APSR Sustainability Issues Discussion Paper

Kevin Bradley
Australian Partnership for Sustainable Repositories
National Library of Australia
28 January 2005

Purpose

The document is intended as a basis for discussion that will aid in the establishment of the specific tasks required to needed to meet APSR milestones. It is structured around a range of identified issues, includes a summary of the topic and reference to key texts.

Background

This paper was initially envisaged as a higher level document intended for the steering committee through the executive officer. However, the formation of the expert committee at the December steering committee meeting suggests that a task based document would be more appropriate for that forum, referred to them by the steering committee.

The final outcome of the process for which this discussion paper was written is a task based plan that will support the APSR aim of elucidating the critical issues of the access continuity and the sustainability of digital collections. It may be utilised to fulfil the requirement of an issues and strategy paper as outlined in Milestone 1 in the APSR Project Specification 12 February 2004.

APSR Sustainability Issues, Directions and Work Plan Executive Summary.

Digital Sustainability: Overview

The size and scale of growth of digital information is well documented, (SIMS 2003) and sustaining these digital records is a particular issue in the University and higher education sector. Digital sustainability recognises that the continuity of digital information goes well beyond basic storing and managing of data and is integrated into the lifecycle of the information object. It includes both technical, social and economic considerations. This paper includes the following aspects of sustainability.

1. **The sustainability of the raw data**; that is the retention of the byte-stream in its proper and logical order so that it can be delivered as required identically to how it was first deposited. Though all agree it is no simple technical feat, most digital preservation expertise simply cite the Cedars' statement regarding the IT expertise required to undertake this task. Cedars Guide to Digital Preservation Strategies states categorically that "preserving a byte-stream is a relatively routine IT operation." (Cedars 2002) However, it is worth noting that IT expertise identifies a considerable risk in the maintenance and refreshment of data, and only a well managed and designed approach to IT will ensure adequate results.

2. **The sustainability of access to meaning**; Sustainability requires not only that the data is retained, but that it can be rendered in future technical environments, and that the content can be comprehended in a manner commensurate with the creators' intentions and user expectations. This is characterised by the D-Space Federation in the statement that "in the long term content needs to be accessible by a wider audience, who may be using a wide variety of computing equipment. Additionally, file interchange formats and standards, and rendering software and hardware change over time … Referring again to OAIS terminology, we need to define mechanisms for obtaining appropriate Dissemination Information Packages from the Archival Information Package held within DSpace." (DSpace, Michael J. Bass et al. 2002)

3. **The economics of sustainability;** this includes the continued viable existence of the institutions that support the technology, and/or those that own, manage, or gain value from, the digital materials stored therein. The need to evaluate social and economic forces, as well as technical, is a necessary part of sustainable design in order to produce a whole solution. D-Space, for example, is not considered a preservation solution, but only useful in "enabling institutions with a *sustainable ability* to retain information assets and offer services upon them." (emphasis added) (DSpace, Michael J. Bass et al. 2002).

4. **The Organisational Structure of Digital Sustainability;** by examining the relationship between the three primary roles of digital repositories, the rights holder, the archive and the beneficiary, it is possible to typify five types of organisational structures. Lavoie (Lavoie 2003)asserts that by understanding

the structures it is possible to apply economic theory of incentives in such a way as to motivate sustainable resourcing and planning amongst the relevant stakeholders.

5. **The Economics of Participation**; As economics is largely a matter of incentives and inhibitors, the same approach suggested above can be applied to encourage users and depositors to participate in use of a repository. The approach would entail determining whether sufficient benefits were gained by participating in the use of the repository, and applying appropriate incentives where the benefits were insufficient.

6. **Sustainability and the Value of the Data;** that is the period of time in which the digitally encoded information is relevant and valuable to the given community. This is a particular aspect of economic value of the data and integral to sustainable planning. Some digitally encoded information will be valuable in perpetuity, while other information has a much shorter useful life before its value to the scholarly community is entirely superseded. There will be a need to distinguish between data for which there is a long term sustainable need, and data for which there is only a short term reliability requirement. Selection policies, and possibly deselection, play a part in sustainability.

7. **Tools, Software and Sustainability;** the economic incentives and technical solutions find common ground in the investigation of appropriate tools for implementing and sustaining digital repositories. The repository software itself is a tool, as are the various components designed to aid in operation, simplify processes, and automate and validate the harvesting of metadata. A necessary part of the outcomes of the partnership will be an audit of currently available tools and software, and a plan to encourage development of necessary software components.

APSR Sustainability Issues, Directions and Work Plan Discussion Paper

Sustainability Issues and Tasks

The tasks that face APSR are identifying, specifying, developing, testing and documenting those tools or processes that either contribute to providing access to meaningful digital content, or which define and support an institution's sustainable ability, developing guidelines that indicate the period for which data is valuable or defining the procedures necessary for exchanging data with an institution which has sustainable ability. All processes proposed to support access continuity and the sustainability of digital collections must themselves be assessed for sustainability.

The Technology of Sustainability

Repositories and Sustainability

Practical repositories are products of the technology of the day and are consequently as much at risk as the content they manage. The D-Space Federation again make this explicit when they state "It is an overt expectation that information assets managed by the D-Space system will outlive the current system, the current implementation of components within the architecture, as well as external implemented services that access and/or add value to the corpus."(DSpace, Michael J. Bass et al. 2002) In a critique of the Library of Congress' NDIIP Technical Architecture, Gladney (Gladney 2004) claims that NDIIP Technical Architecture fails to distinguish between digital repositories and digital preservation, the former of which he asserts is well developed in comparison to the latter. In focussing on *sustainable* repositories, APSR is addressing an important issue that has not been properly considered elsewhere.

Sustainability approaches must consequently consider, at their outset and through their design and execution, future implementations that may not support or be supported by current standards and technologies. It is clear that no repository will provide a complete solution to the problems of sustainability, though neither will any solution succeed that does not incorporate a viable, well designed, digital repository.

The DPC's (Digital Preservation Coalition) report on institutional repositories includes recommendations on the overall repository structure, suggesting a model that includes modular components and a layer architecture with appropriate interfaces. The intention is to allow any layer to be changed without the process requiring major restructuring of the repository as a whole. In the report, Wheatley states that "choosing a high level design can simplify this inevitable change and hopefully prevent any data loss through the process … The dangers of not addressing this issue are all too apparent." (Wheatley 2004). Involvement in, and influence on the design of repository software can support these aims.

The tasks associated with this:

1. Test to determine whether partner's data structure and architecture allows for future migration to alternate repositories.
2. Develop general guidance for ensuring this requirement is incorporated in all sustainable repositories.
3. Document and make available.

Byte-Stream (or Bit-stream) Sustainability

The Open Archival Information System (OAIS) Reference Model(OAIS 2002), a conceptual framework for an archival system dedicated to preserving and maintaining access to digital information over the long term, distinguishes the byte stream from the object itself. The distinction enables the different aspects of the issue of preservation to be considered separately. The preservation of the byte stream, that is the ability to store and retrieve the data in its logical form, is the base level sustainability issue and in the discussion of digital sustainability is most often considered to be an issue that is the province of the IT systems administrator. This assessment requires review and, in the in the case of each partner institution, a clear allocation of roles and responsibilities which can be used as general guidelines.

The issue of byte stream sustainability has frequently been deemed unproblematic, however the case for this has not been adequately tested. Cedars (CURL Exemplars in Digital ARchives) made an unreferenced declaration in 1998 that preservation of the byte stream is routine (DSpace, Michael J. Bass et al. 2002), and this statement has been a fundamental assumption in such projects as University of Michigan's CAMiLEON (CAMiLEON 2003), Edinburgh University Library Digital Preservation Project (EUL 2004), the DCC (DCC 2004) and many others. In their paper "Emulation, Preservation and Abstraction" Holdsworth and Wheatley state that Cedars explicitly recognises "that the indefinite preservation of a byte-stream is technically straightforward" (Holdsworth and Wheatley 2001). All these references refer only to the Cedars' statement. However, most data management professionals will not make the same confident statement and suggest that careful procedures and practices need to be in place to ensure that the sustained viability of the byte stream is achieved. There is a need to test the assumption against professional practice, and to document appropriate technologies and procedures for byte stream preservation. This will also have the effect of making explicit the expectations that are associated with long term sustainability and which may not necessarily be apparent to the various responsible areas.

