



Request for Proposal

Digital Library Infrastructure Replacement

For the

National Library of Australia

Request for Proposal Number: RFP 11/067

Contact Officer: David Wong

Section A – Introduction

A1. Overview

A1.1 The National Library of Australia (the **Library**) is seeking Proposals from Respondents for the provision of Digital Library Infrastructure systems (the **Services**) in order to meet the outcomes set out in **Section B** (Statement of Outcomes). Respondents are not required to bid to provide all the Services – for example, the Library will accept bids providing solutions for particular material types, such as audio, or solutions for particular activities, such as delivery – but should be aware that the Library may favour Proposals which meet all of the Library's requirements.

A2. Overview of process

A2.1 The Library's current intention is that this Request for Proposal (**RFP**) be the first stage of a two-stage process. As part of this two-stage process, this RFP is being used to:

- (a) provide information to industry about the nature and scope of the Library's requirements with the intention of encouraging innovative solutions; and
- (b) enable the Library to further refine those requirements and eliminate impracticable or inappropriate bids.

A2.2 The Library's present intention is to evaluate Proposals submitted pursuant to this RFP (**stage one**) in order to develop a shortlist of Respondents who may be invited to submit tenders in relation to a select Request for Tender (**RFT**) (**stage two**). The Library would then evaluate the tenders in accordance with the terms of that RFT, with a view to executing a contract with a preferred tenderer. The Library's present intention is that this RFT would be issued approximately 3 months after the Closing Time.

A2.3 However, Respondents should note that the Library reserves the right to:

- (a) change the date for issue of any RFT;
- (b) not proceed to an RFT stage; or
- (c) sole source from one or more Respondents to this RFP all or part of the Services, in which case the Library may choose not to conduct a formal RFT process.

A2.4 The Library is not bound contractually, or in any other way, to any recipient of this RFP unless and until a contract has been executed.

A2.5 The Library's expectation at this point in time is that any RFT and contract resulting from this RFP may include a requirement for the successful tenderer(s) to offer to provide Services to other Government Agencies.

A2.6 Any decision by other Government Agencies to avail themselves of an offer by the successful tenderer(s) (if any) may result in a separate contract with the tenderer(s) in

respect of that Government Agency, on substantially the same terms and conditions as any contract between the Library and the successful tenderer(s).

A3. Evaluation of Proposals

A3.1 The Library will evaluate Proposals based on the process and evaluation criteria set out in **Section C** (Evaluation and Selection). It is in the interests of Respondents to address all of the evaluation criteria.

A4. Format for Proposals

A4.1 Respondents are invited to submit their Proposals in the format at **Section D** (Proforma Proposal). However, Proposals in other formats may be considered so long as they meet the Conditions for Participation set out in **clause C2** and address all of the other evaluation criteria.

A4.2 Given that its requirements are not finalised, the Library does not require Respondents to submit final prices at this stage. However, Respondents are required to provide indicative pricing or price ranges so that the Library can determine which Proposals are worth considering further. If Respondents are selected to proceed further (either as part of a formal RFT or a sole source arrangement), Respondents will be required to provide firm pricing at that stage.

A5. Contact Officer

A5.1 The Library's point of contact (**Contact Officer**) in relation to this RFP is:

David Wong
The National Library of Australia
Parkes Place
Canberra ACT 2600
Fax: +61 2 6273 3648
Email: dwong@nla.gov.au

A5.2 Respondents must not communicate with, or solicit information in relation to this RFP from any employee of the Library, other than through the Contact Officer.

A5.3 Respondents may seek clarification in writing of this RFP from the Contact Officer up to 5pm Tuesday 9 August, 2011.

A6. Deadline and place for lodgement

Proposals must be submitted by 2pm Friday 12 August, 2011, by being placed in or addressed by mail to the following Tender Box:

Tender Box
Lower Ground Floor 2 Security Room
National Library of Australia
CANBERRA ACT 2600

Section B – Statement of Outcomes

B1 Introduction

The National Library of Australia's role is to ensure that documentary resources of national significance relating to Australia and the Australian people, as well as significant non-Australian library materials, are collected, preserved and made accessible either through the Library itself or through collaborative arrangements with other libraries and information providers. By offering a strong national focus in all that the Library does, and cooperating with others, the Library supports learning, creative and intellectual endeavour, and contributes to the continuing vitality of Australia's diverse culture and heritage. Increasingly, library materials are being acquired and demanded in digital form.

The Library proposes to redevelop its digital infrastructure to enable it to collect, generate, manage, preserve and provide ongoing access to the rapidly escalating volume of digital information resources that fall within its collecting mandate. It is also proposing to better meet community expectations for easy and timely access to national collections by converting as much as possible of its traditional content into digital form, and by delivering digital content, whether born-digital or digitised, online.

The Library must increase its capacity to collect digital content, including the development of more robust tools for importing, managing and preserving such content. It must also increase its capacity to digitise its paper and other analogue collections, and to deliver digital content online, while supporting scalable and reliable storage capacity measured in multiple petabytes.

What follows is information covering the Library's preservation systems, current approaches and needs. This is to provide industry with appropriate details of the nature and scope of the Library's requirements in order to encourage innovative solutions within Proposal submissions.

B1.1 Library Goals for Digital Infrastructure

In the context of its digital collecting, the Library is seeking outcomes that enable it to:

- Scale up its collecting of digital materials – to support both increased acquisition of materials already in digital form, and mass digitisation programs for generating digital content from physical sources
- Support a full spectrum of collection workflows – from small-scale, single-item, largely manual workflows for some collection materials, to high volume, high throughput, automated batch workflows for others, such as digitised newspapers
- Ensure ongoing access to digital materials – supporting preservation planning, monitoring and response for digital materials
- Support efficient user-centric content delivery – providing rich user experience of collection content, relationships and context – and enable 'getting'
- Manage content and metadata in an efficient, reliable, sustainable, maintainable and accountable manner
- Supply business workflow efficiencies

B1.2 The Digital Library Environment

The Library has a long history in and ongoing commitment to providing innovative digital library services across the breadth of its collections. It is seeking to maintain and expand the activities supported through its existing digital infrastructure to cover:

- Digitisation of print publications and physical materials – for example, books, journals, newspapers, pictures, manuscripts, maps, music, ephemera, realia;
- Preservation capture of sound and audiovisual materials – sound and video recordings;
- Web archiving, including support for alternative collecting scopes – highly selective harvests, harvests based on selected seed URL lists, and harvests of domains or sub-domains;
- Acquisition of ‘born-digital’ materials, deposited online or via a range of physical carriers – tapes, discs and drives – and including current or recent materials, legacy materials, materials produced in uncontrolled environments or controlled in a records management system; and
- Management, preservation and delivery of complex content and metadata with complex relationships and multiple delivery contexts.

