

## **Digital Repositories and the Australian Higher Education Sector: Where to Next?**

Chris Blackall  
Development Portfolio Coordinator  
Australian Partnership for Sustainable Repositories

## Digital Repositories and the Australian Higher Education Sector: Where to Next?

### Abstract

In 2001 the Australian Government announced a raft of digital initiatives as part of 'Backing Australia's Ability – An Innovation Action Plan for the Future'. These were funded under the Systemic Infrastructure Initiative (SII) and have contributed significantly to building the capacity of information infrastructure in the Australian higher education sector. Coinciding with the SII, and specifically encouraged by programs funded by it, has been the arrival of digital (or institutional) repositories as key components of the higher education information landscape. Funding for the SII ends in December 2007, however, and no clear strategy has been proposed by the Government to build on its many achievements. Nevertheless, in December 2006, the Prime Minister's Science, Engineering and Innovation Council, Data for Science Working Group released a Report that strongly recommended that a "sustainable publicly funded national network of federated digital repositories" should be established for the purpose of managing and preserving research data. In this paper I argue why implementing the recommendations of the Data for Science Working Group provides the right strategy for moving forwards, and how refunding the SII program beyond 2007 will make it succeed.

### 1. Introduction

In this paper I argue that the Australian Government should continue its strategic investment in building the capacity of our information infrastructure, in the pioneering way that it did with the Systemic Infrastructure Initiative (SII) through 2003-7. More specifically, because the SII funding cycle will end in December 2007, the Government should clarify the future of its various repository and data management programs, and work toward a cohesive view of a national 'information infrastructure'. Indeed, it should move forward by implementing the strong recommendations of the Prime Minister's Science, Engineering and Innovation Council, Data for Science Working Group (PMSEIC, 2006).

Arguably, the failure to implement the recommendations PMSEIC Data for Science Working Group would be a mistake of major proportions. Conversely, implementing its recommendations will complement, if not multiply, the impact of the multimillion-dollar investments made over several years in the building the capacity of our national information infrastructure<sup>1</sup>.

---

<sup>1</sup> By information infrastructure, I mean hardware-based infrastructure (communications networks, data management facilities, computers etc.); software-based infrastructure (productivity software, repositories, e-learning systems,

I would be disappointed if this paper was simply read as yet another plea for Government handouts; rather, I want to establish that the SII is an example of good public policy that has provided an excellent return on the investment of public funds. It follows, that if SII is good public policy, and a good investment, then continuing it makes good sense beyond the end its funding cycle in 2007.

The aims of this paper then are to explain how the SII has been critical to establishing digital repositories in the Australian higher education sector, justify why funding of the SII should continue, and argue why implementing the recommendations of the PMSEIC Data for Science Working Group should provide the rationale for moving forwards.

I shall begin with a brief commentary on the report of the Data for Science Working Group and then move onto describing some of the achievements of the SII programs. I conclude the paper by linking the recommendations of the Data for Science Working Group to the activities of the SII, with a view to arguing that a refunded SII program provides the foundation for the strategic roadmap for moving forwards.

Before doing so, I want to emphasise that I am writing here as a practitioner who has worked in the higher education sector for seventeen years. During this time I have by witnessed at first-hand the impact of Government policies on my work as an information specialist, and consequently have come to appreciate the importance of good policy in the area. This paper therefore is a personal response to current information infrastructure policy environment and it does not represent the opinions of my current employer, the Australian Partnership for Sustainable Repositories.

## **2. Data for Science**

The Report of the Data for Science Working Group, published in December 2006, provides a clear and particularly strong set of recommendations to improve the management and preservation of research data in Australia.

The Working Group was formed to “advise on directions for managing the vast amounts of data now being generated from scientific research, observational projects, instruments, national and international collaborations, data mining and analysis.” These terms of reference restricted discussion to the needs of ‘science’, but the data management needs of science are almost identical to those of other cognate disciplines, such as the social sciences and the humanities, hence it has broad relevance.

The Data for Science Working Group comprised Australian scientists,

---

operating systems etc.); and, data-centric infrastructure (data sets, web sites, publications, digital objects etc.). Taken together, this hardware/software/data infrastructure underpins our scholarly communications system.

business people, public servants and information specialists who made eleven final recommendations. The first four recommendations are listed below (the others are less relevant to my argument, but no less important to the strengths of the overall Report).