The issues associated with this task

1. Document risks to byte stream preservation, both general and specific, through discussion with APAC, NLA and University partners.
2. Determine the respective roles and responsibilities.
3. Identify any existing appropriate data integrity protocols for both technology and procedures (e.g. AS/NZS ISO/IEC 17799:2001, Information technology—Code of practice, though not entirely applicable to the issue of data sustainability, is an example of the type of data standard which might be considered appropriate).

4. Determine whether those protocols have applicability to the issue of sustainability.
5. Define an appropriate level of certitude for the particular data.
6. Assess whether the risks are ameliorated by the procedures.
7. Test procedures against business model (to ensure testing practices themselves are not unsustainable).
8. Document procedures and protocols.

Sustainability of Access

APSR has an "overall focus on the critical issues of the access continuity and the sustainability of digital collections". Access by the designated community is a critical measure of sustainability; access to an authentic rendering of the informational content in both the short and long term must be the primary aim of any digital continuity program. This is stated succinctly in the UNESCO Guidelines for the Preservation of Digital Heritage: "Digital materials cannot be said to be preserved if access is lost."(Webb 2003)

Access issues may either be those associated with the ability to access and render the content of a digital object, or may be those associated with finding, retrieving, using and re-using a digital object. Access, in terms of finding and retrieving are discussed below under the heading of Persistent Identification and Access and Retrieval, and Sustainability (below): sustainable access, in terms of presenting material such that there is no technical impediment to gaining the intended meaning from the intellectual content of the digital objects, is the primary aim of digital preservation.

An agreed level of sustainable access by the repositories is required as guidance for the designated community of users. The OAIS model suggests that there is a responsibility on the archive to deliver an archival information package that contains all that is "needed to make the content data object understandable to the designated community." (OAIS 2002) page 2-5. This requires that the repository managers "must understand the Knowledge Base of its Designated Community to understand the minimum Representation Information that must be maintained." (OAIS 2002) page 2-4. It may be quite reasonable to assume that a designated community will have particular expertise and technology in providing access to digital objects encoded in a specific way, especially for specialist users, however, more general users of digital materials may not have particular format expertise nor the need to maintain it. It is also very likely that the knowledge base of any designated community will change with time, so that the subject expertise may remain but the technical expertise associated with particular digital delivery technologies may change. It will be necessary to determine in which cases the ongoing support of access is the responsibility of the repository or that of the designated community. Decisions made with regard to designated communities and their knowledge base will inform other areas of sustainability such as format support (see Standard Formats and Format Support, below).

Issues associated with this task:

1. Determine methodology and/or guidelines for defining designated communities with respect to particular academic communities, technical expertise, and collections.
2. Provide guidelines for defining knowledge base of designated communities.
3. Determine roles and responsibilities for keeping this information current.
4. Use this debate to inform the discussion on access and format support.

Persistent Identification, Access and Retrieval, and Sustainability

The ability to find, retrieve and access information is a critical component of sustainability, the ability to cite an object with an unchanging identification is a necessary part of academic study. Various technologies and approaches exist to identify, locate and retrieve data and various schemas for encoding it. Most approaches can be grouped into two types, those that implement a single universal identification system, and those implement a locally unique number attached to a unique institutional identification number(NLA 2002; NLA 2004). The former approach is difficult to gain sufficient agreement on, the latter depends on the sustainability of the institution (see Organisational Sustainability below).

Even assuming the universities are sustainable organisations, there is likely to be frequent movement of data location for many reasons, technical, logistical and economic, even where a single institutional repository has been established. It is not only necessary to resolve the location of the digital object to the persistent identifier, it is also necessary to ensure that the required item remains locatable.

Some universities require access to a number of subject, discipline or faculty specific repositories, preferably, and increasingly, through a single search function. The ability to search a campus wide range of data bases can be managed in three ways; centralised data store, metadata repository (eg OAI Harvested) with distributed data, or direct access (middleware) to distributed data and metadata. Each have different PI implementation requirements, but it is very likely that, within any particular University campus, there will be a mixture of all three approaches.

The tasks with this issue are:

1. Investigate and implement an appropriate persistent identification approach.
2. Debate the issues of resolver versus universal schemes in the higher education sector.
3. Document, distribute and recommend.
4. Investigate the need for a resolver service
5. Ensure a reliable update and mapping service
6. Ensure that approaches work across many repository and access models.
7. Document and recommend.

Preservation Metadata

Digital Preservation, described as the “sustained, direct action” (Webb 2003) necessary to retain digital continuity in the medium and long term, is the processes and procedures put in place to ensure that authentic digitally encoded information can

be reliably accessed for as long as is required. Digital preservation incorporates an essential and significant subset of the issues raised by digital sustainability. The most useful tool in this process is considered to be the use of appropriate preservation metadata. Much of the digital archiving community considers that the preservation metadata elements are adequately described in the document "OCLC/RLG Metadata Framework to Support the Preservation of Digital Objects"(PREMIS 2003). Though this document is under review, APSR and the NLA are members of the review committee and have immediate access to any changes.

The OCLC/RLG Preservation Metadata group supplied the following definition of preservation metadata:

"Preservation metadata is the information necessary to carry out, document, and evaluate the processes that support the long-term retention and accessibility of digital materials. Preservation metadata includes information ranging from a description of the hardware/software environment needed to render a particular class of digital object to a record of the migration of a digital object through successive formats over time, and is an essential component of most digital archiving strategies."(PREMIS 2003)

Preservation metadata is not a technical solution, but rather a mechanism by which future solutions and procedures can be implemented. Any action intended to sustain digital objects must be informed by appropriate and relevant data about the objects. There will undoubtedly be new technical solutions to sustainability issues and preservation metadata is designed to take advantage of such solutions.

Human action and non technical decisions have been identified by APSR partners as constituting a major risk to digital collection sustainability. Decisions which may cut funding, support, or resources for important digital collections can have a more critical and immediate impact than encroaching technological obsolescence. Issues associated with this are considered under business models and economic decisions (below), however, it is worth noting that preservation metadata is intended to enable sustainable long term preservation of digital materials regardless of where they have been held. If encoded in an interoperable form preservation metadata allow for rapid and efficient transfer to alternate storage repositories. Preservation metadata is the key to the survival of the content, and as such is a tool for the management of risk to digital collections due to hostile economic or managerial decisions.

Authenticity and informational integrity is a general concern to users of any records, however the inherent mutability of digital records has made this issue a major concern in digital sustainability, and various methods and measures have been posited to manage this risk (InterPARES 2004). The degree to which the risk of corruption of records is an issue to be managed will be dependent on the types of content as well as its technical form. Bureaucratic and governmental records require a different type of risk management to heritage or technical documents (Gladney 2004), however, all records require some sort of preservation and change audit trail to document the basic validity and provenance of the digital record. Preservation metadata also plays a major role in this process.

APSR has initiated a series of tasks to address the issue of preservation metadata which includes the following:

1. Document the digital material in selected APSR collections in the test-bed partners, including storage systems and existing metadata. Material that is planned to be added can be included.
2. Engage with the research communities in the partners to understand the short-term and long-term value placed on the material and so determine the type and level of preservation metadata required.
3. Identify and assess the areas of likely failure in the sustainability of the material to which preservation metadata would be appropriate.
4. Compare the areas of risk against the OCLC/RLG Framework to identify the preservation metadata elements that will ameliorate the risks and identify any requirements overlooked in the Framework.
5. Compare the elements against current practice in the test-beds to identify any gaps and propose strategies to bridge any gaps. This will include metadata that needs to be captured, workflows that will assist the research communities to work efficiently, tools that assist capture of preservation metadata, and the design of repositories to store and manage the metadata.

The calculation and documentation of existing and potential collection material outlined above, is a necessary part of determining the appropriate level and type of metadata. The collected data however, may also form part of a larger risk assessment project on all sustainability aspects of APSR test bed repositories, data storage technologies and business models.

On completion of the above tasks, the resultant issues will be:

1. The testing of the support for preservation metadata currently available in partner and potential repository systems and test-bed datasets.
2. The incorporation of the requisite preservation metadata into the data models of the various test bed repositories.
3. The identification of the point in the data life cycle where the various aspects of preservation metadata needs to be acquired.
4. Development and/or implementation of tools for the acquisition and creation of preservation metadata.
5. Determination of the level of interoperability in preservation metadata applicable to distributed test bed repositories.
6. Documentation of preservation metadata recommendations and technologies.