The Library currently maintains a suite of systems to provide its digital library services. The following systems are not in scope for replacement under the current request for proposals, and proposed final solutions must integrate with these systems and their future replacements:

- Acquisition management system – RefTracker;
- Archival management system for manuscript materials – Archivists' Toolkit;
- Authentication services – Active Directory;
- Equipment and services in the Library's audio conversion facility, for audio preservation, capture, analysis and processing – QUADRIGA AudioCube, DOBBIN;
- Integrated Library Management System (ILMS) – Voyager;
- Interlibrary loan management and document delivery system – Relais;
- Issue tracking system - Atlassian JIRA;
- Online ordering system, for clients to request copies of collection material requiring mediated access - Copies Direct (in-house system);
- Online Public Access Catalogue (OPAC) - based on VuFind;
- Payment services – eWay payment gateway;
- Records Management System – TRIM;
- Rights Management System – Sprightly (in-house system);
- Search and discovery service – Trove (in-house system);
- Storage system – Digital Object Storage System (DOSS) (in-house system); and
- User Management System – in-house system;

The following systems are in scope for replacement:

- Digital Collections Manager (DCM) – workflow, management and delivery for digitised and born-digital pictorial and audio materials;
- Newspapers Content Manager (NCM) – workflow and management system supporting mass digitisation of Australian newspapers and journals;
- PANDAS/PANDORA – harvest, management and delivery system for selective web archiving;

- Whole Australian Web Domain Harvest service – outsourced harvest, returned content stored and managed by the Library; and
- Prometheus – workflow and management system for capture of digital content from physical carriers.

B1.3 Proposals

The Library invites proposals for innovative solutions addressing its requirements, detailed in the following sections, with the following in mind:

- The requirements expressed in this document are a high level view of what we require, to invite responses as set out in clause A2.1;
- The requirements are not intended to imply a specific architecture;
- Proposed solutions will need to be flexible enough to accommodate additional or revised requirements in future, as the Library's business and workflows change over time;
- The Library is not seeking an overall collection management system, nor is it seeking a storage management system;
- The Library is not necessarily looking for a single software system covering these requirements, but rather a suite of software that can be integrated in a modular, adaptable way;
- The Library is seeking to address all its requirements, but not necessarily from a single Respondent. Proposals covering modular portions of the requirements are also invited; and
- Solutions will need to integrate with some existing systems (for example the Voyager Integrated Library Management System, the Library's existing storage management system, and Rights Management System). Proposals should include details on how solutions may be integrated with existing systems and their future replacements.

B2 Outcome Requirements

The Library's digital library infrastructure broadly operates across four main activities:

- Collect;
- Manage;
- Preserve; and
- Deliver.

High-level requirements for each of these activities are presented in subsequent sections.

The Library is seeking solutions which are scalable, configurable, extensible, flexible and adaptable to the Library's changing environment and needs.

B2.1 Collect

Collecting digital library material includes the activities of selection, assessment, arrangement, description, applying intellectual control (identification of individual objects) and access control for born digital material as well as externally and internally digitised material. It also includes processes for undertaking, tracking and reporting on internal digitisation activities.

The collecting activity also covers storage of digital library material, which is the activity of ingesting content files and metadata to a repository. Storage can include automated system processes such as creation of derivative copies on ingest.

B2.1.1 Collect – Context

The Library currently collects and creates a variety of born digital and digitised content via a number of pathways:

- Curated in-house digitisation of the Library's collection material, including creation of images, text, sound and audiovisual representations of books, serials, maps, manuscript items, sheet music and oral history and folklore recordings;
- Curated harvesting and archiving of websites, by Library staff and by partner organisations;
- Mass digitisation of images, text, sound and audiovisual material by external suppliers;
- Bulk harvesting of websites, including whole domain harvests by an external supplier;
- Collection of digital publications and unpublished material on a variety of physical media, e.g. CD-ROMS, floppy disks, ZIP disks; and
- Creation of new digital content in-house or by external contractors – for example: digital photography; digital sound and audiovisual recording; production of transcripts, summaries and finding aids; creation of image-based finding aids for maps.

The collected material may include complex digital content, consisting of multiple files in different formats, for example:

- A newspaper represented by a set of images and OCR text transcript;
- An interview represented by a sound recording with multiple sessions, a time-coded XML summary and time-coded XML transcript; and
- A website consisting of many pages of text, HTML markup, Javascript code, images and video clips.

In addition, proposed solutions will need to support the Library's current or anticipated need to collect a wider variety of digital material, including:

- Publishers' pre-publication files for printed books, serials and newspapers;
- Large born-digital collections of images and documents;
- Sound, video and other audiovisual files, including multiple recording formats, multi-track recordings and surround sound formats;
- Music, including industry standard notation, scoring, sound design, editing and production files, plus other non-standard application outputs;
- Electronic journals, e-books in a range of design, production-ready and publication output formats;
- Databases and content from record keeping systems;
- Scientific and other research data sets;
- Architectural, CAD and other design files;
- Interactive and synchronous digital artworks;
- Social media content; and
- Small representative samples of games and software applications.

B2.1.2 Collect – Requirements

The Library must be able to:

- Selectively collect a large quantity of born-digital material in a wide variety of formats, considering curatorial value and preservation risks when deciding what to collect;
- Digitise material from physical and analogue electronic collections to create image-based, sound, audiovisual and textual representations of the original material, internally and through outsourced mass digitisation;
- Represent an intellectual work (for example, book, map, musical recording) as a complex digital object, which may have many related representations and many files per representation; and
- Store metadata relevant to the management, preservation and delivery of the digital object.

We are, therefore, looking for solutions that will:

1. Enable staff to manage internal development of digital collections, including:
 - Identifying material to be digitised or captured;
 - Following standardised workflows to create born-digital material; to capture existing born-digital material from a range of carriers, file systems, or network accessible locations; or to convert non-digital materials to digital form;
 - Record or import descriptive, technical, preservation, access rights and structural metadata about digital objects; and
 - Tracking status of digital objects through these workflows and verifying successful storage of files.
2. Enable staff or authorised remote agents to manage development of digital collections from external sources, especially web content, including:
 - Identifying and selecting material to be acquired;
 - Following standardised workflows to harvest existing born-digital materials;

- Defining harvest scope and schedules across a range of approaches – highly selective site-based collection; wider curated harvesting based on url seed lists; domain harvesting
 - Recording licences or permissions for collection;
 - Recording or importing descriptive, structural, technical, preservation and access rights metadata about digital objects;
 - Tracking status of content and processes in collection workflows; and
 - Assessing quality of harvested content and metadata and harvest scope and taking remedial action as required.
3. Enable staff to manage development of digital collections via online deposit from external sources, including:
 - Managing set up of authorised contributors;
 - Manage set up of content requirements and acceptance criteria for authorised contributors;
 - Recording licenses, permissions or transfer of rights for deposited material;
 - Recording or importing descriptive, structural, technical, preservation and access rights metadata about digital objects;
 - Tracking status of content and processes in collection workflows; and
 - Assessing quality of transferred content and metadata and taking remedial action as required.
 4. Accept digital content, metadata or supplementary documentation deposited by external contributors - contractors, authorised depositors.
 5. Accept digital content prepared for submission in standardised ways, including large and complex objects that consist of one or more files in known or unknown formats.
 6. Actively harvest content and metadata from external sources.
 7. Accept (import) or harvest records or sets of existing descriptive, administrative, structural, technical or provenance metadata with or without accompanying digital content.
 8. Enable staff to appraise and select individual digital content items or groups of items for inclusion in our digital collections, including assessing content against the library's digital content standards.
 9. Enable staff to troubleshoot problems in collection workflows, including routing of problem content or metadata for assessment and remediation, restart of workflows from appropriate points and re-introduction of remediated content to standard workflows.
 10. Enable staff to arrange and describe digital content, including creating or supplementing metadata or creating an alternative structure for a complex object or series of objects.
 11. Support arrangement and description of content, including:
 - Creation or import of content guides, finding aids or other descriptive lists providing access points to collection content;
 - Support description of and references to content at various aggregation levels from multiple descriptive sources or contexts; and
 - Support description of and references to content at various aggregation levels across multiple items, sets of items or collections.
 12. Support creation or collection of digital content representing only part of an item, part of a set of items or part of a collection, and defining the part's context within the whole.