- That Australia's government, science, research and business communities establish a nationally supported long-term strategic framework for scientific data management, including guiding principles, policies, best practices and infrastructure.
- That a high-level expert committee be established to provide the leadership role in progressing the formation of the long-term strategic framework for scientific data management.
- That the necessary policy and programmes be implemented with a view to establishing a sustainable publicly funded national network of federated digital repositories.
- That the expert committee consider the development of a strategic roadmap for the implementation and evolution of the national network of federated digital repositories.

These recommendations should be self-explanatory, but there are a number of concepts that require comment.

What, for example, is a "long-term strategic framework for scientific data management"? There are International precedents to follow here that establish the context. JISC (Joint Information Systems Committee), the peak information infrastructure body for higher education in the UK, for instance, has a well developed national framework for data management named the Information Environment (Awre, 2004). Similarly, in the USA, the National Science Foundation, Office for Cyberinfrastructure, is planning for a National Digital Data Framework (Greer, 2006). Clearly, a strategic framework for managing research data is not voodoo. We have the expertise in Australia to know what to do, and how to do it; but it would require planning, resources and sustained effort to bring it to a successful conclusion.

This leads me to comment on another string of words repeatedly used in the recommendations—"sustainable publicly funded national network of federated digital repositories". A mouthful to be sure, but for such a network to be "sustainable" and "publicly funded", it would need the Australian Government and States to fund it recurrently for an indefinite period. Neither the Australian Government, nor the States, has a track record of funding such infrastructure initiatives in this way; indeed, such initiatives are plagued by episodic and ad hoc funding arrangements. Clearly, a significant change of Government funding policy would be required before the recommendations could be fully carried out.

As Kevin Bradley (2005) has pointed out, the notion of 'sustainability' used by

repository specialists is not about throwing more and more money at data management and preservation problems—although nothing happens without money. He argues, instead, that for data management and preservation to be sustainable it must be valued as part of a broader set of cultural values, such as scientific norms, or the desire to preserve the cultural heritage of the nation. The PMSIEC Report illustrates this point with a diagram of a pyramid with four layers; ‘data’ at the bottom, followed by ‘information’, then ‘knowledge’, and topped by ‘wisdom’. In other words, to be sustainable, data must be discoverable, reusable and shared as an integral part of our social life and should contribute to the common good—to knowledge and wisdom perhaps. Data management and preservation, therefore, has to be integrated into our overall systems of scholarly communications for it to be truly sustainable. For example, it is now possible to electronically publish a research paper, store an archival version in a repository, and then link it directly to the relevant data set stored elsewhere. This level of integration is within reach, and the PMSIEC Report is to be applauded for taking into account the broader contexts of data sustainability.

The next recommendation that needs comment is the idea of “federated digital repositories”. Admittedly, we already have a national network of digital repositories that are connected through the Internet, but they are far from being ‘federated’. As I shall go on to explain later, ‘federating’ repositories, technically speaking, is not a straightforward exercise. We have a good idea how it might be done, but we currently have only the odd brick or two—not the whole wall (D. Rehak, Dodds, & Lannom, 2005; D. R. Rehak, 2004).

Finally, a question worth posing in response to the Report is this: why should repository ‘infrastructure’ matter to the workaday academic researcher or student? Recent studies about research data management in Australia that are cited in the Report indicate that many researchers manage their data poorly and are unsure where to seek assistance (McNamara & Buchhorn, 2006; Shadbolt, Knijff, Young, & Winton, 2006). These studies also report that researchers nowadays typically generate so much data that is well beyond their IT skills or facilities to manage effectively. Moreover, they report some lines of research generate so much data that human intervention in the data management process is neither practicable nor advised. Indeed, we are reaching the point where our data management problems that can only be solved systemically, using new methods and technologies at the infrastructure level. This message comes across loud and clear in the Report and has been echoed recently in major reports on the same topic published in the USA (ARL, 2006; National Science Board, 2005).

Once again, from the perspective of the workaday scholarly practitioner, I appreciate that talk of ‘infrastructure’ could be a profoundly uninteresting topic. After all, a sewerage system is infrastructure. Infrastructure is generally

designed not to draw attention; it just has to stay in the background and work without interruption. Similarly, information infrastructure only becomes visible when it doesn't work, or, in the case of broadband connectivity in Australia, when it is overpriced and slow by comparison to other regions. That some people find infrastructure uninteresting is understandable—finding it unimportant is a serious mistake.