Technology Watch

As all digital information, including the repository itself, is subject to format obsolescence and the vagaries of market or technological changes in software, operating systems and hardware, there will be a need to be aware of the changes that might be occurring within the technical market so as to plan and undertake any consequent tasks in a timely manner. Virtually all technical decisions regarding the long term sustainability of digital objects are made on a format-specific basis initiated by a change, however there is no impetus for format owners and suppliers to provide that information. Manual technology watches are considered both time consuming

and haphazard in their results. Various schemes are being developed and systems such as Harvard University Library's global digital format registry (Abrams 2003), or the UK's National Archive's project PRONOM (TNA 2004) when linked with an automatic preservation notification system like DSTC's PANIC (DSTC 2004) presents the type possibility of the type of notification system that may be integrated into sustainable repository planning and design. Such systems are still in the development stage and APSR can benefit by monitoring developments over the next 12 months and then determining if any such system can be integrated into the test bed repositories.

The list generated through the documentation of digital material in the partner collections, described in Preservation Metadata (above), will be a necessary prerequisite to involvement in, or use of, a format registry.

1. monitor developments over the next 12 months
2. determine if any such system can be integrated into the test bed repositories
3. consolidate format list from Preservation Metadata (above)

Tools, Software and Sustainability.

There is a technical, workflow and economic requirement for various software tools to undertake the general and specialised tasks. Many of the necessary software tools are, or at least may be, available or under development. A survey and audit of available software tools will be a necessary part of the APSR project. The repository itself is a software tool which is vital to the task of maintaining data for organisations with sustainable ability, and Wheatley (2004) lists six open source institutional repositories and assesses them for their digital preservation value. A continued watch on the development of these systems is required, as well as the addition of any new initiatives in this area.

In developing a set of sustainability requirements and principles, and formulating necessary workflows and practices in response to them, the need for various tools to support them will become obvious. This will develop through the life of the APSR project. There are, however, many workflow requirements that are already obvious and for which tools exist. The National Library of New Zealand's Preservation metadata extraction tool is a response to a requirement that should be evaluated for the APSR project, as should the National Archives of Australia's XML conversion system XENA. There also exist the potential, within the life of the project, to influence the design of particular tools, or to commission or undertake the development of tools specifically for the project.

Tasks associated with these issues.

1. Identify needs, practices and work flows for which software tools would be required.
2. Document requirements.
3. Survey and audit available tools.
4. Assess whether the potential exists to influence tool design.

5. Determine whether development of tools to meet special requirements could or should be undertaken.
6. Make all developments available.
7. Document.

Certitude[1] (authenticity), Significant Properties and Measurement of Processes.

If formats are to be rendered in standard formats, or if migration of the file format is necessary, then there needs to be some way of specifying the level and measuring the performance of the outcome of the translation. This necessitates some way of describing or measuring the characteristics of the particular object. This approach of identifying the essential information characteristics of an object has been labelled "significant properties"(Cedars 2001; Hedstrom and Lee 2002) or "essence" (Heslop, Davis et al. 2002; Wilson 2003) Outside of the archives community it has not rendered much useful, practical processes suitable for sustainability, however, little or no work has been done on university based collections. The approach may be beneficial for some classes of objects, for example those text documents where colour scheme, embedded images, page layout or internal hyperlinks can be discounted as valuable, or scientific data sets where the output can be clearly specified, and those components successfully sustained. Lavoie considers it likely that the increase in activities associated with sustainability "will encourage the development of a consensus in terms of what 'successful preservation' means in regard to particular classes of information resources"(Lavoie 2004).

APSR partners would need to:

1. Discuss and determine measures for "successful preservation".
2. Determine whether this approach is in any way practical and whether appropriate business models could be developed to support it.
3. Specify classes of objects where such action would be valuable.
4. Encourage the definition of 'successful preservation'.
5. Consider and propose alternate paths to ensure accuracy, certitude and authenticity in the case of other formats of preservation procedures.

Standard Storage Encoding

As a means of reducing the impact of format obsolescence some archives and repositories advocate the rendering of all files in a structured format that is less dependent on current operating systems and programs. The National Archives of Australia (NAA) has developed a software system known as XENA which is used to convert specific file types and files to XML. The "essence" of the document type is identified theoretically, the designated content is encoded in XML and the specific

[1] A meaning of certitude is "something that is assured or unfailing". It is deployed here in place of the way authenticity has been used in the digital debate. Instead of authenticity of a record in the archives sense, which may be critical to important legal and statutory documents, it is intended to imply the idea of accuracy and authenticity. Most users of academic record need to be as certain as possible that the document represents what it says it does, and that the process of encoding and archiving has not altered its representational value. A user of a legal document needs to be sure that the document is authentic.

document is re-rendered using browsers specific to the current systems ensuring that the essence is not lost or compromised. This approach requires significant development of rendering software, at ingest of each specific file format and also continued development of access and rendering software in the longer term for each specific document type. However, this may be necessary in the long term for documents that have not been so encoded, and the structured nature of the encoded documents reduces the resources required to undertake this when compared with currently usable and often proprietary formats.

There is a risk in this strategy that some essential characteristics may be lost, and as a consequence this approach lends itself only to particular types of documents, though the concept is more difficult to implement in other, more complex cases. For example, a simple email message without attachment may be stored in a structured form; the resultant potential loss of formatting is unlikely to have any effect on the informational content as the text is sufficient to convey meaning. A more complex document where the structuring is an integral part of the information will require a higher level of metadata to ensure ongoing ability to render the information as required, and less, or no, freedom to render the document in an alternate form without loss. The approach is most beneficial where a large number of similar documents are ingested into a repository such as a bureaucratic, government or business archive, but is labour intensive for more general repositories which have less control over format type.

An encoding and structuring process that addresses the issue of very long term sustainability is proposed by Gladney and Lorie (Gladney and Lorie 2003), and Gladney (Gladney 2003). Briefly, it is suggested that the digital object be encoded in a Universal Virtual Computer (UVC) code (Lorie 2002), a virtual computer of relatively simple specification which can be fully specified and tested. Objects so encoded are described by Gladney as Trustworthy Digital Objects (TDO). The TDO is, in the current form, packaged with XML and Unicode, including relevant metadata and specifications. Compilers exist which allow the document to be rendered in the current environment, and may be developed relatively simply, according to Gladney, for future machine languages. It is claimed that the approach is equally appropriate to complex digital objects (such as executables or programs) as it is to simple objects. The disadvantage, assuming the process is successful, is that it is probably more cost effective to manage the information through available technologies for all but the most complex objects, a fact acknowledged in “Preserving Digital Records” (Gladney 2004)

It is likely that the use of the standardised storage encodings has significant cost attached, both now and in the future, and are probably only justified for particular approaches and formats intended for long term sustainability.

Tasks arising from this issue include:

1. Review notions of essence of significant properties
2. Examine available encoding systems
3. Determine if any intended content would benefit from this approach.
4. Consider whether the Trusted Digital Object or Universal Virtual Computer is relevant to the APSR partners’ sustainability intentions.

Standard Formats and Format Support

There is an extraordinarily large range of potential formats and versions which may be submitted for sustainable storage in a repository. Pragmatically, it is probably unavoidable for any repository institution to make a decision about supporting a finite range of those formats and either defining the selected formats as preferred, or rendering the submitted data in specified formats. DSpace suggests a three level support scheme, while stipulating that it is the responsibility of the host institution to "determine the exact meaning of each support level, after careful consideration of costs and requirements" (DSpace 2004). The three levels of the scheme as implemented by MIT are labelled *Supported*, in which the format is recognised and future access and usability is guaranteed, *Known*, in which the format is recognised, can be retrieved and it is hoped will be accessible in the future, and *Unsupported* in which usability is not guaranteed and only the retrieval of the bit-stream is possible.

Supporting particular formats and relegating the others to *known* and *unsupported* impacts on both the depositor and the repository differently according to the type of resource. Some aspects of sustainability are easily implemented, while other are less well defined. The difference between a sustainable format, and a distributable format is often very marked, and frequently both are required to maintain a sustainable repository. The types of standard formats may be grouped under the following categories: images, sound recordings, textual content, and video, which are the basic components for more complex objects such as reports, web sites and multimedia works.