13. Enable staff to manage contract consignment workflows for creation of content or parallel versions of content.
14. Enable staff to link system records about digital objects to records in other systems, for example the library management system or rights management system.
15. Enable staff to manage collection workflows that involve external contributors or multiple internal contributors.
16. Enable staff to report on progress of internal digitisation and external collection or deposit processes and projects.
17. Enable staff to associate pre-defined system policies, acceptance criteria, or metadata sets with a set of digital objects.
18. Support a wide range of formats for preservation master copies and access copies.
19. Support multiple master copies and multiple access copies for a digital object.
20. Support different editions of a digital object (for example, edited and unedited versions of transcript, map with and without contour markings).
21. Support deposit of standard types of content directly from other systems.
22. Validate content files and metadata against acceptance criteria on submission.
23. Automatically identify formats and transform content to standard preservation formats on submission, if appropriate, according to defined system policies.
24. Automatically create access copies from preservation master copies.
25. Support import of access copies and alternative or parallel representations of content (e.g. Transcripts, ocr text) created via external processes.
26. Automatically extract or create descriptive, administrative, structural, technical or provenance metadata from files and from system processes.
27. Enable authorised staff to inspect and update automatically created metadata.
28. Support recording complex relationships between digital copies that represent the same content in different ways.
29. Support recording a variety of relationships between catalogue or collection management system records and digital copies, alternative representations and versions.
30. Record version history and process history for preservation master copies.
31. Automatically provide a persistent identifier for internal use and index metadata so that copies can be found quickly.
32. Support assignment of access control policies for content and metadata at the time of collection.
33. Support enforcement of access control policies within collection workflows, limiting staff access to content in process as well as under management – for example, for sensitive or sacred material.
34. Provide workflow patterns for collecting digital material that are as consistent as possible, while catering for the different requirements of various material and collection types.
35. Automatically assign unique and persistent identifiers to each item, that are independent of alternative identifiers for an item or set of items.

B2.1.3 Collect – Scenarios

The following workflows and scenarios are intended to provide context and illustrate expected complexities to be addressed. They are not exhaustive.

B2.1.3.1 Workflow overview examples

The following examples show some of the typical activities undertaken in collection workflows for different sets of digital collection material.

Collection workflow 1

Digitisation workflow: The Newspaper Digitisation Program (NDP) has contracts with companies for scanning microfilm and/or print copies of newspapers and magazines to produce TIFF images for each page. NDP has separate contracts with other companies for content analysis and OCR of the TIFF images. NDP staff ingest and quality assure the scanned images, then batch them up and send them to OCR contractors. The deliverables contain OCR data and article and page metadata, which are extracted and indexed. The results are quality assured and the pages and articles are made available in the search and discovery system.

An alternative scenario is the same as the above except that page metadata (e.g. issue date and page numbers) are assigned by NDP staff during quality assurance of the scanned images, instead of by the OCR contractor. NDP staff identify duplicate and missing pages as part of this process.

Contractors re-supply any scan jobs and OCR jobs which fail the ingest process (e.g. because of validation errors) or quality assurance (e.g. because of poor image quality or inaccurate zoning). NDP staff track the status of scan jobs and OCR jobs. After scanned pages are quality assured there can be a delay of weeks or months before they are OCR'd.

Volume of work varies but upwards of 200,000 scanned images and 200,000 pages OCR'd text per month is not unusual. Some scanning and OCR may also be done in-house at the Library.

A comparable workflow would apply for large scale digitisation of other print materials – books, journals, ephemera – but, potentially, a proportion of digitisation and OCR work would be performed in-house.

Collection workflow 2

Sound preservation: The Sound Preservation branch creates preservation quality digital audio of oral history and folkloric recordings through transfer from physical analogue and digital audio recordings and through live born-digital recording.

Audio content is captured as high quality WAV audio files using professional quality equipment, and content is then quality assured, analysed and converted to Broadcast WAV Format (BWF) files, incorporating descriptive, technical and provenance metadata from the management system into the file headers. Additional process and analysis metadata is recorded in an accompanying metadata file. MP3 derivative access copies for supply or delivery may be generated from the recordings.

Audio preservation transfer may alternatively be consigned to external contractors in batches with consignment lists, with batches of WAV audio files and accompanying transfer metadata returned for quality assurance and import for analysis and creation of BWF files.

For oral histories, MP3 audio may be sent to contractors for transcription or creation of time-coded summaries of content. Time-coded transcripts and summaries are returned as TEI XML for quality assurance and subsequent import against the descriptive control records for the audio work and its copies.

The Sound Preservation branch creates 1000-2000 recorded hours of audio per year.

Collection workflow 3

Web archiving: The Web Archiving branch undertakes both selective harvesting and broader scope harvesting of Web content, including harvest of whole domains or sub-domains.

For highly curated selective harvesting, Web Archiving staff evaluate Web sites proposed as being within the Library's collecting scope and make a selection decision. Permission to collect and provide public access is sought from publishers of sites selected for collection and licences are recorded. Staff set the collection scope, schedule and frequency for harvest and may initiate gathering.

For broader scoped harvests, URL seed lists and scope rules may be developed in consultation with curators or specialists and be used to target gathering.

Once initiated, staff may monitor progress of the harvest and may adjust or cease gathering as required. Harvested content may be inspected and excess content deleted or missing content added via supplemental harvests where necessary.

Approved Web content may then be indexed for discovery, described and associated with control records, have access controls applied at various levels of granularity, and be prepared for publishing via the delivery system.

Collection workflow 4

Collect other born-digital content via harvest, deposit or on physical carriers:

Born-digital content may be supplied or collected via several methods:

- Via 'push' mechanisms where content and metadata is deposited or remotely transferred to the Library online;
- Via 'pull' mechanisms where the Library sets up arrangements to select and collect content and metadata from networked locations (for example, publisher FTP sites or Flickr); and
- Via supply on physical carriers (tapes, drives, disc).

Content and metadata may be supplied in agreed packages or formats and may be comprised of simple content (e.g. 1:1 files and metadata) or complex content (e.g. composite, related or interdependent files and complex metadata).

For deposit, content and associated metadata may be deposited or transferred online for appraisal from authorised contributors such as publishers, donors, contractors, partner agencies, or staff acting remotely.