### **3. Backing Australia's Ability**

At the beginning of this paper I argued that the Australian Government should continue its strategic investment in building the capacity of our national information infrastructures in the pioneering way it did in the Systemic Infrastructure Initiative. What was this Initiative, and why was it pioneering?

Systemic Infrastructure Initiative had its genesis in a long series of policy debates in Australia about the future of the national economy in a globalising world, culminating in a range of initiatives to promote science and innovation as drivers of economic prosperity. The critical point came in January, 2001, when the Australian Government launched 'Backing Australia's Ability – An Innovation Action Plan for the Future' (Howard, 2001). *Backing Australia's Ability* crystallized the Australian Government's research and innovation policy and demonstrated its commitment to it by earmarking \$246 million to be allocated over five years to fund a wide range of programs related to research infrastructure.

One of the central strategies of *Backing Australia's Ability* was to invest new resources in building the overall capacity of the Australian 'national innovation system' (OECD, 1997). The salient outcome from *Backing Australia's Ability* relevant here was the importance placed on 'information infrastructure' as part of the national innovation system—made real for everyone at the time by the dramatic growth of Internet use and the emergence of the World Wide Web as a socio-economic force to be reckoned with. On a practical level, this policy led to new funding streams for high performance computing facilities and networks, but also a range of information and communications technology based initiatives, including the SII.

The SII didn't just appear out of the blue however. In many respects it was a result of a concerted lobbying effort by research librarians to convince the Government to seize the opportunities afforded by the Internet for scholarly communications (McLean, 1999a, 1999b; Steele, 2000). They deserve our gratitude. The Government departments (and consultants) responsible for education, science and economic policy sympathetically responded to this lobbying and followed up with a raft of consultative committees and reports recommending funding for new infrastructure (HEBAC, 2002; HEIAC, 2002; HEIISC, 2003). Finally, the Australian Government decided to channel funds to the higher education sector to "upgrade the basic infrastructure of

universities” as it was written in the Minister’s press release.

### The Systemic Infrastructure Initiative

The SII officially kicked off in 2003 with DEST funding of \$12m for four major projects, collectively named FRODO (Federated Repositories of Online Digital Objects). Of the FRODO projects, APSR (Australian Partnership for Sustainable Repositories) and ARROW (Australian Research Repositories Online to the World) have emerged as the mainstays of repository development in Australia. It should be noted that the initial SII projects represented a very diverse set of interests, with some not directly related to repositories. I do not have to space to discuss them all here in detail, however, they are fully described online.<sup>2</sup>

ARROW [[www.arrow.edu.au](http://www.arrow.edu.au)] is an institutional consortium comprising: Monash University (lead institution), National Library of Australia, the University of New South Wales, and Swinburne University of Technology. University of Southern Queensland became a member in September 2006 through the PILIN Project. It has developed a federated discovery service in partnership with the National Library of Australia and collaborated with VTLS, a commercial library management software provider, to create the VITAL user-interface software for the Fedora repository. ARROW has established the ARROW Community to provide support for all 16 institutions that have installed the VITAL software to date. It also provides advice to all Australian higher education institutions on important aspects of repository implementation and management.

APSR [[www.apsr.edu.au](http://www.apsr.edu.au)] is also an institutional partnership comprising: Australian National University (lead institution), University of Sydney, University of Queensland, National Library of Australia and the Australian Partnership for Advanced Computing. It is a national centre of excellence for the management of scholarly assets in digital format and, more specifically, provides a focus on the critical issues of the access continuity and the sustainability of digital collections. It has build expertise in digital sustainability through the support of repository facilities within partner institutions. Partner institutions currently use DSpace and/or Fedora as their main repository platform. APSR has supported the development of the Fez user-interface for the Fedora repository (developed by the University of Queensland) and has an ambitious program in 2007 to improve the integration and interoperability of DSpace and Fedora. APSR also contributes to building the national capacity in this area by encouraging the development of skills and expertise, staging events, and providing coordination throughout the sector.

