LIBER Case Study:

The Lithuanian National Open Access Research Data Archive (MIDAS)

AUTHOR: Jūratė Kuprienė, Vilnius University Library, jurate.kupriene@mb.vu.lt

KEYWORDS: generic, national

1 What was the starting point?

There are a number of different local registries, databases, and other local repositories used to store various scientific data in Lithuania. Diverse software tools are used to store data in various formats. The collection and storage of scientific data are often limited to the internal regulations and procedures of certain institutions. Access to scientific data is only available directly from the collecting institution. Therefore there is a huge risk of losing highly valuable data, and usage of this data in inter-institutional and multidisciplinary levels becomes rather complicated. As a consequence, Vilnius University decided to implement the National Open Access Research Data Archive (MIDAS) project.

The purpose of the project is to establish the infrastructure of a national research data archive that enables the collection and storage of research and empirical data, and ensures free, easy and convenient access to the data.

The project is led by Vilnius University in collaboration with the Vilnius University Hospital and another 13 science, study and medical institutions. The project duration is 30 months (January 2012 –June 2014) and funded by EU structural funds and a national budget (budget of the project: almost Lt 15 million, approx. EUR 4.34 million).

2 What kind of research data is targeted?

This is a versatile research data archive where scholars and researchers from Lithuania will be able to store their collected data.

The primary responsibility for data management lies with Vilnius University and consists of the following:

- Registration of users complying with the rules of the MIDAS user administration;
- Coordination and administration of activities concerning management, sharing and accounting of data stored in MIDAS;
- Foresight and provision of tools that prevent accidental loss, unauthorised destruction, alteration, disclosure or any other misuse of data;
- Determination of principles and rules as well as ensuring authorised and legal data management;
- Administration of the MIDAS data storage and archive; ensuring the safety and security of data stored in MIDAS, and guaranteeing continuous operation of the infrastructure itself, etc.

Responsibilities are shared with partner institutions, in particular the Vilnius University Hospital Santariškių Klinikos (Santariskes Clinics) has to:

- Administer biomedical research data in MIDAS and ensure the operation of the MIDAS research data management sub-system;

June 2014

- Ensure safe and secure biomedical research data administration by implementing organisational, technical, technological and methodological tools;
- Foresee and provide tools that prevent accidental loss, unauthorised destruction, alteration, disclosure or any other misuse of biomedical research data;
- Determine principles and rules as well as ensure authorised and legal biomedical research data management, etc.

In addition, science, study and other research institutions in Lithuania have to

- Administer their own research data by complying with the rules determined by the MIDAS governance institution;
- Ensure safe and secure data administration by implementing organisational, technical, technological and methodological tools;
- Take responsibility and guarantee that administration and disclosure of their own research data are authorised;
- Also ensure reliability of their own data stored in MIDAS by complying with legal acts under their power of attorney, etc.

Data will cover different subject areas and be preserved in various formats. Research data, its metadata as well as authorship data (if published) are preserved for an indefinite period as long as it is an indispensable part of pursuing MIDAS objectives.
The Vilnius University Hospital Santariškių Klinikos (Santariskes Clinics) will ensure that data stored in the MIDAS are compiled as a logically coherent collection of biomedical information in a database that is anonymised or granted written consent from the patient to allow biomedical research to be conducted, biomedical research data to be administered in MIDAS, and data access for MIDAS users and data reuse in other research projects.

The anonymisation module will ensure that data used in research are immediately modified in such a way that the patient cannot be identified.

Personal data collected in biomedical or other research are preserved until the end of that specific project. After the research is completed, the collected data are anonymised and then preserved for an indefinite period.

3 What is the organisational framework?

Roles and responsibilities

The lead institution is Vilnius University while the project partner is the Vilnius University Hospital Santariškių Klinikos (Santariskes Clinics). Further project participants comprise 13 science, study and medical institutions.

Data governance and administration will be managed by Vilnius University; another administrator is the Vilnius University Hospital Santariškių Klinikos (Santariskes Clinics) which is responsible for biomedical data management. The science, study and research institutions in Lithuania are responsible for their own data management.

Policies

The MIDAS information system policy and data protection for the MIDAS information system policy have already been approved. The development and adoption of other policies will be initiated after the project has been completed. Open access to research outputs is regulated by the Republic of Lithuania Law on Higher Education and Research (Article 45).

The main reasons for archiving / storing data include:

- Guaranteed safety of scientific data and effective sharing among science and study institutions in Lithuania.
- Increased quality of research outputs:
  a) Researchers will be able to explore and examine more data;
  b) Easier access to research data will involve and encourage more academics and other people to participate in discussions about the accuracy, reliability and relevancy, etc. of scientific results
- Increased efficiency of research performance:
  a) Researchers will be prevented from duplicating work in terms of research data collection;
  b) Time and material resources will be saved while collecting the same type of data from different institutions;
  c) Time and material resources will be saved significantly by collecting data preserved in medical institutions and providing virtual access with no special software or hardware requirements.
- Increased variety of research outputs:
  a) Easier access to data will foster research that is unrealisable or economically unviable in practice due to time constraints and other resources for data collection.
  b) Scholars, PhD students and others will have an interest in creating original and exclusive papers by formulating more diverse aims and objectives.
- Increased quality of education:
  a) Lecturers, associate professors and professors will have more options to exploit the newest and most relevant data as well as be more informed while sharing knowledge and giving assignments to students.

Information infrastructure tools will be developed and implemented in MIDAS to ensure that scholarly communication flows smoothly. Scholarly communication is taken to mean communication among researchers as well as information sharing between science institutions and the dissemination of scholarly outputs both nationally and globally.

Services and tools implemented in MIDAS will allow users to measure and analyse research data that are uploaded to the archive.