For harvest, third parties may make content and metadata network accessible for harvest by the Library for appraisal.

For content received on physical carriers, content in directories and files or archives files (e.g. ZIP, TAR) might be captured directly or a block or disc image of the carrier might be captured and analysed for file systems. Block or disc images and archive files will be unpacked to extract file systems and contained files for appraisal.

All collected content will be validated according to appropriate criteria defined for submission processes. File and directory structures will be recorded, where appropriate. Files will be analysed for viruses, file format identification and validation, technical and descriptive metadata extraction or generation, and creation of integrity data (e.g. checksums). View copies of content might be automatically created for inspection, where possible.

Library staff might view and appraise content and metadata for selection. Selected content and metadata might be quality assured and might be arranged and described before import, or might simply be imported and further arranged and described, if required, while under management. Relationships between content are recorded and maintained. Access copies of content might be generated, and access controls will be applied.

Contributors might be notified of rejected content and reason for rejection.

B2.1.3.2 Specific workflow aspects or variations examples

The following scenarios highlight aspects of collection workflows or exceptions to the usual workflows that imply certain system requirements.

Collection scenario 1

Bulk import of metadata: Collection areas receive lists describing collections in an agreed form and would like to import the metadata to a system. This will often be a separate process to uploading content files. Examples include:

- Pictures and Manuscripts branch receives a spreadsheet describing the contents of a set of boxes, including information about born digital material. They want to import the whole box list as a document that can be made available to discovery and delivery systems, and import separate records about born digital material that can be associated with content files when they are uploaded;
- Oral History branch receives a spreadsheet including descriptive and technical metadata about a series of 300 interviews on reels and cassette tapes. They would like to bulk import descriptive metadata to the catalogue and bulk import technical metadata to a digitisation system so that this information can be associated with the audio files created through internal digitisation processes; and
- A hierarchical finding aid for a manuscript collection in EAD XML is imported to provide additional description of the collection content at various aggregate levels, including collection level. Content from the collection might be digitised and associated with descriptive records at appropriate levels.

Collection scenario 2

Collect content from physical carriers: Content on physical carriers may present particular complexities related to characteristics, capacity or limitations of the carriers involved. Examples include:

- Pictures and Manuscripts branch receive a collection of CD-ROMS, plus some old 3.5 inch and 5.25 inch floppy discs as part of an accession. The 5.25 in floppy discs cannot be captured by Manuscripts and need to be captured by specialist staff or contractors. Disc images of the 5.25 in floppy discs are captured and returned to Manuscripts. Manuscripts then need to load existing disc images for appraisal and processing as usual;
- Asian Collections receives a set of DVDs and VCDs they wish to capture for preservation. The file systems, filenames and file content contain Asian characters which must be maintained, and the discs must be captured to enable playback; and
- A set of multimedia CD-ROMs from the late 1990s is distributed on 'hybrid' discs, containing separate file systems for use on PC and Apple Macintosh, with different executable files for running the multimedia publications on the different platforms. These publications might be captured and processed such that access through the alternative platforms is retained.

Collection scenario 3

Collect born-digital content with print equivalents: Collection areas may collect digital content in parallel with or instead of equivalent print materials. This content must be able to be collected, described and made available for access alongside similar print materials. Examples include:

- Australian Collections branch receives a set of digital files for an A2 printed exhibition poster and banner for inclusion in the collection from an exhibition designer. This digital version needs to be stored and delivered with links from a catalogue record in Voyager. The files may arrive as artwork in Adobe InDesign or Portable Document Format (PDF), with rights for online delivery. Digitising from the printed copy would yield a lesser quality image, so it is preferable to upload and deliver the original file; and
- Music and Dance branch wants to be able to collect and archive e-ephemera and e-programmes in conjunction with Ephemera and Web Publishing. This is to replace the traditional, hard-copy material which currently provides valuable artistic, social, historical and biographical data. The collection area has received digital files for printed items sought for the Library ephemera collection (for example, handbills, fliers, posters etc. from political campaigns; CDs with digital files or individual PDFs of programs for special events).

Collection scenario 4

Select, arrange, describe, and maintain relationships: The Library is likely to receive mixed collections of complex content and must be able to appraise and select content, arrange and describe where necessary, and maintain complex contextual, structural or technical relationships between content, including between physical and digital materials. Examples include:

- The Library receives a massive photographic collection of over 150,000 items (transparencies, print and born-digital material) produced from the 1950s through to 2010. The collection includes photographic project notebooks and spreadsheets, but due to ill-health some of the photographer's later work has not been well described or documented. The Library would like to bulk import the digital existing digital content, including any file structures and metadata (including embedded descriptive, technical or geospatial metadata) to allow appraisal and selection. Photographic prints and transparencies may also be appraised and selected for inclusion in the Library's collection and may be scheduled for digitisation. The Library wishes to be able to catalogue, arrange and describe the selected

content at appropriate aggregate levels to aid discovery, associate available metadata with corresponding material, and maintain contextual relationships between the physical original material, any digital surrogates created, and the born-digital content and metadata;

- Maps branch has received a collection of GIS data on a 16 GB USB drive. The digital mapping data is in a range of formats, including JPEG2000 images, GIS shapefile data, ECW image files and related co-ordinate XML data, PDF, Access Databases and corresponding CSV files, requiring several viewers to enable viewing of the maps and geophysical data. Maps wishes to capture and appraise the material, viewing and selecting which material to collect. They also need to maintain associations between images and any related co-ordinate data, and the relationships between other files which reference each other to render the maps;
- Australian Serials branch receives a journal publication which, with every issue, includes one or more CD-ROM discs of supplementary material in digital form. Australian Serials wish to create preservation surrogates of this material and associate these with holdings records for each issue of the journal over time to provide access to onsite and offsite clients;
- The Pictures and Manuscripts branch receives a novelist's laptop containing a 300 GB hard drive, which, as well as containing the sequence of drafts and working notes for each of the novelist's recent works, contains the novelist's recent email archive and other personal documents associated with their writing life. Manuscripts wants to capture the contents of the laptop, to retain a copy of the artefact in original order, but then to view, appraise and select or reject content for inclusion in the Manuscript collection - for example, they might choose to reject software utilities, games and football tipping documents, but select novel drafts, personal documents and email archive. The contents of the email archive similarly might be appraised, with selection or rejection of groups of emails (or in rare cases, at individual item level). Manuscripts might also choose to arrange and describe the material in an order different from the original order, while maintaining a record of the original order. Arrangement and description might occur during the process of importing the material, or at any time later, once the material is in storage. Manuscripts would consult with the donor about appraisal decisions and would either destroy or return to the donor any nominated groups of records; and
- As per the previous example, but in this case a composer's laptop is received, containing a variety of music notation files and sound files created in various versions of proprietary software packages over 20 years. Manuscripts and Music staff will need to examine the files, determine and record what software will be required to access the content in the original format, and decide whether to create access copies in alternative formats, if available. It will be necessary to record relationships between files representing progressive versions of the same musical works and how content files may need to be combined for presentation, any associated documentation, and any alternative access copies.