---

<sup>2</sup> For a complete listing of the SII funded projects, see the DEST web site: [http://www.dest.gov.au/sectors/research\\_sector/](http://www.dest.gov.au/sectors/research_sector/)

In 2007, APSR and ARROW are collaborating on a range of repository development projects and outreach events.

The FRODO projects were followed in 2005 with \$19.5m for MERRI (Managed Environment for Research Repository Infrastructure). This included RUBRIC (Regional Universities Building Research Infrastructure Collaboratively) that has developed repository infrastructure and capability across regional and smaller universities. Once again, please refer to the DEST website for detailed descriptions of these projects.

In August 2006, the Australian Government announced a further \$15m funding to the SII programs. The funds largely went to groups that had demonstrated progress in previous funding cycles, including 'extension' funding for the APSR and ARROW through to December 2007. Significantly, funding was also directed toward projects that are strategically important for the future of the national information infrastructure. These included funding for security and authentication infrastructure through the Meta Access Systems Management (MAMS) Project [[www.melcoe.mq.edu.au/projects/MAMS](http://www.melcoe.mq.edu.au/projects/MAMS)], and the ARCHER [[www.archer.edu.au](http://www.archer.edu.au)] project to develop robust software tools developed under the earlier DART [[dart.edu.au](http://dart.edu.au)] project and adapt them to fit the needs of NCRIS priority research capabilities (see next section) and more generally to assist researchers collecting and storing research data, and sharing it securely with collaborators within the emerging global e-research environment.

Despite this extra injection of funding, DEST has not identified a mechanism that could fund SII work beyond December 2007. Nevertheless, while SII activity tapers off in 2007, there are other infrastructure initiatives in the pipeline that may carry this work forward, albeit in quite different configurations. The following section briefly details these initiatives.

### **The National Collaborative Research Infrastructure Strategy**

In November 2006, the Australian Government announced that it was providing \$542m through to 2011 for a program named NCRIS (National Collaborative Research Infrastructure Strategy) [[www.ncris.dest.gov.au](http://www.ncris.dest.gov.au)]. Most of this money has been earmarked for bespoke 'e-science' research initiatives covering a set of 'capabilities' identified by a consultative process. There is a funding stream, however, named Platforms for Collaboration that has a budget of \$75m to cover the costs of shared information infrastructure. This includes the provision of data storage and management services for NCRIS projects, and it is assumed that the Platforms for Collaboration will draw on the existing expertise and resources of the SII projects, although this has not been confirmed.

### **The Research Quality Framework**

Another DEST program that will have a major impact on the future of repositories in Australia is the Research Quality Framework (RQF). In December 2006, the Minister for Education Science and Training, Hon Julie Bishop MP, announced more than \$87.3m in financial support for the implementation of the first cycle of the Research Quality Framework. Part of this announcement included \$25.5m for ASHER (Australian Scheme for Higher Education Repositories) programme. The role of ASHER will be to ensure that each Australian higher education institution has a repository in place for the RQF, and that it will be configured to play a role in storing research publications for external assessment, and presumably to help gather and communicate other research management data. Importantly, ASHER provides no opt out clause for universities; establishing a repository will be mandatory in order to participate in the RQF.

#### **4. Data for Science and the future of SII**

Through the rich diversity of SII projects and events, the Australian higher education sector has undoubtedly learnt a great deal about implementing, managing and enhancing institutional repositories. But perhaps the feature of the SII that strikes one most is that it has provided a framework for successful collaboration and partnering between universities—no mean achievement given their normally competitive postures. Through collaboration and partnerships, the SII stakeholders have enjoyed the benefits of knowledge transfers and a range of synergistic research and development opportunities with other universities and business. Thus the SII has resulted in a classic win-win for its stakeholders. In the final analysis, each university has derived tangible and intangible benefits from SII, the overall infrastructure capacity of the higher education sector has been improved, and the Australian public now enjoys access to a range of new information services and publications that it might otherwise not have had.

Based on an analysis of the achievements of the SII, I argue that funding for it should continue beyond 2007, however, this funding should be conditional on how each project can assist the implementation of the PMSEIC Data for Science Working Group recommendations. Furthermore—and this is a potentially controversial suggestion—the funds should be channeled through and coordinated by the “high-level expert committee” recommended by the Working Group. Australia could do well to create a new institution that would emulate the Joint Information Systems Committee functions with the UK higher education system.