The criteria that must be met in order to sustain image or sound is fairly well known. It should be a linear format, that is a complete encoding of the original image or sound, and must have no lossy compression applied. For the purposes of long term sustainability the Technical Committee of the International Association of Sound and Audio Visual Archives (IASA-TC) recommends .wav in its BWF form (Bradley 2004) (EBU Tech 3285), and National Information Standards Organization formulated its recommendations around TIFF Revision 6.0, Final—June 3, 1992 for images (NISO 2004). In order to make such processes practical APSR may need to get involved in recommending specific technology which produces the information in the required format.

Moving image presents a slightly more complex set of problems. Though developments in an encoding wrapper for digitally encoded video are proceeding (Material eXchange Format), and standards for video encoding which will support the linear criteria specified above are also underway, the market need and available technology are not supporting the requirement extensively. No institution is making unequivocal statement on the sustainability of video formats and it may be necessary for APSR to develop some aspirational guidance, along with current interim solutions, even though those solutions may compromise long term sustainability. Currently most archival facilities are setting a minimum standard of MPEG-2, 4:2:2 profile while awaiting developments (MPEG 2004).

Text presents another set of issues, surprisingly complex for the long used means of communication it is. The Library of Congress have sought to cluster the array of significant textual characteristics under three integrity factors: integrity of document structure and navigation; integrity of layout, font, and other design features; and integrity of rendering for mathematics, chemical formulae, diagrams, etc. (Library of Congress 2004). They list a range of suitable formats, from generic XML, through to PDF/A. Also noted for special types of material is the News Industry Text Format (NITF), an XML standard designed to structure independent news articles and so allow exchange(IPTC 2004). Moving from purely textual objects to primarily textual digital objects containing some mixed data, the NISO standard NISO Z39.18-200X Scientific and Technical Reports – Preparation, Presentation and Preservation (NISO 2004), addresses the issue of exchange and preservation for news reports by proposing a standard XML schema.

Complex multimedia digital objects are probably the most likely to encounter problems due to format obsolescence and are also the least likely to be adaptable to an automated standard conversion process. As most of the learning object may be incorporated into a complex meta-object, there is the added incentive to be able to re-purpose the content so that it may be reused, or restructured as a response to technological changes, or possible intellectual or teaching changes. As the conversion of such material is, by necessity, hand crafted and probably unsustainable as a long term strategy, it would be beneficial to provide guidance to the depositors. This would enable the initial documents to be structured in such a way as to reduce the load on programmers. Sufficient expertise exists amongst the APSR partners to provide such advice.

APSR partners will have to determine as a sustainability issue;

1. If they are going to support particular formats,
2. What are the technical and economic implications of that support
3. Whether open source or proprietary formats are preferred
4. The extent of the support (years and technological)
5. Which formats are to be supported

There will be a need to

6. Survey the developing standards which may support the work.
7. develop guidelines for depositors
8. ensure the deposit requirement do not produce too great an impediment to lodgement.

Risk and Risk Management

The Australian Risk Management Standard (Australian Standard AS/NZS 4360:1999), specifies 5 steps in risk management: Establish the context; Identify the risks; Analyse the risks; Evaluate the risks; and finally treat the risks.

Establishing the context is the process of defining what it is that is at risk and what are the aims or objectives of, in this case, the repository's owners. This may well be the establishment of a listing of the content of a repository, or of the formats are identified as a major category for risk, a listing of formats. The processes of identifying the risk is describing what may happen, while analysing the risk is the process of determining

what are the likelihood and consequence of that happening. Evaluating the risk is the process of awarding a risk rating based on the previous categories. Treating the risk is the primary aim of APSR, facilitated by focusing on the critical issues of the access continuity and the sustainability of digital collections.

A common aspect of risk management methodologies is that the middle three steps, identify the risks, analyse the risks and evaluate the risks, are more a creative process than objective, necessary though, to reveal context specific risks. Co-Author of the Australian standard, Dale Cooper suggests that checklists and other prescriptive methods "block identification of the risks that go beyond the experience encapsulated in the list" (Cooper 2004). Likewise, he states assessment of the likelihood is a process of informed, though still subjective, judgement. The primary recommendation to reduce the risk of poor risk identification and assessment is to use structured workshops, or failing that, individual structured interviews of highly skilled or informed experts in the area. This is critically necessary in determining the risk to sustainability of digital objects as the most likely failure mechanisms are not well known.

The INFORM (INvestigation of FOrmats based on Risk Management) Methodology (Stanescu 2004), provides an approach intended to reduce the likelihood of inadequate judgement by recommending a panel of reviewers to analyse the risks. The larger and more informed the group, the greater the likelihood of accurate results. The intention of the INFORM methodology is to associate risks with particular formats and make that information, as a measure of preservation action, publicly available and reviewable.

The National Library of Australia Risk strategy (unpublished) is similar in the identification of the risks and categories of risk, but relied on a workshop of in-house and consultant expertise to analyse and determine risks. The NLA methodology was useful in identifying the time at which action should be taken rather than when the consequences of a risk might transpire. The intention of the NLA risk methodology is to manage the risks in an in-house repository.

The need for a risk assessment is critical to determining the action necessary for the APSR project. It is implied in the section on The Technology of Sustainability, though quite clearly in the section on Preservation Metadata (above) and is explicit in the section on the Economics of Sustainability (below).

The tasks associated with this issue:

1. Compare the INFORM, NLA and any other appropriate risk assessment approaches.
2. Based on these, develop a risk assessment approach appropriate to the higher education sector.
3. Apply the risk assessment to information gathered under other areas of concern (eg format list in Preservation Metadata and Technology Watch)
4. Document and use to inform APSR aims and priorities.

The Economics of Sustainability.

One of the maxims that accompanies the formation of digital collections is that they will not survive the type of benign neglect that is appropriate for the preservation of paper based materials. In other words digital materials require persistent attention, and no one-off action will sustain them for extended periods of time. An alternate view is presented by Henry Gladney (Gladney 2004), who claims that data will survive with minimal back up strategies, but that the cost of providing meaningful access to the content through the use of developing digital archaeology skills will be cripplingly high, and not necessarily successful. Though Gladney consequently defines digital preservation action as an economic issue, it is one where he sees immediate action as necessary and accordingly advocates investment in the present to ensure access in the future. Lavoie notes that with the advent of digital authoring and distribution technologies, our developing capability to manage and sustain such information is being outstripped by our ability to produce it. He argues that, along with the necessary technological infrastructure for sustainability "must come the development of the associated *economic infrastructure*" (Lavoie 2004).

Understood in these terms, digital preservation is as much an economic issue as a technical one. The requirements of ongoing sustainability demand at their base a source of reliable funding, necessary to ensure that the constant, albeit potentially low level, support for the sustainability of the digital content and its supporting repositories, technologies and systems can be maintained for as long as it is required. Such constant funding is not at all typical of the university based communities that build these digital collections, many of which tend to be grant funded on an episodic basis. There is therefore a need to develop costing models for sustainability of digital materials according to the specific requirements of the various classes of content, access and sustainability.

However, though it makes economic sense that costing models precede business models (Lavoie 2003), not all economic planning is centred around monetary considerations, nor are all repository models those of a grants generated cluster of digital information. There is therefore a requirement to determine the other non-monetary economic, social and attitudinal aspects that may hinder or contribute to the sustainability of digital content, including incentives and disincentives to participation and use, the expected life and research value of the content, and the usage model, value and profile of the repository. There will probably be a need to develop indicators that codify the relative viability of the repository based on broad economic characteristics.

Costing Models

Ownership and Rights

The economics of the establishment of digital repositories has been a part of discussion since their inception. The resultant costing models have largely clustered around notions of ownership and copyright, purchase costs and access, and division of income where publishers own rights and repositories want to provide access. Failure to resolve this economic issue can cripple an archive of published materials and this concern was expressed with regard to the Yale project. "When we determined in the Yale project that we could not identify a point at which a publisher could say automatically that it was no longer in the publisher's business interest to maintain the

archive and, therefore, the archival institution could open the archive to the public, we were faced with the prospect of building an archive that might remain dark or at best "dim" for years, perhaps decades. The archival institution would not have the opportunity to recover costs by charging users in any way. Nevertheless, the conviction was that the archive should still be built now, as there were too many risks associated with postponing the archive's creation until the publisher no longer wished to protect it for business reasons." (Lavoie 2004) Measures can and have been taken with regard to managing these published resources and this is useful from a commercial point of view (see for example JSTOR www.jstor.org/jstor/ or e-prints http://www.eprints.org/).