Collection scenario 5

Associate content with various levels of aggregate description: Collection content can be described at various levels of aggregation, ranging from description of individual items or parts of items, to groups of items and series, up to description at collection level. The same collection item may need to be referenced simultaneously from various levels of description. Staff need to be able to associate collection content and metadata with appropriate and available levels of description, submit content without being required to exhaustively arrange and describe it first, and revise associations of content and descriptive levels as these become available. An example is:

- Manuscripts staff have received a request for a copy of a letter within a large collection that has not been fully arranged or described. After the letter is digitised for the request, collection staff choose to retain the digitised copy. The collection only has a catalogue record at the collection level. The copy will need to be linked to this collection record and also be described as an individual item. The collection is later arranged and described down to sub-series level. The copy of the letter may then be associated with the appropriate sub-series in the finding aid.

Collection scenario 6

Extract, generate, update and embed metadata: Collected content may contain embedded metadata or enable useful metadata to be derived. Authorised staff may also need to update or create additional metadata as required. Inclusion of embedded delivered content would also help maintain contextual references to the Library's collection. Examples include:

- Import commissioned video interviews in high quality lossless video format (Motion JPEG 2000), with embedded headers containing metadata. Generate several derivative delivery formats with metadata headers;
- Extract descriptive, technical and camera setting metadata from born-digital images.
- Extract, derive or record geospatial metadata to support location or map based searching – for example: from image files where this is recorded by the camera; from location information in catalogue records for any type of digital object - books, pictures, maps, manuscripts, music, oral history, newspapers; or from geo-tags added by staff or library users; and
- Incorporate geospatial (GPS) data to identify location of audio/video recordings, especially in the context of environmental sound recordings.

Collection scenario 7

Error handling: Sometimes jobs become 'stuck' in automated processing stages for various reasons, ranging from problems with specific files to server memory issues. Library staff need be able to check the status of jobs in progress, with options for troubleshooting what went wrong and recovering, resetting or aborting jobs to try import processing again.

B2.2 Manage

Managing digital library objects and collections includes the activities of finding objects in a repository by a wide range of criteria including complex searches, updating and versioning content files, updating descriptive and technical metadata, arrangement and rearrangement of objects in the repository, access control, and creating collection management reports.

Administration of digital library infrastructure can include management of users, groups and permissions; configuration of automated processes, user workflows and metadata control lists in systems; monitoring systems and analysing system problems.

B2.2.1 Manage – Context

The Library needs to manage, update, monitor and report on the content and metadata it collects, preserves and delivers, as well as manage system operations and administration.

Solutions must offer flexibility to react and adapt to change, enabling updates to managed content and metadata in response to new information, requests, policy changes or resource opportunities, and enabling addition or adaptation of workflows and operations in response to changing technologies, capabilities and expectations.

B2.2.2 Manage – Requirements

The Library must be able to:

- Configure systems to facilitate efficient workflows for collecting, managing, preserving and delivering digital collection items or entire collections;
- Find and report on digital objects based on complex criteria as well as simple searches;
- Control internal as well as external access to digital objects; and
- Update content files and metadata, and rearrange complex objects and ordered series.

We are, therefore, looking for solutions that will:

36. Enable staff to find records describing digital objects by a wide range of criteria (including advanced search with multiple criteria), covering descriptive, technical, preservation, administrative, access rights and structural metadata.
37. Enable staff to generate regular pre-set and custom one-off reports on digital collections based on a range of criteria.
38. Enable staff to schedule or run ad hoc analyses on content and metadata.
39. Enable staff to create, update or import descriptive, technical, preservation, access rights and structural metadata for digital objects, individually or in batches.
40. Enable staff to upload new versions of master copies and access copies, or generate new versions of access copies from masters, individually or in batches.
41. Enable staff to view version history and manage which versions are available for access.
42. Enable staff to manage relationships between different representations of the same content - for example, the audio recording, summary and transcript of an interview.
43. Enable staff to manage logical as well as 'physical' structure of digital objects, and associate a digital item with more than one set of structural or logical relationships.
44. Enable staff to produce alternative arrangements of digital content, including content from multiple items.
45. Enable staff to assign digital items to any number of arbitrarily defined sets for management and further action.
46. Enable staff to manage relationships between objects that belong to a set - for example, pages in a book.
47. Enable staff to assign one or more persistent identifiers down to file level for items or parts of items, or for sets of items.

48. Enable authorised staff to manage content files individually or in batches, including permanently deleting items, parts of items or access copies.
49. Enable staff to manage information about projects, tools and other controlled lists of metadata.
50. Enable staff and authorised external agents to manage users and user permissions, including permissions for external agencies, contributors and depositors.
51. Enable staff to link digital object records to records in other systems, for example catalogue, rights management system.
52. Enable staff to define sets and assign digital items to sets for managing them according to various criteria.
53. Enable staff to define and approve system policies and default values – for example collection groups, functional roles, preservation intent, access policies, and risk profiles.
54. Enable staff to define and approve acceptance criteria for deposited content - including format, well-formedness and quality of content and metadata.
55. Enable staff to define and approve metadata sets to be supplied or extracted for defined sets of digital items, processes or workflows.
56. Enable staff to quality assure and approve updates to digital items or metadata.
57. Support links to specific records from other systems, make metadata available for use by other systems and pull in metadata from other systems.
58. Support complex user permissions and access restrictions to content and metadata based on multiple criteria.
59. Support access restrictions on content down to file level for both external and internal users.
60. Support access restrictions on metadata for both external and internal users.
61. Assign appropriate storage locations for deposited content and metadata and for content and metadata generated, transformed or derived as part of deposit, management or preservation processes, according to system policies.
62. Manage file locations and maintain associations between storage locations and item identifiers if files are relocated.
63. Version digital items after an action which updates or changes the item, update associated metadata and relationships, keep and manage previous as well as current versions.
64. Support a range of descriptive, technical, administrative, structural and provenance metadata needs for different materials or collections.
65. Support export and migration of managed content and metadata in standardised ways to and from the system.
66. Enable staff to selectively restore content and metadata from backup.
67. Enable scheduling of backup and disaster recovery regimens.
68. Monitor and report on storage usage and growth, and monitor and report on system performance.

69. Extensibly support integration of additional or alternative tools to enhance collection, quality assurance, processing and analysis, discovery, access or delivery.
70. Extensibly support collection and management of additional formats.
71. Enable staff to manage physical inventory for those collections requiring it.

B2.2.3 Manage – Scenarios

The following scenarios are intended to provide context and illustrate expected complexities to be addressed. They are not exhaustive.

B2.2.3.1 *Specific workflow aspects or variations examples*

The following scenarios highlight aspects of management workflows or exceptions to the usual workflows that imply certain system requirements.

Management scenario 1

Update of metadata: Library staff need to be able to update metadata as required, and to do so efficiently. Examples include:

- Collection staff are notified that the spelling of an interviewee's name needs to be corrected in the catalogue record and also in metadata associated with digitised files and in the rights management system. They would like to update the information in one place and have the correction reflected in the other systems; and
- Sound Preservation staff realise that some technical metadata imported from a previous system is incorrect for a particular series of recordings and want to batch update a replacement set of corrected metadata recorded in a spreadsheet.

