

# APSR Road Map 2007

PRELIMINARY DOCUMENT

Version 1.5

24 October 2006



Contact: Dr Adrian Burton  
Phone: +61 2 6125 6659  
Email: [adrian.burton@apsr.edu.au](mailto:adrian.burton@apsr.edu.au)

## Table of Contents

A. Looking Ahead .....	1
The Emerging Research Information Environment .....	1
Human Infrastructures .....	1
Cohesion and Collaboration in the Sector .....	1
Whither the Digital Collection?.....	2
B. Action.....	3
1. Infrastructure and Services for Digital Collections.....	3
2. Repository Integration and Interoperability .....	3
3. eResearch Facilitation.....	4
4. Research Reporting .....	4
C. Approaches .....	5

The purpose of this APSR Road Map 2007 (Preliminary Document) is to serve as a framework for more detailed planning. More comprehensive planning documents are envisaged for Q1 2007. The roadmap document has three parts:

- A. Looking Ahead
- B. Action
- C. Approaches

It begins with a scan of the horizon to map the important landmark needs of the foreseeable future (A). This then leads to sketching out important areas of activity (B) and appropriate approaches (C) to best respond to those needs.

## **A. Looking Ahead**

### **The Emerging Research Information Environment**

Substantial investments at the institutional, national and international levels have seen the emergence of a sophisticated information environment for Australian researchers. This information environment is rapidly improving and enabling new kinds of knowledge transfer in the Australian research and innovations sector. The information environment is enabled by an access infrastructure of networked data centres, repositories, online collections, portals, registries, collaboration and personal work environments, data manipulation services, access protocols, and so on. The access infrastructures in turn depend on a substrate of robust networking, computing, and storage facilities. The combination of these enabling elements ("cyberinfrastructure") needs to be coordinated and sustained into the future.

A sophisticated information environment in Australia will also need systemic planning & coordination of the whole lifecycle of research collections from creation through collaboration to curation.

There is an ongoing need for new services and infrastructure elements to bring the benefits of the research information environment to the broad community of Australian researchers.

### **Human Infrastructures**

As well as sophisticated technical capacities enabled by software, servers, networks and so on, the emerging research information environment requires sophisticated researchers, students, data scientists, technology support staff, librarians, archivists, and managers. This is the human element of the infrastructure for the new research information environment. The people who design, build, maintain and use these online collections are a critical element of the new research information environment.

There is an ongoing need for the provision of services, consultancies, training, workshops, conferences, forums, communications lists, publications, and online information to build the 'capability' of the Australian researchers to take advantage of the technical infrastructure. The research community needs the ability to make effective and efficient use of the infrastructure.

### **Cohesion and Collaboration in the Sector**

Significant cohesion within the sector will be necessary in order to create an information environment for "joined-up" research. The environment will need to provide sector-wide, information services and infrastructure. Broad engagement with and adoption in the sector will be a critical element of the success and relevance of such services.

A seamlessly interoperable national and international research information environment requires the adoption of open standards, common approaches, and shared definitions. Making this happen will require a high level of cohesion within the sector.

Such cohesion extends to the users of the infrastructure too. Research "communities" (cohesive groups of scholars) are an important factor for the data management of long-lived collections. These research communities are a critical element in the process of assigning value to scholarly communications and research data. Moreover, the standards for such data are often built around these communities.

There is an ongoing need for an inclusive approach that encourages cohesion and collaboration amongst institutional initiatives.

### **Whither the Digital Collection?**

The network of data centres, repositories, and online collections is an integral part of the new research information environment. This spectrum of different types of digital collections reflects a range of administrative and research dynamics.

Large institutions with a centralized aggregation of many varied collections typically use repository software packages such as DSpace, EPrints, Fez/Fedora, or VTLS/Fedora to provide a collection and information management framework. Data centres will generally provide more specialised solutions for reference data collections widely used by a research community. Other research groups will maintain their own tailor-made online collections. A holistic approach to this spectrum of information management needs and solutions is required to respond in a coordinated manner to the heterogeneous requirements of the Australian researchers.

The centralised "Institutional Repository" represents one point on the spectrum of solutions. The institutional approach aims to deliver sophisticated information management tools sustainably to large numbers of researchers and research groups. But the challenge for institutional repositories is to remain relevant and properly integrated into the overall academic enterprise. Repository development and management will need to focus not only on preservation and access, but also on increased applicability to scholarly work practices and integration with university information systems.