However, there is an often expressed concern amongst academics who state there is some ambiguity in rights in material generated and published whilst in University employment. If ownership and rights are not to be a disincentive to deposit, it may be prudent to clarify these issues. This may be as simple as citing university policy or legal opinion, or may include a statement of intent. The issue of access and rights will be informed by the work of MAMS (Meta Access Management System http://www.melcoe.mq.edu.au/projects/MAMS/)

Tasks arising from this issue include:

1. Investigate rights and ownership in deposited materials from university academics.
2. If necessary develop appropriate policy, statement or advice
3. Document and publicise.

Sustainability, Resources and Costs

Costing models will be an integral part of sustaining digital repositories in the medium to long term. It is only with sound estimates of the ongoing costs that bids for sustainable resources can be made. A number of broad level costing approaches are available and virtually all start with the caveat that as technology suitable for digital sustainability develops, so too will the costing models. There is consequently a limit to the accuracy of long term prediction regarding long term sustainability. The UNESCO guidelines suggest starting with a short-to-medium term plan, probably five years, and review within that period developments for the next stage(Webb 2003)

The National Library of New Zealand Digital Library Development Review (Ross 2003), bases its costing model on a model being developed for NASA by US company SGT inc (http://www.sgt-inc.com/) which was developed to help in "estimating the life cycle costs of future ESE data service providers and supporting systems" (Hunolt 2003) p.6. Ross adapted the categories to suit a digital library model, identifying various areas where costs would be incurred and combining all the attributed costs to show how this could produce an item level and an annual digital library cost figure (Ross 2003) p.45-7]. However, in the end Ross states "the estimates vary and cost models are not very consistent" (ibid:51), and concludes with a detailed recommendation regarding the NLNZ's economic modelling of the collection of digital objects.

The ERPA Guidance Cost Orientation Tool (ERPANET 2003) rather unsurprisingly takes a similar approach, listing categories and sub categories where costs may be incurred along with guidance on how to collect cost information and apply the tool. In the application of such costing models the types of preservation service should be known, for though quantity, quality and size of the digital materials ingested has an impact on scale, the cost of long term digital sustainability correlates more to the range of digital services offered (Chapman 2003).

Seeing a need for simple yet scalable metrics, Sannett's model for costing proposes categories such as costs for preserving, costs for use, costs for acquisition, and costs associated with institutional use. Each category is divided into detailed parts, often under the broad accounting headings of capital, direct and indirect costs. The intention is to provide a framework that allows customisation by users by the adding or deleting specific components. "It is clear that the soft-funding scenario of the past and present is not sufficient to fund present and projected activities to preserve electronic materials. The issue of institutional sustainability in preservation must be discussed and resolved. Who will pay for the costs involved with acquiring, preserving, and accessing the materials? A number of strategies have been proposed, some of which are continued institutional support, fee for use, fee from the author, fee from the publisher, and legislative support." (Sanett 2003).

Significant work has been done in investigating costing models and efficient procedures in the digitisation and ingest of particular media which is of relevance to the university based repository. Realistic estimates of time and resources needed to acquire and ingest digital content is required. PRESTO has undertaken cost comparisons on video and audio materials (Wright 2002; Wright 2004), IASA's latest publication includes critical time estimates(Bradley 2004), and many image digitisation projects have long been costed eg (Puglia 1999).

Tasks arising from this issue include those identified in the costing models:

1. Investigate costing models
2. develop a university-sector-appropriate approach to cost collection and economic modelling with regard to digital materials.
3. identify business needs and scope of preservation (policy and risk questions)
4. identify types of digital objects that will be created and need to be preserved
5. identify how long they need to be preserved
6. identify consequences for people and organisation
7. identify methods, standards, tools, technologies, systems to be used.
8. apply costing models to the described categories
9. determine accuracy
10. Review and document

The Effect of Funding Models on Sustainability

Due to the nature of research and research funding, the provision of funds and other resources to projects which generate digital material requiring repository storage is often episodic. This frequently results in the initial or set up costs for such projects being funded, but not the ongoing sustainability costs. Very significant collections

are often built this way, but their continued viability may subsequently be in jeopardy. Modification or supplementation should be made to the grant or project based funding model, or alternate strategies will need to be developed, to support the sustainability of such materials.

The APSR could make representation to the funding bodies, the universities and any other responsible bodies to ensure that appropriate steps are taken to deposit such materials in appropriate, sustainable repositories.

1. Identify appropriate bodies
2. Develop a recommendation for the funding bodies
3. Make appropriate representation

Organisational Sustainability

The complex technical infrastructure that supports digital sustainability, the dependency on continued funding, and the likelihood that digital data will not survive extended periods of neglect means that digital repositories need stable technical support as well as resources. It follows that digital repositories are dependent on the ongoing viability of the sponsoring organisation. This is a requirement not only for the technical structure but the selection of an appropriate persistent identifier scheme, and the ability to manage a resolver service that continues to locate digital objects intended for long term use are also dependent on the sustainability of the institution or organisation. If a repository is at risk because of the viability of the organisational structure that supports it, then structure of the repository, the interoperability of the metadata and data formats, and the ability to seamlessly migrate to alternate repositories is an integral part of any plans to manage sustainability. In identifying these points as an area of concern, APSR are addressing what appears to be a critical issue that, with the exception of the DPC paper (Wheatley 2004), has received scant regard within the current literature.

The sustainability of the repository within a given university might be measured by its relationship to the organisational structure. A repository may be a policy initiative of the university, of the faculty or department, an individual academic on behalf of a faculty, or be an independent repository housed within a section of the university not necessarily funded by that section. A policy initiative of the University is more likely to receive ongoing funding than and independent repository. It could be stated that the economic risk to the sustainability of the repository is inversely proportional to its position within the university hierarchy.

It is most likely that the owners of a repository at risk of failure due to loss of organisational support would seek to deposit the content of that repository with an appropriate alternative facility. It follows that the higher the potential risk to the sustainability of the repository, the greater the need to adhere to interoperable and exchange standards (see Repository structure, interoperability and exchange, above). It might be argued that, under these circumstances, a decentralised model is more at risk than centralised model (for the alternate view, see Incentives and Motivations, below).

Tasks Associated with these issues.

1. Explore the issue of organisational sustainability
2. Develop risk indicators for particular organisation structures
3. Provide arguments for, and guidelines on, appropriate repository structure, interoperability and exchange which is applicable at the initial stages.
4. Compare the risks to centralised and decentralised models.
5. Document and publicise

Business Models

The Organisational Structure of Digital Sustainability

Lavoie (Lavoie 2003) makes the point that the formation of repositories and associated sustainability activities will produce limited types of organisational structure to support it. The type of structure is dependent on three key roles; the rights holder, who holds the intellectual property rights to the digital materials; the archive, which is the area responsible for providing services to ensure the long-term preservation and accessibility of digital materials; and the beneficiary, those that benefits from the long-term retention of digital materials. The relationship of the entities, or roles, within that structure tends to lead towards particular types of natural incentives to undertake sustainability or preservation action, and precludes those incentives, or introduces disincentives in others. Lavoie argues that by understanding the economic models we are in a position to encourage preservation action by applying an incentive where such incentives do not naturally occur.

The obvious assumption behind Lavoie's work is that the motivation for the creation and use of a repository transpires due to a particular business need, and the continued sustainability of that repository is dependent, at least in part, on their being a continuing business need. The important conclusion is that, though there may be a business need, and even a cultural or intellectual imperative to sustain a digital repository, their may not be sufficient incentive for any individual entity to undertake the tasks that ensure sustainability due to the nature of the relationship between roles and the consequent organisational model. In such cases it is economically necessary to apply corrective economic policy measures and provide appropriate incentives.