Management scenario 2

Batch processing and update of content: For sound recordings, it is decided that the delivery system will deliver streaming MP3 derivatives. For existing BWF master files which do not already have a corresponding MP3 access copy, Sound Preservation staff need to batch download master files from the repository to feed them through an audio processing workflow system that will create the MP3s, and then batch ingest the new MP3 files.

Management scenario 3

Manage access controls: Staff need to create and update access controls for content and metadata as required, with immediate application of restrictions. An example is:

- Oral History branch has received advice that certain content in a session of a long interview might be regarded as defamatory. They need to restrict that session from public access, while still allowing the other sessions to remain accessible through a web delivery interface. The change should take effect in the delivery system as soon as it is made in the management system.

Solutions also need to support access restrictions, even for internal access, for example:

- The Pictures and Manuscripts branch have loaded a collection of personal papers from a set of disks supplied by a donor. Some of the material - several files in separate directories and

a number of whole directories - is sensitive and has different access conditions to other parts of the material. Manuscripts want to impose different access restrictions to these individual files and directories, such that only some authorised Manuscripts staff can access and view the material;

- Print material with copyright and access conditions may be digitised for preservation purposes. Information about restrictions like ‘onsite viewing only’ needs to be available to clients and staff. It should be possible for staff to record conditions under which material was copied and indicate an end date for restrictions. There should be automatic processes to report when copyright and other restrictions expire;
- The Pictures Collection contains images that have been identified by representatives of Indigenous communities as having secret or sacred content that should be restricted. The items can be digitised for preservation purposes and to supply copies to Indigenous people of the appropriate groups and gender, but will never be placed online. We would like to be able to restrict their digitised copies so that only selected staff of the appropriate gender who have been trained in protocols surrounding this material can view the images. This group should be managed by a high level administrator so people can not add themselves to the group; and
- A significant number of interviews in the Oral History Collection are restricted for the life of the interviewee and may contain sensitive content. It is important that access is limited, even within the library. For restricted material in this collection, only authorised staff within Sound Preservation and Oral History should be able to access the content files.

Management scenario 4

Create or update relationships between representations: Oral History staff need to define relationships between files that represent the same content in different ways, so that these can be delivered together. For example a sound recording could have three audio sessions (three master BWF files and nine MP3 files for different bit rates), a full-text transcript (one Word document covering the three sessions), full-text transcript with time-points (XML document), edited version of XML transcript, and a summary with time-points (XML document).

Management scenario 5

Create or update relationships between works: Oral History staff want to link related objects for delivery, for example linking a sound recording with photographs of participants taken during the sound recording session.

Management scenario 7

Support ongoing arrangement and description: Manuscripts are upgrading the level of arrangement and description for the collection of a prominent Australian academic, which includes both paper-based materials and digital items previously captured and stored for preservation. The original order of the material is subsequently re-arranged to include additional series and sub-series, and some new groupings of previously disparate material have been created. Manuscripts want to maintain details of the original arrangement, as well as the new arrangement, which includes more extensive description.

Management scenario 8

Track audit history of original object: A donor requests a copy of the material previously donated to the Pictures and Manuscripts branch. Manuscripts staff retrieve the requested material

and supply it to the donor. The donor disputes that the copy of the material which Manuscripts staff supplied was what was originally donated to the Library. A Manuscripts staff member extracts an audit report for the material, which provides a detailed history of events affecting each file in the set, from the point of receipt, through any processing undertaken as part of bringing the material into storage, and while in permanent storage, such that Manuscripts can demonstrate that the material represents a true record of what was deposited.

Management scenario 9

Track project progress: Digitisation branch are digitising a collection of 40,000 images by a significant Australian photographer. This work will be spread over several years and they need to report on status of jobs in this project – for example, which jobs and items have been requested by Pictures branch; which jobs are in various stages of progress through the digitisation and ingest process, completed, and acquitted.

B2.3 Preserve

Preservation of digital library material includes the activities of collecting and maintaining preservation metadata, reporting on digital collections to identify formats and preservation risks, and performing actions on sets of digital objects to ensure that the objects continue to be accessible.

B2.3.1 Preserve – Context

The Library's digital collections are diverse in terms of genres, sources, and levels of control or influence over their creation. In addition, the Library's digital collections are currently approaching petabyte scale, and are expected to continue growing.

From a preservation point of view, the Library's digital collections present:

- A mix of materials needing to be kept in perpetuity, along with materials that can be discarded after specified periods or events;
- Mixed levels of complexity in terms of object structure and relationships;
- Mixed levels of intellectual control;
- A wide range of received file formats (and carrier formats);
- Different levels of complexity in preservation planning and processing;
- Different timetables for preservation action;
- A need for different preservation approaches, often at different scales; and
- A need for recurring – and possibly changing - preservation action cycles over time, using a changing suite of tools.

Preservation responsibilities

Preservation of the Library's digital collections involves three main goals:

- Maintaining access to reliable data at bit-stream level;
- Maintaining access to content encoded in the bit streams; and
- Maintaining access to the intended and available meaning of the content.

While specific preservation activities may focus on one or more of these goals, the Library's preservation responsibility is only fulfilled when all three goals have been adequately addressed.

This responsibility applies across all digital collections, subject to policy decisions for specific groups of digital objects.

Likely preservation treatment actions

Broad preservation action approaches that are likely to be required will include:

- Format migration at the point of collecting;
- Format migration on recognition of risks;
- Format migration at the point of delivery;
- Emulation of various levels of software and hardware environments;
- Maintenance or supply of appropriate software or hardware;
- Documenting known problems for which no other action can be taken; and
- Deletion.

B2.3.2 Preserve – Requirements

Systems and arrangements

Systems and arrangements to address these preservation challenges will have to enable a small human resource base to deal efficiently, effectively and accountably with billions of objects, in diverse formats and collection contexts for access in wide ranging end user environments. Such systems and arrangements will need to support mass processes while offering the necessary flexibility to tailor both large- and small-scale preservation actions.

Role of preservation metadata

Consistent metadata is a critical enabler of preservation planning and action. Regardless of where metadata is stored, adequate and consistent preservation metadata must be recorded and maintained as a part of every digital object. The metadata must be efficiently accessible and reportable in order to support monitoring, grouping, and processing of objects, and to maintain an auditable record of what has happened to each object.

Required preservation processes

The Library must be able to:

- Understand what it holds in its collections;
- Understand what its preservation intentions are for every digital object and what it is entitled to do to realise its intentions;
- Understand what is required to provide access, and the current level of support available within the Library;
- Anticipate the effects of changes in that support;
- Recognise planning triggers, and plan and take appropriate action on a scale appropriate to the size of the target; and
- Audit the effectiveness of its preservation arrangements and modify the arrangements if necessary.