As the new research information environment matures into a sophisticated grid of services and infrastructure, a key characteristic of any digital collection will be its ability to integrate and interoperate with the wider research information environment.

Researchers are more and more dependent on data throughout the whole process of research and scholarship. Cyberinfrastructure for digital collections needs to support the changing nature of those needs throughout the life cycle of a research project; the research data itself is transformed, the management approaches change and stewardship responsibilities also change. A spectrum of cyberinfrastructure is emerging to satisfy that lifecycle of needs.

The overall opportunity is to bring data centres, repositories, and other online collections closer together so that services can be run across these forms of data management.

There is an ongoing need for common services and infrastructure for a wide range of digital collections of Australian research. The integration and interoperability of these collections is therefore a priority.

## B. Action

Based on the above broad needs, APSR's strategic agenda has identified four areas of activity:

1. Infrastructure & Services for Digital Collections
2. Repository interoperability & integration
3. eResearch Facilitation
4. Research Reporting

### 1. Infrastructure and Services for Digital Collections

This first (and key) area of activity aims to improve the infrastructures and services for managing digital collections. This theme stems from APSR's mission as part of Strategic Infrastructure Initiative (SII) to provide infrastructure and services to meet the ongoing requirements of the Australian research information environment.

From a national perspective, these infrastructure elements are often *shared* services (e.g., registries, aggregators, federations, brokers) for use across many repositories or digital collections. As shared services, they also contribute to an information environment beyond individual repositories, or collections, to include international frameworks for research collaboration. These services are also relevant to the National Collaborative Research Infrastructure Strategy (NCRIS) initiative.

These infrastructure and services should apply as broadly as possible across the spectrum of digital collections, data centres, and different repository platforms.

Within this theme, the following priority areas are proposed for 2007 and beyond: a collections service registry; digital preservation infrastructure and services; and collections statistics infrastructure. APSR also proposes to offer technical support services to satisfy the need for technical expertise within the emerging research information environment. This includes the establishment of the *Repository Technical Support Service* that will provide targeted technical support for Australian adopters of common open-source repository software solutions [DSpace, Fez, EPrints, etc.].

In this area of activity, APSR also has an ongoing agenda of information exchange and training events.

### 2. Repository Integration and Interoperability

This second area of activity is focused on improving how repositories interact with the wide range of information systems typically used for education and research.

This theme is based on the broad acknowledgment that if repositories are not sufficiently integrated into the everyday computing activities of the education and research community they will lack relevance. Similarly, most institutions have not integrated repositories with their key 'enterprise' information systems (such as student administration, finance, e-learning systems and so on). This could marginalise their institutional relevance and raise barriers to attracting ongoing policy and funding support.

To overcome these barriers, APSR seeks to improve the integration and interoperability of repositories. Improving the *integration* of repositories is addressed through the adoption of common, platform-independent, standards and protocols for packaging and exchanging data. This is aimed at improving the articulation points between the repository and the rest of the applications used in research, teaching and administration, for example, by seamlessly connecting word-processors, e-journal publishing software, repositories into a single unified workflow.

The other important aspect of integration is to combine repositories with 'disseminators' or collection presentation engines to provide richer graphical representations of collections and complex digital objects held within the repository.

As a broad-based heterogeneous partnership, APSR is uniquely positioned to undertake cross-platform development. Popular repository software, such as DSpace, Fedora, and EPrints are technically incompatible on many levels and APSR continues to focus instead on developing standards and protocols for seamlessly exchanging data between them.

Improving the *interoperability* of repositories is more complex and involves the adoption of common system architectures, standards and protocols that will guide the development of new repository platforms. It is envisaged that interoperability will be achieved, in the near-term, by using 'service-oriented' architectures and standards, and in the long-term, through the adoption of a common service interfaces (such as API's) for repositories.

This integration and interoperability agenda complements the work being done in the eFramework. In fact the APSR work is designed to offer real-life scenarios for the testing, implementation and instantiation of eFramework principles.

Within this theme, the following priority areas are proposed for the medium term future: repository integration with common research tools; standard repository content models and presentation definitions; an APSR repository interchange profile; and, defining standard software interfaces for repositories.