Tasks which may occur from this issue (see also Lavoie 2003:41-44)

1. Consider the types of organisational structures that accompany the formation and maintenance of digital repositories.
2. Understand how the organisational structure of digital preservation activities comes about, and how it evolves over time.
3. Determine where failures are likely to occur
4. Develop appropriate corrective policies (incentives).
5. Match corrective policy measures with the types of market failures for which they are the most effective.
6. Document and publicise

Incentives and Motivations

Lavoie's model is primarily concerned with motivating particular entities to undertake appropriate action or supply sufficient resources to ensure the preservation and subsequent sustainability of the repository. However, as "economics are fundamentally about incentives" (Lavoie 2003)p.i), the concept may be applied to other aspects of sustainable repositories, particularly incentives for participation and use.

A fundamental economic, as well as technical, distinction between repositories is whether they reside as a centrally organised and funded initiative, or are distributed throughout the institution. It is generally assumed in a distributed digital repository model that, as responsibility for each separate repository probably resides with an individual, the motivation to participate in lodging material, entering appropriate metadata, and taking responsibility for the ongoing maintenance of the repository is very high. However, where responsibility for the repository is vested in a more centralised entity, there may be a need to provide incentives for participation, especially where any value added work, such as metadata entry, is required. If the repository is designed to benefit researchers and scholars other than the depositor, and if there is not sufficient direct benefit to the depositor, then it is likely that the supply of material deposited will eventually falter. It will be necessary to develop tools to bring the relationship between benefit and cost to as near as neutral as possible, even then economic thinking suggests that this will not ultimately be successful unless there is a perceived positive benefit to outweigh the work in submitting material.

This can be achieved in three ways; the repository assume all costs and labour associated with acquiring the material; the repository or institution can associate significant benefits with depositing material; or the appropriate authority can mandate deposit.

PARADISEC, for example, undertakes an upload and metadata service for contributors to the repository. The digitisation and storage of such materials is also perceived as a positive benefit to depositors, as the issue of format obsolescence which has been well publicised in both audio and video materials and concerns collection owners. And finally there is a positive benefit to participation in the archive through more sophisticated use of the material. Without all three incentives it is unlikely that there would be sufficient deposit materials, even though the repository serves a demonstrable good.

Researchers at the University of Rochester who deposit scholarly content in the institutional digital repository are provided with a searchable researchers page. The researchers page is the access point for the scholarly content, and increases the visibility of the researcher within the user community by simplifying access to the content and providing a citable reference to the item. It also cuts across the community model developed by D-Space which has been identified in this context as an impediment to depositing content. The depositor gain benefit through use of the repository (Foster and Gibbons 2005).

The intention of many of the Australian universities to mandate the lodgement of theses for higher degrees in digital form is an example of mandated deposit.

Similarly, users of the repositories will only be motivated by significant benefits. Consequently access tools must be designed which encourage use of the material, or at least are not so cumbersome that the disincentive they introduce does not outweigh the benefits of using the material, and material should be selected which is of the highest possible value.

Tasks that arise out of these issues:

1. Determine whether this type of economic modelling is appropriate to APSR partners.
2. Assess and document the inhibitors that may discourage participation
3. Assess the tools and workflows that may reduce or neutralise disincentives
4. Develop tools to determine whether the benefits outweigh the inhibitors.

Where inhibitors exists

1. Develop possible incentives that might encourage participation.
2. Model them to determine economic suitability
3. Collect and maintain records of use and other factors which might help to inform this process.

Sustainability and the Value of the Data

Not all data will, or should, be sustained in perpetuity. Though costs are a function of many variables, not least the range of archival services, the archival period of retention is a significant factor (Lavoie 2003)p.7), therefore it is essential to plan retention of digital materials for the appropriate period. Certain datasets or learning objects may only have intellectual, teaching or research value for a short period of time, possibly shorter than the life of the target sustainable repository. If sustainability is the primary aim of the repository it may be valid to exclude such materials, or to provide a limited type of service. Other materials may be considered valuable for a medium period of time, in which case the time between ingest and access may not be so great as to have incurred the problems caused by format obsolescence and impaired access. It may be possible to attach a reviewable lifetime rating to identified digital objects, and so reduce estimates of costs on objects so designated. The decision to delete after a given period can be reviewed, or the material can be assessed and deselected. It is worth considering though, that the cost of expert review may well exceed the cost benefits of deselection in the individual case.

For the purposes of economic planning there is little or no difference between an object retained for a long archival period and perpetuity. In this case the critical economic issue is to ensure that the material acquired is of high academic significance to future researchers. It is very unlikely that a collection of low research significance will survive in the long term as resources will always be allocated to high value materials first.

The need to sustain the digitally published scholarly output of researchers and academic staff is well understood, and will be necessary to sustain the model of intellectual progress that has underpinned academia. E-Prints, JSTOR and Digital Thesis Online are examples that address the acquisition and selection of such materials for potentially sustainable digital repositories, and these and other approaches will need to supported to ensure that the full range of material is captured and retained appropriately. However, for many disciplines the potential exists to generate supporting material which may in itself be more valuable in the long term than the products of its academic evaluation. Datasets from some areas of scientific and medical research, field recordings from linguistics, anthropology, (ethno)-musicology and history, images from all areas of study, are all by-products of research that have the potential to eclipse the significance of the original research for which they were carried out. This primary source material constitutes a major source for future researchers. This generates a selection policy quandary in deciding what materials should be sustained. Prof Anthony Seeger (UCLA) made explicit this incongruity at the SIMS (ethno)-musicological research conference when he asked: "What is more valuable in the long run, researchers' theories or the by products of research, like recordings and other collections? How many important theoretical articles published between 1900 and 1920 influence your current work? Wax cylinders recorded during that period are extremely valuable to both their original communities and contemporary researchers. Ironically, the by-products of our research may be more significant than our soon dated theoretical insights." (Seeger 2004).

Digital repositories must seek out, acquire and incorporate such materials. Traditionally, this information has been acquired by archives and libraries long after publication of the resources, and the paper records retrieved and made available. If created or stored in digital form it is imperative that such information be ingested into a sustainable repository much earlier in the informational life cycle than was the case with paper based records. It may be necessary to allow access constraints related to authorship and ownership over a given period of time to provide sufficient incentive to deposit such materials.

Tasks arising from this issue include:

1. Develop criteria for selection guidelines
2. Develop and document procedures that encourage the lodgement of digital materials, both primary and secondary where appropriate.
3. Develop selection policy which can be used to determine appropriate useful life of collection materials.
4. Determine minimum life span for which each item or class of material will be accepted (i.e. guidelines to exclude materials of transient value)
5. Develop review mechanisms and procedures.

Recommendations

1. That the APSR Steering Committee consider and approve the framework that underpins APSR Sustainability Issues, Directions and Work Plan Discussion Paper.
2. That the APSR Steering Committee refer the APSR Sustainability Issues, Directions and Work Plan Discussion Paper to the Expert Committee for detailed discussion of the suggested tasks and their relative priority.
3. That the APSR Expert Committee develop an action plan using a table (see appendix A) to indicate priority, timeline and/or allocation of responsibility.
4. That the APSR Expert Committee refer the action plan to the steering committee for consideration and discussion of budgetary and prioritisation issues.

References

Abrams, S. (2003). Towards a global digital format registry. World Library and Information Congress: 69th IFLA General Conference and Council, Berlin http://www.ifla.org/IV/ifla69/papers/128e-Abrams_Seaman.pdf Accessed January 2005.

Bradley, K. (2004). IASA TC 04 Guidelines on the Production and Preservation of Digital Audio Objects. Denmark, International Association of Sound and AudioVisual Archives.

