The current state of CRIS-IR interoperability at the University of Porto

Lígia Maria Ribeiro, Universidade do Porto
Praça de Gomes Teixeira, 4099-002 Porto PORTUGAL
lmr@reit.up.pt

Maria Eugénia Matos Fernandes, Universidade do Porto
Praça de Gomes Teixeira, 4099-002 Porto PORTUGAL
efernand@reit.up.pt

University of Porto (www.up.pt) is a public research university in Portugal. By the time of the celebration of its 100th anniversary, its mission is clearly linked to the 'generation of scientific, cultural and artistic knowledge, the high level education strongly anchored in research, the economic and social value of knowledge and the active participation in the progress of communities in which it operates'. U.Porto serves nearly 32,000 students, 2,400 teachers and researchers along with 1,700 non-teaching staff distributed by 14 schools and 60 scientific research units, spread across 3 university campuses located in the city of Porto.

The valorization of the skills and knowledge generated within U.Porto as well as the establishment of linkages with other actors are more and more perceived by the university as key tools to respond to the current challenges deriving from the globalization of markets or the technological change. U.Porto is a key player in what comes to the aims of increasing regional economic competitiveness through research, technological development activities and innovation and promoting the visibility of regional research-driven clusters.

The University of Porto is a vast and complex organizational and information macro-system. The amount and variety of information that reflect the dimension and complexity of the organization behind, emphasises the need of an integrated approach to support information management.

The management of this systemic universe relies on SIGARRA (Information System for the Aggregated Management of Resources and Academic Records), a transverse and integrated information system developed in house to facilitate the production, storage, flow, access and dissemination of the information generated within the different organizational processes, of pedagogical, scientific, technical and administrative nature. SIGARRA promotes internal cooperation and cooperation with external academic, scientific and business communities, and gives significant contribution to the cohesion of the community itself, encouraging good practices and compliance with procedures. SIGARRA started at the Engineering Faculty of U.Porto in 1996 and is being used since 2003 by all faculties of the University[i]. It includes the student and human recourses management systems and interacts with other systems used by the University, such as the financial management system, library and e-learning management systems and the DSPACE platform that supports the U.Porto Repository. We may say that SIGARRA it a kind of backbone that supports the full activity of the University.

In respect to research, SIGARRA includes the components that are usually seen on CRIS and its scientific production module is linked with the Open Access Repository of U.Porto[ii].

Recognizing the benefits of open access and having subscribed the Berlin Declaration, U.Porto decided to create an Open Repository for its scientific production, available in the Internet since November 2007[iii]. One year later, on December 2008, the Open Repository of the U.Porto was certified as being compliant with the directives contained in the Digital Repositories Infrastructure Vision for European Research (DRIVER) and became part of the repository Directory of this European project. At the same time, the Senate of U.Porto approved the Open Access Policy of the University, applied to the scientific production of the academic

community, covering journal articles, proceedings, master and doctoral theses, reports and other publications, as well. The importance of this regulation was rather significant because it embodies the institutional support and commitment of U.Porto to the principles of open access to scientific literature.

From the moment it was created, the number of publications of the Open Repository has grown steadily. At the beginning it registered less than 1.000 full-text and open access publications and now it has more than 25.000. Several factors contribute to its growth. Beyond the scientific production de per se, the open access policy of U.Porto and the interoperability between the information system (SIGARRA) and the repository stimulated the registration and publication of the intellectual works produced by the academic community[iv].

U.Porto conceived its repository based on organizational assumptions and its management takes place centrally in spite of a distributed access. Plus the open repository, where scientific literature is delivered, it includes the Thematic Repository[v] - a set of special collections that consist of digital libraries, legacies from ancient teachers and alumni and learning materials addressed to students with special and specific needs - and the Digital Archive of U.Porto that was born in the sequence of the need of gathering, describing, preserving and disseminating series of scanned documents. These contents are in full text and the great majority of them are in open access. Faculties and other units of the University self-archive in this platform their digital objects and corresponding metadata.