APSR intends to organise international forums to showcase Australian work in this area and to expose the Australian community to international approaches.

### 3. eResearch Facilitation

The third activity area is eResearch facilitation. APSR accepts that eResearch is not a well-defined concept, but nevertheless sees the need for the repository network to be aligned more closely with the ICT-enabled environments for research and scholarship. Development activities here consist of supporting researchers in their immediate research-driven data management needs while simultaneously establishing connections with longer-term archival environments, such as institutional repositories and large-scale data storage facilities, such as the APAC National Facility.

Under this general theme, APSR identifies the ongoing need for information infrastructure specialists to work with research groups to address generic eResearch infrastructure needs. This has been the APSR approach with developing infrastructure and data management tools for groups like PARADISEC (Pacific and Regional Archive for Digital Sources in Endangered Cultures) and individual field-based ethnographers.

APSR's work in this area tends to address the 'long tail' of the 80% of researchers whose more modestly-scaled eResearch initiatives are not easily accommodated by existing national support services in APAC and ASSDA. These areas typically include small-scale science groups and humanities scholars.

There is also an ongoing need to address the issues of sustainability of Australian eResearch data collections, including technical, financial, and administrative aspects.

APSR sees a continuing demand for its existing national services in this broad area: an *eResearch Technical Consultancy Service* to provide data sustainability and management advice to research groups; a technical training agenda; a 'guild' of eResearch technologists; and an ongoing series of major national and international 'forum' events.

### 4. Research Reporting

The final activity area is support for research reporting. This area of activity responds to emerging policy imperatives to identify, assess and reward research excellence. The planned RQF (Research Quality Framework) would provide a focus for activity in this area.

The *research reporting* theme has close links back to the *integration and interoperability* theme, because the challenge for repositories in research reporting is to integrate the archival and dissemination capabilities of the repository with the research reporting mechanisms and workflows of the rest of the university. This will require ongoing development of integration approaches, data definitions, and standard presentation templates for research portfolios.

Activity in this area aims to improve the generic potential for institutions and research groups to manage more efficiently and portray more effectively their research output.

As the Australian Government's policy becomes firmer in this area, we assume the need for some targeted APSR activity will emerge. In the short term we have pencilled in the probability of APSR-coordinated information events through 2007, as well as some responsive software development.

## C. Approaches

Having identified future needs and proposed activity to address those needs, it remains to identify what approaches will be most effective for building the infrastructure for the emerging research information environment.

The successful APSR experience has been based upon these approaches:

1. Collaboration
2. Leveraging off existing institutional resources
3. Working closely with communities of practice
4. Training, skills, & outreach
5. Cross-platform development
6. Emphasis on sustainability
7. International linkages
8. Focus on the immediate needs of key stakeholders

A number of these approaches derive from the nature of a partnership. We can expect considerable advantages from partnerships that can leverage off and encourage cohesion amongst institutional initiatives. These cohesive partnerships significantly lower the barriers to collaboration and encourage cross-platform approaches to information technology development based on open standards and open source software. This 'grass roots coalition' approach enables the partnership to work closely with communities of practice and focus on the immediate needs of the key stakeholders.

A broad coalition of institutions encourages these approaches, and without such a coalition many of these successful approaches will be impossible to emulate. Without the right approach, even the noblest agenda becomes difficult to bring to fruition.

The APSR approach has been an effective use of Commonwealth infrastructure funding, as APSR projects leverage off institutional initiatives. APSR builds upon existing institutional efforts and allocates APSR funds to collaborative activity with generic application to the national research infrastructure.

In response to the need for a sophisticated human infrastructure, a focus on training, skills, and outreach is critical. A broad alliance can certainly generate the motivation and the national reach required to pursue a significant national agenda in training and skills development.

The wider coalition is particularly important when developing an interoperable infrastructure for scholarly communication.

Another fundamental APSR approach is the ability to successfully move from testbeds to infrastructure and services. The testbed approach is an important way of bootstrapping experience and capability in emerging fields, but there has to be some path for these testbeds to mature into production. APSR has fostered expertise in the management of digital collections through its testbed programs and has subsequently been able to build services and infrastructure upon that expertise.

Ongoing structures are needed that encourage commitment, cohesion, and coordination within the sector.