

CONNECTING THE DOTS: AN INTEGRATED APPROACH TO DIGITAL CONTENT MANAGEMENT IN AN INSTITUTIONAL CONTEXT

Context

There is a growing interest among academic institutions in collecting, preserving, reusing and creating value-added services from digital content produced in and for research, teaching and learning. The emphasis on research outputs and collaboration, and distance, flexible and online learning, together with developments in information technology, has led to an increased awareness that the digital content being created by members of the academic community is an institutional asset. It is also increasingly being recognised as an institutional challenge, requiring tactical management and a strategic response.

At the same time many academic libraries are responding to the challenges of new technologies by taking the opportunity to redefine their fundamental role in the creation, distribution and provision of access to information. Over the past decade libraries have moved almost completely towards a digital platform for management of the information they acquire or subscribe to. They have built significant digital collections of material published by others, and they are, increasingly, producing new content themselves¹. Often this content originates from, or is the intellectual property of, their own institutions.

All around the world universities, their libraries, faculties, research centres and information technology and course development units, are trying to cope with the digital revolution. There is a growing recognition and articulation of the convergence that is occurring among the various digital initiatives with which universities are engaged, and the opportunities for potential synergies and more significant outcomes through collaboration².

Neil McLean³ stresses the need for this growing convergence to be viewed from a service perspective, rather than a delivery perspective. He argues that no online learning or research environment can be successful without relatively seamless access to information resources at the point of need, and says further that the challenge remains to find a balance between systems support, "learning containers", information resources and sound pedagogical principles.

Through COLIS (the Collaborative Online Learning and Information Services model), McLean demonstrates that the new electronic environment must comprise a complex interactive matrix that is dependent on the information resources mentioned above, as well as on user databases, content and rights management software, metadata repositories and so on.

Sally A Rogers, from Ohio State University, argues that the full array of a university's digital assets and information services should be broadly defined, and should include the library's catalogue, the electronic journals, reference databases and other electronic resources available through the library, as well as institutional repositories and resources created or collated elsewhere in the university, such as course material.

¹ Refer Harboe-Ree, *The library as digitorium*.

² Refer Rogers, *Developing an institutional Knowledge Bank at Ohio State University*.

³ Refer McLean, *Libraries and e-learning*.

Brandon-hall.com⁴ staff use the concept of “enterprise-wide systems framework for the management of intellectual capital”. Following a survey of emerging trends in 2001, brandon-hall.com staff concluded that, while the merging of learning management and knowledge management is in its early stages, there are numerous signs of the coming integration. Rogers in 2002 noted the overlapping of such initiatives as digital collections, course web sites, electronic course packs and learning objects, the desirability of integration to search across these repositories and the development of standards to promote interoperability. Rogers also highlights the potential of increased interoperability and connectivity to generate innovation in research, teaching and learning.

An integrated approach

In this context, the move by academic libraries to establish e-print repositories and e-publishing capability should be seen as part of an environment containing a complex and converging suite of intellectual property sources required to support the contemporary research and learning environment.

Ohio State University has coined the phrase Knowledge Bank to refer to this complex environment. In developing its Knowledge Bank concept Ohio State University identified a number of expected benefits, which include:

- improved access to scholarly communication throughout its lifecycle (e.g., informal communications about initial research questions and data, classroom presentations, preprints, technical reports, formal publications);
- integration of content to drive knowledge;
- synergies and economies of scale realised through cooperative effort;
- archiving and preservation of digital output to secure long-term access;
- leveraging of institutional knowledge capital (e.g. innovative re-use of research and learning materials);
- revenue potential with selected items; and
- increased visibility for Ohio State University.

While this list does not specifically identify the benefits of collaboration with other institutions embarking on similar strategies, the Knowledge Bank report stresses the importance of this, and Ohio State University is now one of a group of North American and European universities participating in the DSpace Federation⁵. These institutions will test software already developed by Massachusetts Institute of Technology (MIT), and offer suggestions about how to further develop and improve it.

The October 2002 draft report of the Higher Education Information Infrastructure Advisory Committee (HEIIAC) of the Department of Education, Science and Training (DEST)⁶ identifies the following critical features of an enhanced research infrastructure:

- information infrastructure resources should optimise the efforts of researchers in the higher education sector to create, manage, discover, access and disseminate knowledge;

⁴ Refer *Learning management and knowledge management*.