We are, therefore, looking for solutions that will:

72. Enable staff to define and approve sets of managed content, based on content characteristics, metadata or system policies, including preservation intent.
73. Enable staff to define and approve risk criteria for sets of managed content, and to define and approve more general system risk conditions that affect content or the management of content.
74. Enable staff to search and report on managed content makeup, characteristics, metadata, risks, system conditions, schedules, events and outcomes, including:
 - Summary reports of content and metadata characteristics across collections or defined sets of managed content;
 - Detailed metadata and audit history reports for individual items or sets of items;
 - Regular scheduled reports; and
 - Ad hoc reports.
75. Enable staff to create, update and maintain reference information on:
 - File formats and versions;
 - Software and hardware access pathways for maintaining managed content; and
 - Levels of support and fidelity for access pathways, and associated risks.
76. Enable staff to create, update and maintain resource inventories of locally available software and hardware access pathways that may be used for preservation action or for access to managed content.
77. Support import and update of resource inventory information and reference information through exchange with external services.
78. Support analysis of metadata, preservation intentions and risks for managed content, comparing these against reference information and resource inventories to:
 - Help staff identify priorities for preservation planning, monitoring and action; and
 - Detect defined conditions and trigger approved actions.
79. Enable staff to perform 'what if' simulations to determine impact of changes to availability of access pathways, including:
 - Removal of access pathways, to assess risks or impacts on access; and
 - Addition or revision of access pathways, to assess proposed remedial preservation action plans.
80. Enable staff to adjust preservation priorities for sets of managed content.
81. Enable definition of monitoring tasks and reporting of outcomes, including:
 - Verifying and reporting on data integrity (fixity checks) for managed content and metadata;
 - Monitoring for the occurrence of other conditions defined through system policies;
 - Scheduling performance of monitoring tasks; and
 - Performing monitoring tasks in response to ad hoc requests.
82. Enable staff to define and approve actions to be performed on detection of conditions which:
 - Affect content;
 - Affect the ability to monitor content or evaluate conditions; and
 - Meet other system policy criteria (e.g. Formats to be converted on deposit).
83. Notify, report, or take defined actions upon detection of conditions which:

- Affect content;
 - Affect the ability to monitor content or evaluate conditions; and
 - Meet other system policy criteria.
84. Create and manage multiple backup copies of content and metadata.
 85. Enable staff to inspect digital content or metadata which has failed data integrity checks and restore content or metadata from backup.
 86. Enable staff to inspect digital content or metadata for which risk conditions or other system policy conditions have been met.
 87. Enable staff to define and approve types of preservation action for use within preservation planning and evaluation.
 88. Enable staff to define, approve and prioritise candidate preservation action plans for sets of managed content.
 89. Enable staff to define and approve acceptance criteria for preservation action outcomes for sets of managed content.
 90. Enable staff to develop, test, quality assure and approve executable preservation action plans for sets of managed content, using approved action steps, tools and test sets of content.
 91. Support preservation action plans which include multiple steps.
 92. Support preservation action plans combining manual and automated workflows.
 93. Enable staff to estimate performance and capacity requirements for executing preservation action plans.
 94. Enable staff to add and configure tools for use in analysis, preservation action and quality assurance.
 95. Enable staff to schedule and execute approved preservation action plans against sets of managed content in the production environment, including:
 - Execution of preservation action plans within the repository;
 - Extraction of content for preservation action via external processes; and
 - Action on individual items or sets of items.
 96. Support import and integration of preservation-treated content and metadata (from either internal or external processes) including:
 - Verifying that preservation-treated digital content conforms to acceptance criteria for preservation outcomes for designated sets of digital content;
 - Enabling staff to quality assure and approve preservation-treated digital content for incorporation into the collection; and
 - After approval, updating versions of content, metadata and associated relationships.
 97. Support comprehensive reporting for auditing purposes of content and metadata composition, conditions, system policies, risks and monitoring, events, actions and outcomes.
 98. Support export of managed content and/or metadata for transfer to other systems, including options to export all data for simple migration to future replacement systems.

B2.3.3 Preserve – Scenarios

The following workflows and scenarios are intended to provide context and illustrate expected complexities to be addressed. They are not exhaustive.

B2.3.3.1 Workflow overview example

The following example shows some of the typical activities undertaken in preservation workflows.

Preservation workflow 1

Preservation planning, monitoring and action: Preservation staff negotiate the type and level of preservation support required for digital content in collections with individual collection managers. They periodically generate reports on the file format makeup of collection content and identify sets of content to be managed according to agreed preservation intent. Preservation staff define sets of conditions to be monitored for defined sets of content which, when detected, will indicate that preservation action is required to maintain access to digital content in line with agreed preservation intent.

Preservation staff also use content analysis reports to assist prioritisation of preservation action planning, for example, assessing the impact of loss of support for a particular software application, or assessing the relative value of taking preservation action on a small but high value set of content versus action on a larger proportion of the collection.

Preservation staff draft and assess preservation action plans to remedy loss of technical access pathways to digital collection content. For example, preservation plans may involve acquiring alternative software to display or replay collection content, emulating software environments, or transforming content formats to enable display with current software.

On detection of conditions requiring preservation action, Preservation staff prepare and test candidate action plans against a test set of content and evaluate the results for effectiveness in line with agreed preservation intent. Action plans agreed as suitable may then be scheduled for execution against batches of content in the affected set and results may be quality assured. If action plans result in creation of alternative versions of content, the content is incorporated into collection content, and content versions, metadata and relationships are updated accordingly.

B2.3.3.2 Specific workflow aspects or variations examples

The following scenarios highlight aspects of preservation workflows or exceptions to the usual workflows that imply certain system requirements.

Preservation scenario 1

Maintaining preservation planning knowledgebases: Preservation staff maintain a set of knowledgebases providing technical data and relationship information about file formats, software applications and hardware of relevance to the Library, plus inventories of software and hardware locally available to support collection content. Staff periodically update information in these knowledgebases as additional information becomes available or as inventory changes occur. Preservation staff use information in knowledgebases to inform setup of preservation monitoring criteria and to assist planning using information on suitable technical access pathways to types of content.

Preservation scenario 2

Preservation planning and monitoring – Identify content which requires preservation

planning: Preservation staff run analyses, extract reports and set monitoring criteria to inform or alert them to circumstances requiring preservation planning.

Examples include:

- Preservation staff request a summary report on format types that are in the Manuscript collection. The report includes detailing what software is currently available to render these formats, current levels of support for accessing the required software for those formats, and the preservation intentions agreed with the Manuscripts branch. This information will be used to assess the degree and priority of preservation planning required for the Manuscript digital content; and
- Preservation staff set a monitoring criterion for when the number of available instances of access software for any format falls below a specified level. Staff are alerted when new or unidentified formats are deposited for which minimum software support is not available, or when updates to the inventory of available software remove support for existing content below the set support threshold.

