN Foundation).

6 What have you learned so far? What's next?

Further developments:

- Release of the Study "Process Management for Research Data towards Sharing of Data", publication of the study and presentation at conferences (dissemination work)

- Introduction of a "Temporary Files Repository" (Phaidra TEMP) prior to Phaidra in order to enhance e-learning issues
- Release of Phaidra Plus, a docked application conceived for digital humanities http://phaidraplus.univie.ac.at/
- Controlled vocabulary (content) for the development of a terminology server (technical)
- Expansion of knowledge and resources on a nationwide basis via the e-Infrastructures Austria project

Further information

Website: http://phaidraservice.univie.ac.at/en/phaidra/what-can-phaidra-do/

Contact: Phaidra Repository Management: Paolo Budroni, paolo.budroni@univie.ac.at; Phaidra Customer Management: Susanne Blumesberger, susanne.blumesberger@univie.ac.at; Phaidra Plus (Digital Humanities): Ralf Pausz, ralf.pausz@univie.ac.at; European Projects: Gerda McNeill, Gerda, gerda.mcneill@univie.ac.at; Research Data and project e-Infrastructures Austria: Barbara Sanchez Solis, barbara.sanchez.solis@univie.ac.at

References

FWF. Open Access Policy for FWF-funded projects. http://www.fwf.ac.at/en/public_relations/oai/index.html

Budroni, P.; Höckner, M. (2010). Phaidra, a Repository Project of the University of Vienna; INFOtheca, No 1, vol XII, August 2011

Sánchez Solís, B. (2014). Factors for Enabling Sharing and Reuse of Research Data, study performed by NOAD Austria, https://phaidra.univie.ac.at/detail_object/o:344008