⁵ For DSpace, refer p. 4 of this report. The DSpace Federation comprises Massachusetts Institute of Technology, Columbia University, Cornell University, Ohio State University, the University of Rochester, the University of Toronto and the University of Washington. Cambridge University is also a participant.

⁶ DEST [Draft] *research information infrastructure framework for Australian higher education*.

- access to the research information infrastructure should not be constrained by institutional affiliations, geographic locations or disciplines of individual researchers;
- collaboration among libraries has improved the effectiveness of individual institutions, and further collaboration, clear strategies and a shared vision would significantly improve the coordination of the national research infrastructure;
- opportunities should be sought for the academic community to regain control of scholarly publishing; and
- computing and communication technologies provide new opportunities for the creation, management, storage and dissemination of information.

The HEIAC report is primarily concerned with managing the current problems associated with scholarly communication and publishing, and it stresses the need to adopt a national collaborative approach. Neil McLean, Ohio State University and others embrace scholarly communication strategies and then argue that they should be incorporated into a more holistic approach to the management of institutional digital content and intellectual capital.

The merging of these two approaches would yield substantial benefits to Australian university communities, consistent with the following statements of principle:

1. Australian universities have a commitment to support and promote their institutions' research activity through the creation and preservation of digital content, especially institutional repositories and electronic publishing.
2. Australian universities have a commitment to help their institutions achieve their goals more effectively by assisting with the integration of digital resources.
3. Australian universities have a commitment to collaborating nationally and internationally in the achievement of a more integrated approach to the management and interoperability of digital content.

These statements reflect the HEIAC objectives and place them into a framework that, if implemented, would improve institutional and national efficiency and effectiveness.

There are many barriers to the adoption of an integrated approach, not least the size of the task, the lack of integration software, the cost and challenge of undertaking such a task in its entirety and the confusing plethora of possible approaches. Also to be overcome in many instances are significant cultural differences, especially within institutions, and the apparent durability and acceptance of current patterns of information creation and dissemination. Notwithstanding, there are significant risks in not adopting this approach, including the cost of *not* managing information effectively and without duplication of effort, and the potential alienation of end users.

Libraries' role

Libraries are natural, although not exclusive, information management leaders within their universities, based on their traditional print and growing digital content management expertise. This is not to say that academic libraries can or should assume responsibility for overall information management within their institutions. However it does suggest that they should be well positioned to exercise that leadership role, either as advisors, managers or practitioners.

There is a question, though, about how well equipped Australian university libraries *actually* are to maintain or extend their leadership role to the advantage of their institutions. The barriers to this include:

- an unwillingness to redefine their role to manage internally produced content (or to produce that content themselves);
- a lack of resources to undertake a broader or redefined role;
- the volume of data that could or should be coordinated or published;
- a lack of institutional support or readiness to assist with the necessary changes; and
- a lack of the particular technical expertise required.

Many Australian university libraries have started the process of redefining their role, but rarely to the extent of their American and European counterparts. We (that is, Australian university libraries) need to examine this issue to determine whether or not we **want** to redefine our role, and, if so, what strategies we need to put in place. We need to examine the issue of the resources required to broaden or fundamentally change our role, noting that many of the most innovative changes overseas have occurred as the direct result of significant government or philanthropic grants, building on an underlying base quite a lot stronger than that enjoyed by most Australian university libraries.

We also need to examine the internal, intra-university and national cultural and management constraints. Libraries are only one of the groups within a university that may see themselves as partial owners of this territory. Information Technology groups may claim technical expertise, and Learning and Teaching groups may see that any collaborative endeavour should be primarily driven by pedagogical considerations. The challenge is to avoid on the one hand “the horse designed by a committee” and, on the other, the “tragedy of the commons”. All the key players need to be involved but no one player should dominate.

One of the most pressing issues we as libraries have to deal with, perhaps surprisingly, is a lack of relevant technical expertise. I say “perhaps surprisingly”, because our natural strength is our ability to effectively manage large and complex information repositories. It can be argued that libraries have a number of areas of weakness technically that are currently preventing them from breaking through into a wider information management role. Libraries need to urgently develop their expertise in the areas of XML (Extensible Markup Language) and metadata, both of which are emerging as essential to the storage, use, integration, dissemination and preservation of information, and other wider information tools such as content management systems.