The most recent development resulted in the conception of a prototype of repository for scientific data. Aware of the importance of data curation and of the access to the scientific data generated by its researchers U.Porto developed a data scoping study in 2011 on available datasets[vi]. The recommendations from pioneering actions in scientific dataset auditing were followed, namely the methodology proposed by Data Asset Framework[vii] and helped to establish the main lines of the data audit experiment. The goal was to gather inputs from researchers regarding the issues of preserving and sharing data.

Several researchers, from different scientific areas and domains - life sciences, engineering, social sciences and arts - were interviewed. These contacts were essential to understand their interests, needs and priorities related to the data. A wide variety of situations were detected, ranging from an early awareness of digital fragilities to a well-sustained practice of depositing data in international databases and sharing them within communities of research partners. The importance of data re-use had already been recognized by some of the researchers. On the other hand, opening data after the conclusion of the research process was an issue not accepted by nearly all. In fact, researchers proved to be very cautious with the exposure of data without any constraints, unlike what happens with scientific publications. Other interviewed researchers were willing to share their data in specialized communities of research and we realized that this restrict and controlled environment was what they preferred. We also observed that the great motivation to data curation was the specific needs of exploring and searching the data collected and produced along the research processes[viii]. However, the incorporation of data curation procedures into the research workflow revealed to be quite distant yet.

The data audit leaded to the development of a prototype data repository built as an extension to the DSPACE platform and to a proposal of a curation workflow that comprises the deposit of datasets by the researchers, an interview with the data curator so that he can convert datasets into spreadsheets and get the required elements to produce contextual and descriptive metadata[ix]. This model was designed taking into account what seemed to be the real needs of researchers, rather than by any abstract data management convenience. At this moment, the repository contains two different types of datasets, some of them in open access, other requiring permission to be delivered. This prototype will contribute to get feedback from the researchers and to the definition of research data management services for U.Porto. The curation workflow is being tested on the sample datasets provided by the researchers.

At the present time, besides continuing the development of the scientific data repository, we are working to foster the CRIS-IR interoperability, exploring the linking between projects, publications and data.

In this work we will present the current state of CRIS-IR interoperability at the University of Porto and the main results achieved.

We strongly believe that open distributed and interoperable e-science infrastructures are nuclear to foster cooperation between academia and industry for leveraging innovation and creating new scientific knowledge with strong impact on economic and social development.

Key words: information systems, repositories, interoperability, scientific production.

[i] Ribeiro, Lígia M., Gabriel David, Ana Azevedo e J. C. Marques dos Santos. 1997. Developing an information system at the Engineering Faculty of Porto University. In Proceedings of the EUNIS 97 - European Cooperation in Higher Education Information Systems, ed. Yves Epelboin and Jean-François Desnos, 282-287. Grenoble, France.
[ii] http://repositorio-aberto.up.pt/
[iii] Ribeiro, Lígia M. e Maria Eugénia Matos Fernandes. 2009. O Repositório Aberto da Universidade do Porto, 4.ª Conferência sobre o Acesso Livre ao Conhecimento, Universidade do Minho, Braga, Portugal.
[iv] Ribeiro, Lígia M. and Maria Eugénia Matos Fernandes. 2010. The U.Porto Open Repository: the role of the Information System. In Proceedings of the EUNIS 2010 – European Cooperation in Higher Education Information Systems.Warsaw, Poland.
[v] http://repositorio-tematico.up.pt/
[vi] http://sciencedata.up.pt/doc/doku.php
[vii] DAF Team: The Data Asset Framework Implementation Guide.
[viii] Ribeiro, Cristina and Maria Eugénia Matos Fernandes. 2011. Data Curation at U.Porto: Identifying current practices across disciplinary domains. IASSIST Quarterly, 35(4):14–17.
[ix] Rocha, João, Cristina Ribeiro, João C. Lopes, Managing multidisciplinary research data: Extending DSpace to enable long-term preservation of tabular datasets. 2012. 9th International Conference on Preservation of Digital Objects, iPRES.