CAMiLEON (2003). "Creative Archiving at Michigan & Leeds: Emulating the Old on the New."http://www.si.umich.edu/CAMILEON/ Accessed January 2005

Cedars (2001). Cedars Test Site (Phase 2) Significant Properties. Curl Exemplars in Digital Archives. http://www.leeds.ac.uk/cedars/testSitePhase2/sigProp1.html Accessed January 2005

Cedars (2002). Cedars Guide to: Digital Preservation Strategies. http://www.leeds.ac.uk/cedars/guideto/dpstrategies/dpstrategies.html Accessed January 2005

Chapman, S. (2003). "Counting the Costs of Digital Preservation: Is Repository Storage Affordable? Article No. 178, 2003-05-07." Journal of Digital Information **Volume**(Volume 4 Issue 2). http://jodi.ecs.soton.ac.uk/Articles/v04/i02/Chapman/ Accessed January 2005

Cooper, D. F. (2004). Tutorial Notes: The Australian and New Zealand Standard on Risk Management, AS/NZS 4360:2004. Broadleaf Capital International Pty Ltd. http://www.broadleaf.com.au/tutorials/Tut_Standard.pdf Accessed January 2005

DCC (2004). Digital Curation Centre. http://www.dcc.ac.uk/index.html Accessed January 2005

DSpace (2004). System Documentation: Functional Overview. MIT & Hewlett Packard. http://dspace.org/technology/system-docs/functional.html Accessed January 2005

DSpace, Michael J. Bass, David Stuve, R. Tansley, Margret Branschofsky, Peter Breton 1, P. Carmichael, Bill Cattey, Dan Chudnov and J. Ng (2002). DSpace-A Sustainable Solution for Institutional Digital Asset Services Spanning the Information Asset Value Chain: Ingest, Manage, Preserve, Disseminate. Internal Reference Specification, Technology & Architecture Version 2002-03-01. Masssachusetts Institute of Technology, Hewlett-Packard Company. http://libraries.mit.edu/dspace-mit/technology/architecture.pdf Accessed January 2005

DSTC (2004). PANIC (Preservation webservices Architecture for Newmedia and Interactive Collections). Distributed Systems Technology Centre. http://metadata.net/newmedia/ Accessed January 2005

ERPANET (2003). erpa guidance Cost Orientation Tool. Electronic Resource Preservation and Access Network. http://www.erpanet.org/guidance/docs/ERPANETCostingTool.pdf Accessed January 2005

EUL (2004). Digital Preservation at Edinburgh University Library. http://www.lib.ed.ac.uk/sites/digpres/ Accessed January 2005

Foster, N. F. and S. Gibbons (2005). "Understanding Faculty to Improve Content Recruitment for Institutional Repositories." D-Lib Magazine 11(1).http://www.dlib.org/dlib/january05/foster/01foster.html Accessed

Gladney, H. M. (2003). "Trustworthy 100 year Old Digital Object: Evidence Even After Every Witness is Dead." **Volume**.http://eprints.erpanet.org/archive/00000008/ Accessed January 2005

Gladney, H. M. (2004). "Persepctives on Trustworthy Information." Digital Document Quarterly **Volume**(Volume 3, Number 4 (4Q2004)).http://home.pacbell.net/hgladney/ddq_3_4.htm Accessed 5 Jan 2005.

Gladney, H. M. (2004). "Persepctives on Trustworthy Information." Digital Document Quarterly **Volume**(4 (4Q2004)).http://home.pacbell.net/hgladney/ddq_3_4.htm Accessed

Gladney, H. M. (2004). "Preserving Digital Records: A Method Guided by Scientific Philosophy." **Volume**.http://eprints.erpanet.org/archive/00000071/ Accessed 6 January 2005

Gladney, H. M. and R. Lorie (2003). "Trustworthy 100 year Old Digital Object, Durable encoding for when its too late to ask." **Volume**: Accessed

Hedstrom, M. and C. A. Lee (2002). "Significant Properties of Digital Objects: Definitions, Applications, Implications." DLM-Forum **Volume**.http://www.dlmforum2002.org/download/margaret_hedstrom.PDF Accessed January 2005

Heslop, H., S. Davis and A. Wilson (2002). An Approach to the Preservation of Digital Records. Canberra, National Archives of Australia. www.naa.gov.au/recordkeeping/er/digital_preservation/summary.html Accessed January 2005

Holdsworth, D. and P. Wheatley (2001). "Emulation, Preservation and Abstraction." **Volume**.http://129.11.152.25/CAMiLEON/dh/ep5.html Accessed January 2005

Hunolt, G. (2003). NASA SEEDS Requirements / LOS & Cost Model, Draft Working Paper. Publisher. http://lennier.gsfc.nasa.gov/seeds/LOS_020116.pdf Accessed

InterPARES (2004). The Long-term Preservation of Authentic Electronic Records: Findings of the InterPARES Project. International Research on Permanent Authentic Records in Electronic Systems. http://www.interpares.org/book/index.cfm Accessed 6 January 2005

IPTC (2004). News Industry Text Format (NITF). The International Press Telecommunications Council. http://www.nitf.org/ Accessed January 2005

Lavoie, B. F. (2003). "The Incentives to Preserve Digital Materials: Roles, Scenarios, and Economic Decision-Making." **Volume**.http://www.oclc.org/research/projects/digipres/incentives-dp.pdf Accessed January 2005.

Lavoie, B. F. (2004). "Of Mice and Memory: Economically Sustainable Preservation for the Twenty-first Century, In." Access in the Future Tense **Volume**.http://www.clir.org/pubs/abstract/pub126abst.htm Accessed January 2005.

Library of Congress (2004). Text, Quality and Functionality Factors. http://www.digitalpreservation.gov/formats/content/text_quality.shtml Accessed January 2005

Lorie, R. (2002). "The UVC, a Method for Preserving Digital Documents." **Volume**(Series Number 4).http://www.kb.nl/hrd/dd/dd_onderzoek/reports/4-uvc.pdf Accessed January 2005

MPEG, M. P. E. G. (2004). Information technology -- Generic coding of moving pictures and associated audio information MPEG-2, 4:2:2 Profile. http://www.mpeg.org/ Accessed January 2005

NISO (2004). Data Dictionary-Technical Metadata for Digital Still Images NISO Z39.87-2002 AIIM 20-2002. City, National Information Standards Organization and AIIM International. http://www.niso.org/standards/resources/Z39_87_trial_use.pdf Accessed 24 January 2005

NISO (2004). NISO Z39.18 - 200x Scientific and Technical Reports - Preparation, Presentation and Preservation. City. http://www.niso.org/standards/resources/Z39-18-200x.pdf?CFID=274272&CFTOKEN=21242365 Accessed January 2005

> This standard outlines the elements, organization and design of scientific and technical reports, including guidance for uniform presentation of front and back matter, text, and visual and tabular matter in print and digital formats, as well as recommendations for multi-media reports.

NLA (2002). Managing Web Resources for Persistent Access. National Library of Australia. http://www.nla.gov.au/guidelines/persistence.html Accessed January 2005

NLA (2004). Persistent Identifiers. National Library of Australia. http://www.nla.gov.au/nlaredirect/pi.html Accessed January 2005

OAIS (2002). Reference Model for an Open Archival Information System (OAIS). City, Management Council of the Consultative Committee for Space Data Systems (CCSDS). http://www.ccsds.org/documents/650x0b1.pdf Accessed January 2005

PREMIS (2003). Preservation Metadata for Digital Materials. PREservation Metadata Implementation Strategies, OCLC. http://www.oclc.org/research/projects/pmwg/background.htm Accessed November 2004

Puglia, S. (1999). "The Costs of Digital Imaging Projects." RLG DigiNews 3(5).http://www.rlg.org/preserv/diginews/diginews3-5.html#feature Accessed

Ross, S. (2003). Digital Library Development Review. National Library of New Zealand Te Puna Māatauranga o Aotearoa. http://www.natlib.govt.nz/files/ross_report.pdf Accessed January 2005

Sanett, S. (2003). "The Cost to Preserve Authentic Electronic Records in Perpetuity: Comparing Costs across Cost Models and Cost Frameworks." RLG DigiNews 7(4).http://www.rlg.org/preserv/diginews/v7_n4_feature2.html Accessed

Seeger, A. (2004). Handout Notes at SIMS. Symposium of the International Musicological Society, Melbourne Accessed

SIMS (2003). How Much Information? Berkeley, CA, School of Information Studies. http://www.sims.berkeley.edu/research/projects/how-much-info-2003/printable_report.pdf Accessed 18 January

Stanescu, A. (2004). "Assessing the Durability of Formats in a Digital Preservation Environment: The INFORM Methodology." D-Lib Magazine 11(10).http://www.dlib.org/dlib/november04/stanescu/11stanescu.html Accessed

TNA (2004). PRONOM: The File Format registry. The National Archives UK. http://www.nationalarchives.gov.uk/pronom/ Accessed January 2005

Webb, C. (2003). UNESCO Guidelines for the Preservation of Digital Heritage. City, UNESCO. http://unesdoc.unesco.org/images/0013/001300/130071e.pdf Accessed 29 November 2004