MIDAS will provide virtual services for researchers and other participants in the education process that lead to more effective and higher-quality research. Users will be able to register, find and cite data, search for other infrastructures or tools which provide data archiving services and then use them. They will also be able to share or integrate data and tools into any other science and study infrastructures.

It is planned that MIDAS will work as an open access archive for research data, but discussions concerning licensing issues are still taking place. In addition, (open) data sharing is actively promoted. The publicity programme include tools such as a permanent information stand, articles in scholarly journals and science magazines, a web banner, national seminars for researchers across various fields of science, a national large-scale conference presenting the archive, an e-publication containing detailed information, etc.

4 What kind of support services are provided to researchers?

The needs assessment was carried out in 2010 before starting the project. Science institutions were asked to fill in a survey about existing practices in research data collection; the type of data they collect, the kind of software they use; whether or not they are satisfied with the current situation.

During the system design process in 2012, a similar survey was conducted again but this time was designed to determine the needs of potential users of MIDAS (15 institutions).

Meetings/discussions with researchers were also held during the system design process.

Furthermore, national seminars are planned for researchers across various fields of science by the end of the project, a national large-scale conference will be held to present the archive, and an e-publication containing detailed information will be issued, etc.

The research data archive MIDAS is intimately related to the Lithuanian Academic e-Library (eLABa) and its current project addressing the creation and development of eLABa integral services. The communication between eLABa and MIDAS systems is one of the activities included in the LITMIS programme (The Development of Lithuanian Information Infrastructure for Science and Studies in 2013-2015).

Training for librarians is not available at the moment.

5 What kind of infrastructure is provided?

The main focus lies on data storage, data analysis, and the re-use and preservation of data. There are plans to use open-source technologies (though it is not easy to provide a clear answer because the project is still ongoing).

The MIDAS portal will provide one-stop access to archive materials and offer all users full functionality of the front-office. The MIDAS services will be accessible to recipients through a single, central portal containing a registration form, an information search box, the option to group and sort search results and download files. In addition, the MIDAS platform will support analytical tools.

Access conditions have not yet been specified but there are plans to make it an open system where certain services depend on the level of authorisation.

Interoperability with other repositories/archives will be taken into account and implemented in a final product. A created archive should be capable of communicating and exchanging data with other systems because of applied standards such as the OAI-PMH mechanism and WS-I specifications. Furthermore, a possibility of sitemap integration into the MIDAS portal will be realised to ensure interoperability with the most popular search engines – Google Scholar, Bing and Yahoo.

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

Designing and creating data archives is usually very expensive, time-consuming and demanding work. Sufficient financial and human resources should be gathered and guaranteed in advance.

Things to keep in mind:

- Be aware of scholars', researchers', and institutions' interests and needs;
- Have a team of well-qualified and competent specialists available;
- Plan properly and follow deadlines.

Create an archive, load it with data, maintain and keep developing; create a legal framework; become part of the electronic infrastructure in Lithuania where all researchers can manage their scholarly outputs, projects, research data.

Further information

Contact: Prof. Dr. Habil. Gintautas Dzemyda (Project Manager), Director of Vilnius University Institute of Mathematics and Informatics, e-mail: gintautas.dzemyda@mii.vu.lt

References

Dzemyda, G. (2013). 'Nacionalinis atviros prieigos mokslo informacijos duomenų archyvas MIDAS', presentation made at the LITMIS konferencija: Informacijos infrastruktūra mokslui ir studijoms, Vilnius, December 12, 2013. Available from conference website: LITMIS konferencija: Vilniaus universiteto biblioteka, http://www.litmis.mb.vu.lt/wp-content/uploads/2013/12/6_Dzemyda_MIDAS.pdf

Jurgutis, M., Petrauskienė, Ž. (2011). Lietuvos Mokslo duomenų archyvas. In: Kompiuterininkų dienos, Vilnius, pp. 29-33.

Ministry of Finance of the Republic of Lithuania (2011). Nacionalinis atviros prieigos mokslo informacijos duomenų archyvas, 2011. Available from: ES struktūrinė parama: LR Finansų ministerija, http://www.esparama.lt/paraiska?id=32177&pgsz=10

Petrauskiene, Z. (2012). National Open Access Research Data Archive (in Lit. MIDAS), presentation at NORDBIB 2012: Structural frameworks for open, digital research: Strategy, policy & infrastructure, Copenhagen, 11-13 June 2012, http://conference.nordbib.net/fileadmin/user_upload/dokumenter/A_Petrauskiene.ppt

Petrauskiene, Z.; Kriviene, I. (2013). Support to and implementation of Open Access ideas at Vilnius University, presentation at the COAR Annual Meeting 2013, https://www.coar-repositories.org/files/6_Vilnius_university.pdf

Republic of Lithuania Law on Higher Education and Research, 30 April 2009 No XI-242, Vilnius, as last amended on 24 April 2012, No XI-1987, http://www3.lrs.lt/pls/inter3/dokpaieska.showdoc_l?p_id=424020 (lt), http://www3.lrs.lt/pls/inter3/dokpaieska.showdoc_l?p_id=438419 (en)

Valionis, A. (Ed.) (2014). Nacionalinis atviros prieigos mokslo informacijos duomenų archyvas (MIDAS): iššūkiai, galimybės, Mokslas ir technika, No. 1, pp. 20 – 21, http://www.mokslasirtechnika.lt/mokslo-naujienos/nacionalinis-atviros-prieigos-mokslo-informacijos-duomen-archyvas-midas-iss-kiai-galimyb-s.html

Vilnius University (2013). Nacionalinio atviros prieigos mokslinių tyrimų duomenų archyvo informacinės sistemos nuostatai, patvirtinta Vilniaus universiteto rektoriaus, pp. 1-12