One of the leading exponents of the use of XML is Roy Tennant from the California Digital Library⁷. Tennant argues that XML has the potential to exceed the impact of MARC on librarianship (and hence on the management of scholarly information). XML is used extensively in the commercial sector, but has so far achieved little penetration into the academic sector, especially in Australia, despite the fact that, as claimed by Tennant, it is “quickly becoming the standard method by which information of a variety of types is encoded, transported and processed”⁸. Dick R. Miller, another commentator on XML, argues that, with its exceptional flexibility, generality, and convergence of functionality, XML is fast gaining favour as the universal format for data and document exchange – in effect becoming the lingua franca of the Information Age⁹. If this is the

⁷ Refer Tennant, *XML in libraries*.

⁸ Ibid, p.viii.

⁹ Refer Miller, *XML*:

case, libraries, and their institutions, need to acquire the appropriate skills, while noting that XML is not a solution in itself, but in combination with dedicated “helper” utilities its power becomes harnessable and apparent¹⁰.

Libraries have stronger metadata than XML expertise, but this expertise is limited to a smaller than required number of staff, given the important place metadata has in the management of web based information. To play a leadership role within their institutions, libraries must increase the number of staff with metadata expertise and the range of schemata to which this expertise can be applied.

Libraries *can* adopt a leadership role, and can overcome their technical deficiencies, but they cannot proceed in isolation and must therefore address the range of barriers to change or the adoption of a collaborative and integrated approach to the management of digital content, for whatever purpose, referred to in this paper.

In discussing libraries’ role it should be noted that the emerging Australian Digital Theses and e-print repository projects have both adopted a national collaborative approach, and both draw on international initiatives and common software, however neither of these projects is currently integrated with other university digital content activities (or, at this stage, with each other, although that is intended in the near future).

Enabling architectures

As already mentioned, North American libraries and universities are at the forefront of developing systems to support interoperable institutional or federated digital repositories. The key initiatives, the Fedora Project and DSpace, demonstrate the strategic directions possible. Also of interest are COLIS (which is working through some of the challenges of institutional interoperability), Greenstone (a New Zealand system for building digital collections) and Shibboleth (a North American initiative exploring specific aspects of interoperability), as well as AARLIN and the various gateways such as AustLit, AVEL, Picture Australia and others. An Australian commercial product, TeraText, is of interest because it appears to fulfil some of the required technical functions.

The Mellon Fedora Project

<http://fedora.info/>

In September 2001 The University of Virginia received a grant of US\$1 million from the Andrew W. Mellon Foundation to enable the Library, in collaboration with Cornell University, to build a sophisticated digital object repository system based on the Flexible Extensible Digital Object and Repository Architecture (Fedora). This new system demonstrates how distributed digital library architecture can be deployed using web-based technologies, including XML and web services. The new system is designed to be a foundation upon which interoperable web-based digital libraries can be built. After evaluation the software will be made available to the public as an open-source release.

DSpace

<http://www.dspace.org/>

MIT has a project under development called DSpace, which MIT Libraries and Hewlett-Packard are jointly developing.

¹⁰ Refer Donnelly, *DigiCULT technology watching brief 7*.

“DSpace is designed to provide stable long-term storage needed to house the digital products of MIT faculty and researchers.

For the user: DSpace enables easy remote access and the ability to read and search DSpace items from one location: the World Wide Web.

For the contributor: DSpace offers the advantages of digital distribution and long-term preservation for a variety of formats including text, audio, video, images, datasets and more. Authors can store their digital works in collections that are maintained by MIT communities.

For the institution: DSpace offers the opportunity to provide access to all the research of the institution through one interface. The repository is organised to accommodate the varying policy and workflow issues inherent in a multi-disciplinary environment. Submission workflow and access policies can be customised to adhere closely to each community's needs.” Use of the DSpace approach is demonstrated in the following web entrance to MIT Libraries.