Why do I think that a refunded SII should be aligned with the recommendations of the Data for Science Working Group? First and foremost, the Working Group fundamentally accepts the proposition that repositories serve an essential function in management of research data and need to be sustainably funded. This is explicitly stated in the recommendation

that “policy and programs be implemented with a view to establishing a sustainable publicly funded national network of federated digital repositories”.

I strongly agree with this recommendation, but the keyword to note here is “federated repositories”. As previously discussed, building a federated system is a completely different order of complexity than installing a standalone repository. It is a relatively trivial exercise for systems administrator to install, for example, DSpace or Fedora, but it is very difficult to establish even a limited degree of interoperability between them. By this I mean a limited level of interoperability might consist of exchanging data between DSpace or Fedora using a standard metadata profile—a method successfully pioneered by APSR—but this is still a far cry from the complete interoperability required for a network of federated repositories.

The missing elements required for federated repositories are the ‘shared’ interoperability infrastructures, such as service registries, metadata registries, persistent identifiers and other key ‘service-oriented’ software components described in the specialist literature (see Lavoie, Henry, & Dempsey, 2006). This interoperability ‘layer’ must be available to all repositories; thus, it must use common standards and be governed in ways that are sustainable and equitable. We already know that agreeing upon common information standards can be a painfully long and complicated process, but we have little experience of how to govern and sustainably fund these new federated services on a national and international basis.

One of the main reasons why the SII projects need to be refunded is that they have already started building the service-oriented software components needed of a network of federated repositories. The MAMS (Meta-Access Management System) project, for example, has developed world-class solutions for authenticating and authorizing users of federated repositories, and other key networked applications. Similarly, the ARROW PILIN (Persistent Identifiers Linking Infrastructures) project is investigating and building a federated persistent identifier ‘service’ for repositories. APSR is building a Collections Services Registry, and is collaborating with SII and NCRIS stakeholders to develop strategic roadmap for federated digital repositories that could form the basis of proposal for the “expert committee” recommended by the Working Group. There are many other examples I could site if space permitted. The point here is that the expertise to develop federated repositories already exists in the SII; thus, it would make good policy, if not economic, sense to build on this achievement.

Through the activities listed above, SII stakeholders have also been able to actively participate at an international level in the forums, conferences and so on that are essential to the development of a federated approach to repositories and e-learning systems. Some SII stakeholders, for instance, are deeply involved in the e-Framework (Wilson, Blinco, & Rehak, 2004), and are

providing much of the intellectual momentum for specifying the information architectures needed for federated repositories and e-learning systems.

Thanks to the SII activities, Australia is well placed to create a sustainable, publicly funded, national network of federated digital repositories—that is, just as long as the current levels of funding and support are continued. The key issue here is continuity. Given the world-class outcomes of many of the SII projects, it would be shame for the knowledge and technical capacity that has been accumulated was to be dissipated by flawed decision-making. It is critically important that the capacity established through the SII is retained and consolidated through ongoing funding beyond December 2007.

Finally, the Government should also take particular note of Working Group recommendation that data management expertise must become a “core skill” for researchers and that they receive data management training as part of their education. So too, the Government needs to take into account that if their strategy of making repositories an essential part of the higher education and research sector is to succeed, then repository managers, educators, developers and analysts will need to be trained, certified and be reasonably confident of their long-term career prospects. Continuity is key.

## 5. Conclusion

In this paper I have explained how the SII has been critical to establishing digital repositories in the Australian higher education sector, justified why funding of the SII should continue beyond December 2007, and argued why implementing the recommendations of the PMSEIC Data for Science Working Group should provide the rationale for going forward. By way of examples, I have established the reasons why the SII has been good public policy and, as such, why it deserves ongoing support.

In conclusion, I strongly recommend that the SII, or a new initiative very much like it, be refunded. More specifically, if refunded, it should focus on implementing the recommendations of the PMSEIC Data for Science Working Group. I am confident that if this strategy were carried out it would benefit the Australian higher education sector, and the public, for generations to come.

ARL. (2006). *To Stand the Test of Time: Long-Term Stewardship of Digital Data Sets in Science and Engineering*. Association of Research Libraries.

Awre, C. (2004). Seeing Is Believing: The JISC Information Environment Presentation Programme. *Ariadne*, 39.

