

The brave new world of libraries and repositories

*Tom Ruthven
Executive Officer
Australian Partnership for Sustainable Repositories*

I thought I'd start with the last word in the title of my paper and work my way backwards, not because of some strange sense of style but because "brave" is such a big word compared to repository.

What is a repository?

H. Frank Cervone has neatly described an institutional repository as "software and services that manage and disseminate digital materials for an entire institution"¹. All things are equal in this definition. Repositories have as much emphasis on dissemination of material as managing the material.

Generally, repositories are being established in research institutions. The type of material they want to keep is scholarly in nature but scholarly in a very broad sense. There may be the output of research such as e-prints and theses. There can be research data such as songs recorded by an anthropologist on a field trip, recordings of music performances or data from an electron microscope. Printed material may have been digitised such as sheet music or geological maps. The repository may also contain teaching material, complete e-books, and research published as websites. Image, sound, data, text and multi-media works can be included in what can end up being a bewildering array of intertwined possibilities.

However, at their heart, repositories are a response to the changes in the way researchers are going about their work. Researchers are creating data digitally, communicating digitally, and increasingly publishing the results of their research digitally.

The management and dissemination of the digital material for scholars and researchers are the two building blocks of repositories. MacKenzie Smith has spoken of repositories as "cumulative and perpetual" and "open and interoperable"², and Clifford Lynch talks of the essential commitment of a repository "to the stewardship of these digital materials, including long-term preservation where appropriate, as well as organization and access or distribution."³

The new world

The new world of repositories, in many ways, looks remarkably like the existing world of libraries. Material or content is obtained, access is provided, and there is a commitment and effort made towards preservation.

Content is one of the keystones of any service, whether it is a library, an archive or your corner store, and this content must be good! As Brian Lavoie and Lorcan Dempsey have

¹ Cervone, H. Frank, The repository adventure. *The library journal*, June 1, 2004

² Smith, MacKenzie, Libraries in the lead: the institutional repository phenomenon.

³ Lynch, Clifford A, Institutional Repositories: Essential Infrastructure for Scholarship in the Digital Age. *ARL Bimonthly Report* 226, February 2003. <http://www.arl.org/newsltr/226/ir.html>

stated for repositories: “where coverage is intermittent and/or unpredictable, usefulness is diminished and stakeholder interest will not grow”⁴.

Developing content for a repository primarily relies on researchers adding their material, and enticements are needed for this to occur. Keeping research material safe and for a long time is a key enticement for researchers, but making it easy to contribute goes hand-in-hand with this. Researchers often (and usually) have large amounts of existing material they want to add. They may have used standard and not-so-standard digital formats or it may be in an analogue format that they need to digitise first. Providing simple loading mechanisms must cover the bulk transfer of existing material and small groups of newly created material.

Providing advice on equipment, standards and the amount of effort needed can be seen as an adjunct to developing the content. For example, you might get asked about the best hand-held digital sound recorder to record folkloric music in the field, or what standards should be used when digitising sheet music, or what is the absolute minimum information needed to put in a description of an e-print journal article. Even providing costing models for their next grant application to cover the creation and description of digital material would be of benefit to a researcher.

Another enticement is the increase in the visibility of their research output, and the consequent value that can be added to their material. Providing one front-end to access material in a repository framework is not enough. Material in a repository is “open”, meaning it is there to be shared. A repository makes it possible for a researcher’s output to be seen through many lenses. A piece of music can be found by searching the repository or that organisation’s website, but it might also be linked to from an article by the composer that appears in an electronic journal published in Italy, a criticism of the piece in an electronic journal published in Canada, or it can be harvested to appear in other services such as MusicAustralia.

A repository extends finding into sharing. Different sound recordings of one piece of music can be on various repositories around the world and a researcher can access and use them for their analysis without leaving their desk. Material from different areas of one repository, or different repositories, can be combined in a new ways to produce new publications or research data. These types of activities rely on repositories working out how to make the material “viewable” without the need for special software or equipment. A researcher does not want to load six different types of sound recording players on their PC to listen to pieces of music from six different repositories.

Finally, repositories lay claim to keeping material safe. I will address this shortly, but first I want to look at where libraries fit into this world that is part archival, part dissemination, part proselytiser.

What is a library?

Libraries have always assisted researchers communicate by getting and providing access to material; assisting them to find the information they want; and providing support services for their research and teaching needs. Repositories build on these strengths.

The research libraries’ current role in collecting and keeping information continues into the digital world. Collecting and making theses available has long been seen as part of a university’s obligation. Having an early version of an electronic score by a current composer in a library’s collection may be of much interest to researchers as a manuscript version of

⁴ Lavoie, Brian and Dempsey, Lorcan, Thirteen ways of looking at...digital preservation, *D-Lib magazine*, vol. 10, no. 7/8, <http://dlib.ejournal.ascc.net/dlib/july04/lavoie/07lavoie.html>

Mozart's Requiem. And, keeping regimental war diaries in a research collection is of interest whether they are written on paper or were stored on a palm-pilot. However, a repository can include much more than would normally be considered "library" material. For example, it is not normal for a library to collect research data such as sound recordings from field trips, artworks produced by members of an arts faculty, or data from an electron microscope.

Another area where repositories stretch the boundaries of libraries is in disseminating information and assisting with research. It is a shift from disseminating to publishing and assisting to collaborating.

