Models for integrating institutional repositories and research information management systems

Michael Day

UKOLN, University of Bath, Bath BA2 7AY, United Kingdom
m.day@ukoln.ac.uk

Currently, institutional repositories and research information management systems - where they exist - are often two distinct components of the wider information infrastructure required by higher education institutions. While there is a degree of overlap between them with regard to content and functionality, each has evolved to fulfil quite distinct business requirements.

Institutional repositories (IRs) are, for the most part, collections of content and metadata designed to support resource discovery and access, some with the additional ambition to ensure that institutional content is managed in a sustainable way. IRs themselves provide basic search and browse facilities, but most are also able to utilise tools like the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) to enable the sharing of metadata about repository content with aggregator services. The main content types handled by IRs to date typically include dissertations and research outputs - chiefly papers and publications at various stages of their lifecycle - but they could potentially also be used to manage a range of other institutional content types, including: technical reports and other grey literature, learning objects, administrative records and research data. The metadata collected about repository content could conceivably be represented by any suitable format, but many IRs currently use a form of Dublin Core. In any case, Simple Dublin Core is the format specified for sharing with aggregators through OAI-PMH.

Research information management (RIM) systems are essentially databases that are designed to help institutions manage the entire research process from initial funding opportunities through to the completion of projects, and the supply of information for research assessment. Information is gathered about many different things, including funding streams, people, as well as research outputs - typically publications, although they could conceivably comprise any tangible artefact of research activity [1]. RIM systems typically derive information from a number of other institutional systems, including finance, human resources, student records and databases specifically developed to support research assessment - a key driver of RIM system development. Because of their mainly institutional focus, many RIM systems have been developed in-house, although there is now a growing market for commercial solutions. The data standards used by institutions tend to differ depending on institutional requirements, but there is a degree of high-level convergence on the CERIF (Common European Research Information Format), especially where such

information needs to be shared with national or international initiatives. The feasibility of developing a national standard to support the exchange of research information is currently an area of active exploration in the UK [2].

While the specific requirements of IRs and RIM systems are different, the degree of overlap between them with regard to content and functionality means that there have been a number of attempts to align the two. A good example is the National Academic Research and Collaborations Information System (NARCIS) in the Netherlands, a national RIM which incorporates the national repository network DAREnet. The technical aspects of interoperability between IRs and RIM systems are being dealt with by initiatives like the Knowledge Exchange CRIS-OAR project as well as a number of UK research projects funded by JISC, e.g. Readiness for REF (R4R), Building the Research Information Infrastructure (BRII) and (more recently) the CRISPool project at St Andrews University [3].

In addition to this essential work on supporting interoperability at the technical and semantic level, there may be a need to explore some other ways in which both IRs and RIM systems might need to evolve in the future. It has already been noted that both systems currently tend to view 'research outputs' as primarily meaning peer-reviewed publications. However, the (so called) 'data-deluge' combined with the promotion of 'open science' principles [4] and the growing tendency of some funding bodies to encourage researchers to curate and provide access to research data means that both IRs and RIM systems will most likely need to be able to deal increasingly with metadata (at least) about extremely complex compound objects. In this connection, the EU DRIVER project has explored the concept of 'enhanced publications' - publications that are enhanced in some way with research data, illustrative materials, annotation, etc - and identified some of the ways in which they might be packaged for handling by repositories [5]. A promising model for packaging such compound objects is the OAI's Object Reuse and Exchange (ORE) data model. Interestingly, the ORE data model has also been used in a recent attempt to represent entire research life-cycles in one particular area of science (sensor network applications in seismology), and was able to establish relationships between publications, data and wider project contexts in a way that should be of interest to the developers of RIM data models [6].

References

1. Joint, N.: Current research information systems, open access repositories and libraries. Library Review 57(8), 56-575 (2008)
2. Rogers, N., Huxley, L., Ferguson, N.: Exchanging research information in the UK (2009). Retrieved April 23, 2010 from: http://ie-repository.jisc.ac.uk/448/
3. JISC: Strategic themes: Research information management. Retrieved April 23, 2010 from: http://www.jisc.ac.uk/whatwedo/themes/informationenvironment/researchinfomgt
4. Hall, W., De Roure, D., Shadbolt, N.: The evolution of the Web and implications for eResearch, Philosophical Transactions of the Royal Society A 367, 99-1001 (2009)
5. Woutersen-Windhouwer, S., et al.: Enhanced publications: Linking publications and research data in digital repositories. Amsterdam University Press, Amsterdam (2009)
6. Pepe, A., Mayernik, M., Borgman, C. L., Van de Sompel, H.: From artifacts to aggregations: Modelling scientific life cycles on the Semantic Web. Journal of the American Society for Information Science and Technology 61(3), 567-582 (2010)