Bradley, K. (2005). *APSR Sustainability Issues Discussion Paper*. Canberra: Australian Partnership for Sustainable Repositories.

Greer, C. (2006, 31 October). *Implementing the Strategic Vision for Digital Data*. Paper presented at the NSF Advisory Committee for

- Cyberinfrastructure, Washington, DC.
- HEBAC. (2002). *A Framework for an Australian Research and Education Network: The Final Report of the Systemic Infrastructure Initiative Higher Education Bandwidth Advisory Committee*. Canberra: Department of Education, Science and Training.  
<http://www.dest.gov.au/highered/research/pdf/aren.pdf>
- HEIIAC. (2002). *Research Information Infrastructure Framework for Australian Higher Education: The Final Report of the Higher Education Information Infrastructure Advisory Committee*. Canberra: Department of Education, Science and Training.  
<http://www.dest.gov.au/highered/otherpub/heiiac/report.pdf>
- HEIIISC. (2003). *Report of the Higher Education Information Infrastructure Implementation Steering Committee*. Canberra: Department of Education, Science and Technology.  
<http://www.dest.gov.au/highered/otherpub/heiiisc03/default.htm#intro>
- Howard, J. (2001). *Backing Australia's Ability: An Innovation Action Plan for the Future*. Canberra: Commonwealth of Australia.
- Lavoie, B., Henry, G., & Dempsey, L. (2006). A Service Framework for Libraries. *D-Lib Magazine*, 12(7/8).
- McLean, N. (1999a). *Investment in Information and Knowledge Infrastructure: A Strategic Framework for Australia's Research Enterprise*. Paper presented at the Australia's Information Future: Innovation and Knowledge Management for the 21st Century Workshop, Canberra, Australian National University, 3-4 March.
- McLean, N. (1999b). The Evolution of Information Resource Sharing Infrastructure: An Australian Perspective. *Library Hi Tech*, 17(3), 256-264.
- McNamara, P., & Buchhorn, M. (2006). *Sustainability Issues for Australian Research Data: The Report of the Australian E-Research Sustainability Survey Project*. Canberra: Australian Partnership for Sustainable Repositories.  
[http://www.apsr.edu.au/documents/APSR\\_Sustainability\\_Issues\\_Paper.pdf](http://www.apsr.edu.au/documents/APSR_Sustainability_Issues_Paper.pdf)
- National Science Board. (2005). *Long-Lived Digital Data Collections: Enabling Research and Education in the 21st Century*. National Science Foundation.
- OECD. (1997). *National Innovation Systems*. Paris: Organisation for Economic Co-operation and Development.  
[http://www.oecd.org/dsti/sti/s\\_t/inte/prod/nispub.pdf](http://www.oecd.org/dsti/sti/s_t/inte/prod/nispub.pdf)
- Rehak, D., Dodds, P., & Lannom, L. (2005, 10 May). *A Model and Infrastructure for Federated Learning Content Repositories*. Paper presented at the Interoperability of Web-Based Educational Systems Workshop, Volume 143 or CEUR Workshop Proceedings.
- Rehak, D. R. (2004). *Cordra: Adl's Federated Content Repository Model*.  
<http://www.lsal.cmu.edu/lsal/expertise/papers/presentations/usu2004/usu20040902.ppt>
- Shadbolt, A., Knijff, D. v. d., Young, E., & Winton, L. (2006). *Sustainable Paths for Data-Intensive Research Communities at the University of Melbourne: A Report for the Australian Partnership for Sustainable*

*Repositories*. Melbourne: Information Services, The University of Melbourne.

[http://eprints.infodiv.unimelb.edu.au/archive/00002731/01/APSR\\_Report\\_Final\\_Dec2006.pdf](http://eprints.infodiv.unimelb.edu.au/archive/00002731/01/APSR_Report_Final_Dec2006.pdf)

Steele, C. (2000). *Library and Information Infrastructure—International Strategic Initiatives: A Discussion Paper*. Coalition for Innovation in Scholarly Communication (CISC).

<http://online.anu.edu.au/caul/cisc/proj4final.doc>

Wilson, S., Blinco, K., & Rehak, D. (2004). *An E-Learning Framework: A Summary*. DEST (Australia), JISC-CETIS (UK), and Industry Canada.