The screenshot shows the MIT Libraries website homepage. At the top, there is a blue header with the text "MIT libraries" and navigation links for "SITE INDEX" and "SEARCH". Below the header is a navigation bar with "INFORMATION ACCESS @ MIT", "ALUMNI", "VISITORS", and "GIVING". The main content area features a "SPOTLIGHT" section with a photograph of a library interior and a bullet point: "Find out about SFX - get to online full text and library services even faster". Below the spotlight are five columns of links: "SEARCH OUR COLLECTIONS" (including catalog: BARTON, e-journals + databases: VERA, finding articles, MIT theses online, MIT research: DSpace, off-campus access, more...), "RESEARCH HELP" (including ask us!, help yourself, information navigator, virtual reference collection, more...), "SUBJECTS + COURSES" (including subjects: art, math, physics..., types: census, maps, patents..., course pages, theses, more...), "BORROWING + ORDERING" (including circulation, remote storage (RSC), interlibrary borrowing, suggest a purchase, order MIT publications, see also MIT Press, more...), and "ABOUT US" (including list of MIT libraries, faq, hours, jobs, news, staff, more...). The footer contains contact information: "webmaster@libraries.mit.edu | Contact us", "Massachusetts Institute of Technology", "77 Massachusetts Avenue, Cambridge, MA 02139-4307 USA", and a "Quick Links" dropdown menu.

Ohio State University and others' participation in the DSpace Federation has already been mentioned. Cambridge University, through the Cambridge-MIT Institute, is also a member of the Dspace Federation¹¹.

DSpace@Cambridge states that it will:

- provide a home for digitised material from the University Library's printed and manuscript collections;

¹¹ Refer <http://www.lib.cam.ac.uk/dspace/>

- have the ability to capture, index, store, disseminate, and preserve digital materials created in any part of the University;
- contribute to the development of the open source DSpace system, working with other members of the DSpace Federation of academic research institutions; and
- act as an exemplar site for UK higher and further education institutions.

COLIS

<http://www.colis.mq.edu.au/>

The Collaborative Online Learning and Information Services (COLIS) project was funded by DETYA to achieve the following objectives:

- establish a test bed for the development of collaborative online learning and information services;
- develop a scalable standards based model for institutional interoperability which enables the seamless sharing of online learning and scholarly information resources;
- contribute more fully to the work of the Instructional Management System (IMS) Global Learning Consortium; and
- link with international software companies, corporate management systems providers, learning management systems, content producers, and national government agencies.

Unlike Fedora and DSpace, COLIS is not intended to generate a product; rather, it is a model for exploring software, hardware, content management and rights management issues impacting on the inter-operability of digital resources.

Greenstone

<http://www.greestone.org/>

Greenstone is described on its website as “a suite of software which has the ability to serve digital library collections and build new collections”. It provides support for the acquisition of documents (either electronically or from print), the creation of navigation hierarchies and tagging of documents with a range of metadata schemata.

Greenstone is produced by the New Zealand Digital Library Project at the University of Waikato, and developed and distributed in cooperation with UNESCO and the Human Info NGO. It is open-source, multilingual software.

Shibboleth Project

<http://shibboleth.internet2.edu/>

Shibboleth, a project of Internet2/MACE, is developing architectures, policy structures, practical technologies, and an open source implementation to support inter-institutional sharing of web resources subject to access controls. In addition, Shibboleth will develop a policy framework that will allow inter-operation within the higher education community.

Shibboleth is a response to the need for organisations to exchange information about their users in a secure, and privacy-preserving manner.

TeraText

<http://www.teratext.com.au/>

Melbourne based InQuirion's product, the TeraText Database System, is designed to support the development and management of a metadata repository. Although it has been designed for text products, it is being used by the National Library for both PictureAustralia (images) and MusicAustralia (sound). It is designed to support XML, SGML, Unicode, Z39.50, RTF, HTTP and other industry standards. The TeraText DBS components are modular. They can be installed as a suite or as individual modules to work with existing database management and document-authoring systems. A content server searches on structural elements, document characteristics or multiple sources scaling from gigabytes to terabytes with the same high performance.

Where to from here?

Most Australian universities have recognised the need to improve their information management capacity, and most Australian university libraries are interested in supporting their institutions' research activity through the establishment of electronic repositories. Some university libraries are interested in taking this support further into electronic publishing, and some are supportive of the idea of linking the various digital content sources of their institutions, as described in this paper. DEST's Higher Education Information Infrastructure Advisory Committee has given in principle support to the proposal that Systemic Information Infrastructure (SII) funds be allocated to support the development of institutional repositories and electronic publishing. However there is concern about the possibility that this could be done in ways that exacerbate the current lack of interoperability between various digital sources. In addition, there are no concrete proposals for the use of these funds.