Wheatley, P. (2004). Institutional Repositories in the context of Digital Preservation. Digital Preservation Coalition. Accessed 19 January

Wheatley, P. (2004). Institutional Repositories in the context of Digital Preservation. Digital Preservation Coalition. Accessed 19 January 2005

Wilson, A. (2003). Access Across Time: How the NAA preserves Digital Records. http://www.erpanet.org/events/2003/rome/presentations/Wilson_presentation.pdf Accessed January 2005

Wright, R. (2002). PRESTO - Preservation Technologies for European Broadcast Archives. PRESTO. http://presto.joanneum.ac.at/Public/D7_2.pdf Accessed January 2005

Wright, R. (2004). Various Articles available at PRESTO outcomes page. PRESTO. http://presto.joanneum.ac.at/outcomes.asp Accessed January 2005

Appendix A List of Issues and Tasks

Issues and Tasks	Partners	Deliverable	Time Frame
Repositories and Sustainability			
1. Test to determine whether partner's data structure and architecture allows for future migration to alternate repositories. 2. Develop general guidance for ensuring this requirement is incorporated in all sustainable repositories. 3. Document and make available.			
Byte-Stream (or Bit-stream) Sustainability			
1. Document risks to byte stream preservation, both general and specific, through discussion with APAC, NLA and University partners. 2. Determine the respective roles and responsibilities. 3. Identify any existing appropriate data integrity protocols for both technology and procedures (e.g. AS/NZS ISO/IEC 17799:2001, Information technology—Code of practice, though not entirely applicable to the issue of data sustainability, is an example of the type of data standard which might be considered appropriate). 4. Determine whether those protocols have applicability to the issue of sustainability. 5. Define an appropriate level of certitude for the particular data. 6. Assess whether the risks are ameliorated by the procedures. 7. Test procedures against business model (to ensure testing practices themselves are not unsustainable). 8. Document procedures and protocols.			
Sustainability of Access			
1. Determine methodology and/or guidelines for defining designated communities with respect to particular academic communities, technical expertise, and collections.			

2. Provide guidelines for defining knowledge base of designated communities. 3. Determine roles and responsibilities for keeping this information current. 4. Use this debate to inform the discussion on access and format support.			
Persistent Identification, Access and Retrieval, and Sustainability			
1. Investigate and implement an appropriate persistent identification approach. 2. Debate the issues of resolver versus universal schemes in the higher education sector. 3. Document, distribute and recommend. 4. Investigate the need for a resolver service 5. Ensure a reliable update and mapping service 6. Ensure that approaches work across many repository and access models. 7. Document and recommend.			
Preservation Metadata stage 1			
1. Document the digital material in selected APSR collections in the test-bed partners, including storage systems and existing metadata. Material that is planned to be added can be included. 2. Engage with the research communities in the partners to understand the short-term and long-term value placed on the material. 3. Identify and assess the areas of likely failure in the sustainability of the material to which preservation metadata would be appropriate. 4. Compare the areas of risk against the OCLC/RLG Framework to identify the preservation metadata elements that will ameliorate the risks and identify any requirements overlooked in the Framework. 5. Compare the elements against current practice in the test-beds to identify any gaps and propose strategies to bridge any gaps. This will include metadata that needs to be captured, workflows that will assist the research communities to work efficiently, tools that assist capture of preservation metadata, and the design of repositories to store and manage the metadata.			
Preservation Metadata stage 2			

1. The testing of the support for preservation metadata currently available in partner and potential repository systems and test-bed datasets. 2. The incorporation of the requisite preservation metadata into the data models of the various test bed repositories. 3. The identification of the point in the data life cycle where the various aspects of preservation metadata needs to be acquired. 4. Development and/or implementation of tools for the acquisition and creation of preservation metadata. 5. Determination of the level of interoperability in preservation metadata applicable to distributed test bed repositories. 6. Documentation of preservation metadata recommendations and technologies.			
Technology Watch			
1. monitor developments over the next 12 months 2. determine if any such system can be integrated into the test bed repositories 3. consolidate format list from Preservation Metadata (above)			
Tools, Software and Sustainability.			
1. Identify needs, practices and work flows for which software tools would be required. 2. Document requirements. 3. Survey and audit available tools. 4. Assess whether the potential exists to influence tools design. 5. Determine whether development of special tools to meet requirements could or should be undertaken. 6. Make all developments available 7. Document.			
Certitude (authenticity), Significant Properties and Measurement of Processes.			
1. Discuss and determine measures for “successful preservation”. 2. Determine whether this approach is in any way practical and whether appropriate business models could be developed to support it.			

3. Specify classes of objects where such action would be valuable. 4. Encourage the definition of 'successful preservation'. 5. Consider and propose alternate paths to ensure accuracy, certitude and authenticity in the case of other formats of preservation procedures.			
Standard Storage Encoding			
1. Review notions of essence of significant properties 2. Examine available encoding systems 3. Determine if any intended content would benefit from this approach. 4. Consider whether the trusted Digital object or Universal Virtual Computer is relevant to the APSR partners' sustainability intentions.			
Standard Formats and Format Support			
APSR partners will have to determine as a sustainability issue; 1. If they are going to support particular formats, 2. What are the technical and economic implications of that support 3. Whether open source or proprietary formats are preferred 4. The extent of the support (years and technological) 5. Which formats are to be supported There will be a need to 6. Survey the developing standards which may support the work. 7. develop guidelines for depositors 8. ensure the deposit requirement do not produce to great an impediment to lodgement.			
Risk and Risk Management			
1. Compare the INFORM, NLA and any other appropriate risk assessment approaches. 2. Based on these, develop a risk assessment approach appropriate to the higher education sector. 3. Apply the risk assessment to information gathered under other areas of concern (eg format list in Preservation Metadata and Technology Watch)			

4. Document and use to inform APSR aims and priorities.			
The Economics of Sustainability. **Costing Models** **Ownership and Rights**			
1. Investigate rights and ownership in deposited materials from university academics. 2. If necessary develop appropriate policy, statement or advice 3. Document and publicise.			
Sustainability, Resources and Costs			
1. Investigate costing models 2. develop a university-sector-appropriate approach to cost collection and economic modelling with regard to digital materials. 3. identify business needs and scope of preservation (policy and risk questions) 4. identify types of digital objects that will be created and need to be preserved 5. identify how long they need to be preserved 6. identify consequences for people and organisation 7. identify methods, standards, tools, technologies, systems to be used. 8. apply costing models to the described categories 9. determine accuracy 10. Review and document			
The Effect of Funding Models on Sustainability			
1. Identify appropriate bodies 2. Develop a recommendation for the funding bodies 3. Make appropriate representation			
Organisational Sustainability			
1. Explore the issue of organisational sustainability 2. Develop risk indicators for particular organisation structures 3. Provide arguments for, and guidelines on, appropriate repository structure, interoperability and exchange which is applicable at the initial stages.			

4. Compare the risks to centralised and decentralised models. 5. Document and publicise			
Business Models **The Organisational Structure of Digital Sustainability**			
1. Consider the types of organisational structures that accompany the formation and maintenance of digital repositories. 2. Understand how the organizational structure of digital preservation activities comes about, and how it evolves over time. 3. Determine where failures are likely to occur 4. Develop appropriate corrective policies (incentives). 5. Match corrective policy measures with the types of market failures for which they are the most effective. 6. Document and publicise			
Incentives and Motivations			
1. Determine whether this type of economic modelling is appropriate to APSR partners. 2. Assess and document the inhibitors that may discourage participation 3. Assess the tools and workflows that may reduce or neutralise disincentives 4. Develop tools to determine whether the benefits outweigh the inhibitors. Where inhibitors exists 5. Develop possible incentives that might encourage participation. 6. Model them to determine economic suitability 7. Collect and maintain records of use and other factors which might help to inform this process.			
Sustainability and the Value of the Data			
1. Develop criteria for selection guidelines 2. Develop and document procedures that encourage the lodgement of digital materials, both primary and secondary where appropriate. 3. Develop selection policy which can be used to determine appropriate useful life			

of collection materials. 4. Determine minimum life span for which each item or class of material will be accepted (i.e. guidelines to exclude materials of transient value) 5. Develop review mechanisms and procedures.			