Enriching CRISs through new services for OA repositories

Raf Dekeyser[†], David Prosser[‡]
[†]Katholieke Universiteit Leuven
[‡]SPARC Europe

Summary

During the past years several proposals have been formulated or developed for adding new functionality to the network of Open Access Repositories of scholarly papers. The workshop is intended to discuss how the content of CRISs can be enriched through linking with these new services.

1 Introduction

Open Access repositories of scientific papers are slowly being filled, and everyday new repositories are added to the existing network. These repositories are either institution-based (containing material from the wide range of subjects covered within an institution) or subject-based (containing material from a wide geographical spread of researchers). In either case, the repositories are built to internationally agreed standards (Open Archive Initiative standards) to allow the sharing of metadata and harvesting of objects. By building repositories to common standards, services can be constructed that make use of the material within the repositories. To date, services have concentrated on alerting and searching functions, but it is possible to envisage a wider range of services as the range of material deposited within repositories increases.

Repositories are being increasingly seen by the funders of research as an essential tool in the dissemination of the results of the research they fund. Both the Wellcome Trust in the UK and the National Institutes of Health (NIH) in the US have enacted policies that require (the Wellcome) or encourage (the NIH) grantees to deposit copies of their final, peer reviewed manuscripts in central, subject-based repositories. Many other funding bodies worldwide are considering similar policies with either strongly encouraged or mandated requests to deposit in suitable subject-based or institutional repositories. This is seen by the funding bodies as a core part of their responsibility as their 'mission of disseminating knowledge is only half complete if the information is not made widely and readily available to society.' (Berlin Declaration, 2003). In addition, a growing number of individual institutions worldwide are adopting policies that require their researchers to deposit within local repositories (e.g. CERN, Queensland University of Technology, Bielefeld University, University of Bremen, University of Hamburg, Universidade

do Minho, University of Southampton, Case Western Reserve University.)
The original goal of the open access repository initiative was to give universal access to all of the world's research literature. However, it is becoming apparent that there will be a series of fascinating and important side-effects of collecting and making available the intellectual output of the world's universities. One such area of benefits lies in the interaction of OA repositories and CRISs. Both systems will contain a vast amount of information about the research activities of an institution and it makes sense that they be interfaced in such a way that they can share data (so reducing the need to input information multiple times into multiple systems). Having a full list of an institution's total research output (in many cases for the first time) will allow for improved research management and tracking, for enhanced reporting, and for the development of improved quality metrics. A marriage between an institution's CRIS and OA repository will provide a powerful management and research tool, as well as a show-case of the institute's intellectual achievements.
In the meantime, various members of the OA community are investigating how the use of the global infrastructure of repositories might be optimalized through the development of new services. Some of these newly proposed or developed services can in a straightforward way be applied for extracting from the repositories elements of information that are relevant for CRISs. The CRIS community may wish to work with the repository community to influence development of repository services in a way that benefits CRISs. In a similar way, there will be developments in the CRIS community that can be transferred to repositories to enhance repository development and to ensure a seamless system of research information from initial project proposals through to final published research papers.

2 Examples

The workshop should start with a series of short communications about recently introduced or proposed new applications or services. We mention only some of the possibilities:

2.1 Academic bibliographies

University administrations are eager to keep track of the scientific output of their staff. The main reasons for doing so are the requirement of reporting to the authorities about the scientific activities of the institute and the wish to dispose of objective criteria for staff evaluation. OAI-compliant software, like EPrints or DSpace, have proven to be excellent tools for the automatic collection of the metadata that constitute the *Academic Bibliography* as a complete inventory of the publications of the institute. Extra services may be implemented, such as linking with external databases (e.g. with the JCR for the impact parameters of journals or with a local authority file of the staff members). Uploading of the full text – an

essential element of the OAI philosophy – is not a real requirement in this context. Nevertheless, more and more universities are feeling the need to preserve a full digital archive of their publications and awareness of the advantages of making available as much content as possible in Open Access is growing.

2.2 The Open DOAR listing

Although most open access repositories have been designed to allow information about themselves to be gathered automatically, discrepancies can creep into the system. In order to build the new DOAR listing (Hubbard & Bjørnshauge 2006), each of the repositories has been visited by project staff to check the information that is gathered. This in-depth approach gives a quality-controlled list of repository features. The aim is to create a bridge between repository administrators and the service providers which "harvest" repositories. General search often brings back too many "junk" results. Information from OpenDOAR will enable the search service to provide a more focussed search by selecting repositories that are of direct interest to the user.

2.3 Open access citation information

The JISC Committee for the Information Environment (JCIE) Scholarly Communications Group commissioned a study (Hardy et al. 2005) to identify a framework for universal citation services for open access materials. The result should become an ideal structure for the collection and distribution of citation information and the main requirements of such services. As an example, this could lead to a harvesting service for new citation indices that include all Open Access material.

2.4 Usage measurement of publications

A new system was recently proposed (Johan Bollen et al 2005) for assessing the impact of publications, based on OAI-PMH harvesting of usage information. The download activity from journals and repositories by researchers at LANL was logged and analysed. Download activities from different institutions can be harvested through OAI-PMH technology. The results have been compared with the IF-ranking of ISI, revealing much similarity but also some striking deviations. The analysis can also be used to map the structure of science through journal relationships.
Discussions are underway to set up a similar measuring activity with a network of European universities.

3 Discussion

The previous contributions should be followed by a broad discussion about the usefulness of the proposed services for CRIS systems, about the needs for additional developments, and an identification of areas of mutual interest and possible joint development. A series of questions will be put to a panel composed of experts and speakers, but it is expected that all participants should contribute.

4 References

Berlin Declaration, 2003, http://www.zim.mpg.de/openaccess-berlin/berlindeclaration.html.

Bollen, R.; Luce, S.; Vemulapalli S.S.; Xu, W. (2003): Usage analysis for the identification of research trends in digital libraries, D-Lib Magazine, Vol.5.

Bollen, J.; Van de Sompel, H.; Smith, J.; Luce, R. (2005): Toward alternative metrics of journal impact: a comparison of download and citation data, Information Processing and Management, Vol. 41, No. 6, p. 1419-1440 (also: arXiv:cs.DL/0503007).

Hardy, R.; Oppenheim, C.; Brody, T.; Hitchcock, S. (2005): Open Access citation information, http://eprints.ecs.soton.ac.uk/11536/.

Hubbard, B.; Bjørnshauge, L. (2006): http://www.opendoar.org/documents/OpenDOAR_Press_Release_Jan06.pdf

5 Contact Information

Raf Dekeyser
Campusbibliotheek Arenberg, de Croylaan 6, B-3001 Leuven-Heverlee, Belgium
e-mail: raf.Dekeyser@bib.kuleuven.be
David Prosser
Director of SPARC Europe, Oxford University Library Services, Osney One Building, Osney Mead, Oxford OX2 0EW, UK
e-mail: david.prosser@bodley.ox.ac.uk