The objectives outlined in this paper could be advanced in a number of ways, in particular through:

1. Encouragement of demonstrator projects

- 1.1 **The trialling of key software products.** SII support could be used for the acquisition of key software, although some products are currently free (DSpace, Fedora, Greenstone). The real need is for the appointment of project teams to install and test the software and then impart this knowledge to others. If a commercial product were selected for trialling there would be costs associated with that, in addition to the project teams. Practically, this means that individual institutions, or collaborating groups of institutions, would undertake to test products and then impart their knowledge to the broader Australian university community.
- 1.2 **The customisation or adaptation of software.** It is probably not practical for "holy grail" software to be developed from scratch, particularly when there are some interesting products now available for testing, however it is likely that a degree of customisation would be essential, and this would need funds.
- 1.3 SII support could be used for the **development of XML and metadata skills within universities**, especially in libraries and information technology departments, or it could take the form of professional development funding, or funding to allow staff with these skills to train other staff in a train-the-trainer model.

2. **Populating digital repositories.** In these early days, from an Australian perspective, there is an urgent need to get critical mass of content into digital repositories. One of the main barriers to this is a lack of staff with appropriate expertise to facilitate the depositing of material. While it should not be a long term goal to rely on project funding for this purpose, a lack of resources, especially staff, is an immediate and major barrier to success.
3. **Scoping, benchmarking and international collaboration.** The most significant developments are occurring overseas, and it would be beneficial for Australian universities to use these developments as models and springboards for our own progress. In addition, Australian universities would benefit from entering into partnerships with overseas institutions.
4. The COLIS team, or others, may wish to develop a framework for institutions to follow, and/or identify elements of the project that can be extended or spun off into further collaborative investigation. Alternatively, other elements of **cross-institutional collaboration** should be developed, including, for example, digital rights management, content sharing, enabling new forms of research, and so on.

Authors: Cathrine Harboe-Ree, University Librarian, Monash University

Dr. Andrew Treloar, Manager Identity and Messaging Services,
Information Technology Services, Monash University

Date: 27 May 2003

Bibliography

Department of Education, Science and Training Higher Education Information Infrastructure Initiative Advisory Committee, [Draft] *research information infrastructure framework for Australian higher education*. October 2002.

Donnelly, Martin & Ross, S., *DigiCULT technology watching brief 7: the XML family of technologies*. <http://www.digicult.info/pages/twb.php>

Greenstein, Daniel and Thorin, Suzanne E., "California Digital Library", *The Digital Library: a biography*. <http://www.clir.org/pubs/reports/pub109/section2.htm>

Harboe-Ree, Cathrine, *The library as digitorium: new modes of information, creation, distribution and access*. <http://eprint.monash.edu.au/archive/00000018/>

Learning management and knowledge management: is the holy grail of integration close at hand?
http://www.brandonhall.com/public/whitepapers/lmkm/whitepaper_lmkm260101.PDF

McLean, Neil, *Libraries and e-learning: organisational and technical interoperability*. <http://www.colis.mq.edu.au/archives.htm>

Miller, Dick R., "XML: libraries' strategic opportunity", in *netConnect*, Summer 2000. <http://xmlmarc.stanford.edu/LJ>

Payette, Sandra & Staples, Thornton, "The Mellon Fedora Project: digital library architecture meets XML and web services", Sixth European Conference on Research and Advanced Technology for Digital Libraries. *Lecture notes in computer science*, vol. 2459. Springer-Verlag, Berlin Heidelberg New York (2002) 406-421. <http://www.fedora.info/documents/ecdl2002final.pdf>

Rogers, Sally A., *Developing an institutional Knowledge Bank at Ohio State University: from concept to action plan*. http://www.lib.ohio-state.edu/Lib_Info/srogersKBdoc.pdf

Staples, Thornton, Wayland, Ross & Payette, Sandra, "The Fedora Project: an open-source digital object repository management system", in *D-lib Magazine*, April 2003. <http://dlib.org/dlib/april03/staples/04staples.htm>

Tennant, Roy, ed., *XML in libraries*. New York, Neal-Schuman Publishers, 2002