There are many well-tested ways that have been developed over many years to tell people what is in the library, from adding information to the catalogue to information literacy classes. However, as noted earlier, the basis of dissemination of material in a repository is to make it available openly. That may be providing an interface to let people search and see the full material or it could be letting other people and institutions include links to the material in the repository in their service, or to harvest material. The material is often not published elsewhere and this, combined with a promotional role to make the material available, makes it more like publishing than the dissemination role normally played by libraries.

Providing assistance to researchers is also a feature of library services, whether that is providing a reference and research service, or copying facilities or tutorial "reading bricks". Repositories stretch this range of tasks into helping researchers choose appropriate ways to create digital material, assisting them to describe and load material to a repository, and working with them to provide access in meaningful ways. This is a more collaborative approach to research output than has previously been experienced, much more than just providing assistance.

To preserve and sustain

Keeping digital material safe and secure is a straightforward though not necessarily trivial task. The difficulty is continuing to make it viewable. An analogy is the U-matic video. You may have a perfectly fine tape but no videoplayer to view it on. It is not just preserving the tape that is important. You need the equipment to view it. The software and hardware used to view digital material becomes obsolescent rapidly and constantly. Keeping digital material means not just making sure bitstreams are all still there but also ensuring that people can still view the material. This is made more complex because of the great number of formats that digital material comes in. Many are proprietary and most that are incompatible with each other. This can range from simple still images in a TIFF format, through complex interactive e-books written for a particular type of hand-held reader, to astounding new sound/music/visual art explorations that combine user-driven sound, text and moving image with software written just for that work.

At this stage in the life of repositories there are more questions than answers. Some of the questions that repositories are grappling with include:

- Are we trying to preserve the full experience or the intellectual content only? For example, does it matter if the look-and-feel of a learned research article changes from browser to browser or over time if all the words are there? Should images and links from icons to sound recordings in the article appear on the screen at exactly the same place every time? Is it absolutely necessary for the moving image in the article to be in AVI format rather than MPEG21?
- There are questions about what we want to preserve. A course lecture and its notes may have a more ephemeral life than sound recordings of a significant singer.

Deciding how much effort to put into preserving different types of research material is a critical choice.

- For repositories, the material begins with the researcher. Choosing the format is a decision often made by the researcher. Capturing and creating technical, administrative and descriptive metadata can be tasks undertaken by the researcher. By the time the material arrives at the repository, it is often too late for the “owner” of the repository to control or find this information, even if they want to do this work. Current thinking is that the more that is known about a digital object, the easier it will be to decide how to preserve it. How do we preserve material in the absence or paucity of information? As an aside here, I would like to mention the value for repositories of all the descriptive metadata that libraries have been creating with catalogue records of such things scores, sound recordings and theses. The re-use of this information in repositories is of great benefit and explains why librarians have been influential in descriptive metadata standards for repositories.
- Much discussion has also occurred around whether it is better to migrate to new formats or develop ways to view old formats. Within these are permutations. Just considering migration, you could migrate everything to standard formats as you add them to the repository with the idea you can continue to migrate from standard format to standard format more easily. Alternatively, you could wait until someone wants to view them and migrate the old format as it is needed.
- Finally, there is the need for economic sustainability of the repository itself. Regardless of what hardware or software is used to operate the repository, the institution needs to commit to keeping digital material for the long haul. This is not substantially different from the commitment libraries have to maintaining paper-based material. A key difference at this point in the life of repositories is that the costs are not clear.

Why be brave?

Notwithstanding the unknowns, difficulties and questions raised in this paper, there is nothing like testing the waters to find out what you do and do not know, and what is easy and what is hard.

The impetus for repositories comes from research needs. The absence of a repository will not stop researchers creating, communicating, and publishing in digital formats. What will be lost is the research output.

The technology to support repositories already exists in open source and commercial forms. Though nascent, the technology is developing rapidly and is being used by a wide range of institutions.

Theoretical models, standards and protocols⁵ have been posited and developed for repositories. But it is only with implementation that flaws become obvious, simplifications are discovered and best practice is determined.

This is not the brave new world that Miranda in Shakespeare’s *Tempest* was thinking of when she said “O! Brave new world, that hath such people in’t”, nor the shocking world created by Aldous Huxley, but one developed by a partnership between repositories and researchers

⁵ See, for example, the Open Archival Open Information Reference Model:
<http://ssdoo.gsfc.nasa.gov/nost/wwwclassic/documents/pdf/CCSDS-650.0-B-1.pdf>

with small, cautious steps towards a way to keep and provide access to digital research material.

A final word

I couldn't let this opportunity go by without having a small plug for the project I am working on! The Australian Partnership for Sustainable Repositories aims to become a centre of excellence for the management of scholarly assets in digital format, particularly in the areas of access continuity and the sustainability of digital collections. APSR is supported by the Commonwealth Government's Backing Australia's Ability program⁶ through the Department of Education, Science and Training⁷. The APSR partners will develop and test ideas with real-life data and issues over three years. The partners are the Australian National University, the University of Queensland, the University of Sydney, the National Library of Australia and the Australian Partnership for Advanced Computing.

It is early days in the three-year program but I hope you will be interested to know that a key outcome is to share expertise, knowledge and skills not just amongst the partners but with other repository owners in Australia and overseas.

APSR's website is <http://www.apsr.edu.au> and I would be very happy for you to visit the site or ask questions of me.

Thank you for your time.

⁶ <http://backingaus.innovation.gov.au/>

⁷ <http://dest.gov.au/>