Preservation scenario 3

Preservation planning and monitoring – update of metadata for planning and monitoring:

Preservation staff can add or edit metadata for single files or groups of content files to modify monitoring and reporting. Examples include:

- Preservation staff inspect files to manually identify formats which remain unidentified through automated processes. Staff update format identity metadata and specify access dependencies for the unidentified files to enable more specific preservation monitoring and reporting;
- A set of files submitted by a depositor are identified as being invalid with respect to their format standard, but are within submission tolerances and on inspection are able to be viewed. A preservation staff member updates the metadata for the files to mark them for later appraisal for likely preservation problems;
- A preservation manager decides that files in the Web content collections containing specific invalid HTML tags should be marked for re-assessment in 5 years, and sets up an automated notification for planning re-assessment;
- Preservation staff notice increased occurrence of a new file format in a collection and, in consultation with the collection manager, determine a new preservation intent policy for the format within that collection. Preservation staff need to update metadata for the new format in the collection to assign the new preservation intent; and
- Preservation staff need to make a batch correction to file format identity metadata for files with a particular extension that were not identified by automated identification tools.

Preservation scenario 4

Preservation planning and monitoring – 'What if...' analyses: Preservation staff perform simulations to assess the impact of expected changes to the technical access environment.

Examples include:

- A planned rollout to replace the Library's standard operating environment in 12 months' time will remove support for the MrSID plugin which currently provides access to some

characteristics of digital objects in numerous collections. A preservation manager runs a simulation to remove MrSID support from the available support software inventory and assesses the impact against the known collection makeup; and

- A preservation manager compares the relative impact of two alternative rendering viewers on access support for collection content.

Preservation scenario 5

Preservation planning and assessment: Preservation staff develop candidate preservation plans, test them against target content and evaluate the effectiveness of the preservation action with respect to agreed preservation intent. An example is:

- The Manuscript curator has requested conversion to a more sustainable format of around 10,000 born-digital master files of donated texts within the Manuscripts collection only. Preservation staff identify the set of content to be converted by searching on format and collection criteria and establish a sample test set. Two alternative format converters have been recommended. Preservation staff set up and run tests against the sample set of content to evaluate conversion quality against agreed preservation intent and to estimate conversion time.

Preservation scenario 6

Preservation action – Transform file formats: Preservation staff may specify conversion of files in specified formats or format versions to other specified formats or versions. Links to and from other files must be maintained, and file characteristics must be maintained at threshold levels in line with preservation intentions agreed with collection managers. Examples include:

- All RealMedia files within a collection of millions of archived web pages inside WARC container files are to be converted to WebM format with a Flash wrapper. Content must be identified, accessed, converted and re-incorporated within WARC container files;
- A preservation manager recommends that all files currently in the MS Word .doc format (any version) are to be converted to PDF/A format, and identifies that there are approximately 2 million such files in four different collections; and
- The Library has a collection of 1000 video-CDs from which disk images have been taken. A collection manager asks for the video content to be extracted and made available as MPEG objects while preserving any contextual information.

Preservation scenario 7

Preservation action – Provide emulation or viewing software support: Preservation staff may specify that a suitable software environment or application be made available as required to provide access within the delivery environment, or identified as a delivery pathway for content in specified formats. Examples include:

- A preservation manager decides that all browser content in our web collections should be emulated in an environment suited to the age of the archived website (i.e. for content harvested in 2006, the emulated environment will contain software contemporary with that time). Files that require an external viewer will not be part of this emulation;
- A preservation manager decides the best way to retain access to content in Sibelius v6 format will be to acquire new software that can view the format. The software is acquired

and the metadata relating to Sibelius v6 files (including the software inventory) is updated to record the availability of the new access pathway; and

- As per the previous example, but a Scorch viewer or alternative rendering tool for Sibelius v6 files is acquired and is able to be marshalled directly into the delivery pathway for affected files.

Preservation scenario 8

Delete unsupported or unwanted file formats: Particular content files may become redundant, or particular formats may become completely un-supportable over time. In consultation with collection managers, Preservation staff may specify that such content or content in such formats should be removed from the collection, and metadata updated accordingly to record deletions. Examples include:

- Of the 90 million files in the PANDORA selective web archive, there are 5000 ‘.exe’ files, which a preservation approvals committee agrees should be deleted. Approximately one quarter of the .exe files are contained within ZIP or TAR files inside WARC container files; and
- It is agreed that previous generations of preservation masters from which onward format migrations are no longer likely should be deleted.

B2.4 Deliver

Delivery of digital library material includes providing direct access to individually identified master and derivative content files, subject to access restrictions; providing contextual web-based delivery of complex objects and objects that belong to collections; and supporting processes for supplying digital files through other methods such as FTP, email, or on physical carriers. Delivery of digital objects can include packaging sets of content files with descriptive, technical, rights management and structural metadata, so that the complex digital object can be identified and understood in contexts outside the repository.

B2.4.1 Deliver – Context

The Library currently delivers a range of content through its delivery systems – digitised pictures, maps, manuscript items, ephemera, realia, sheet music, books, journals and newspapers; born-digital images; sound recordings with accompanying time-pointed summaries; and archived Web content.

The Library wants to improve options for access and use of its collections, with improved utility, usability and support for delivering a wider range of formats and increasingly complex content in its enhanced collecting scope.

B2.4.2 Deliver – Requirements

The Library must be able to:

- Retrieve individual items or batches of items from the digital collections for a variety of purposes, including fulfilling customer orders for copies, use in the Library’s publications and exhibitions, processing master copies to produce access copies and sharing content with other institutions;

- Deliver digital objects to the public through web based interfaces that provide rich contextual displays of content and metadata and enable library users to interact with the objects;
- Deliver digital items and collections in compliance with any access restrictions, including ability to deliver whole items or specific parts; and
- Deliver content files and metadata directly to the public in a range of formats that can be used in multiple contexts, for example access through web browsers, mobile applications, e-readers, sharing through social networking platforms, downloading for offline access.

We are, therefore, looking for solutions that will:

99. Enable staff to search managed content and metadata within the system.
100. Enable staff to quickly access derivative copies of atomic digital objects (images, documents, audio and video) for various purposes, including quality assurance and preservation processes and selecting appropriate items to download for exhibition and publication purposes.
101. Enable staff to quickly download master copies of digital objects, individually or in arbitrary batches, for use in publications and exhibitions, for generating new derivative versions, and for fulfilling copy requests from library clients.
102. Support delivery of packaged content and metadata for viewing offline – for example, viewing at a standalone pc within a reading room; supply to a client at a remote location; supply to donor; supply to valuer; supply of an entire web content harvest or content instance for transfer to another agency.
103. Enable authorised staff to easily see what access restrictions apply to digital objects, and to quickly modify access restrictions when required at multiple levels, for example, series, item, part, specific version or edition.
104. Enable staff to manage relationships between digital objects for delivery purposes and make choices about how they will be delivered in context.
105. Enable staff to select which of multiple versions or editions of a digital object is the preferred copy for delivery.
106. Enable staff to manage persistent identifiers used for delivery of digital objects, including associating more than one contextual identifier with a digital object.
107. Enable staff to create and configure collection ‘views’ of content and metadata.
108. Make a range of metadata available to discovery systems.
109. Make textual versions of digital objects available to discovery systems (for example, full-text transcripts and summaries of oral history interviews and OCR transcripts of published documents).
110. Enable authorised staff to configure availability of metadata for discovery
111. Enable library users to search by free text, URL searches (for web material), geographic coordinate, placename and bounded area searching.
112. Apply access control policies for access and delivery of content and metadata down to file level for items, parts of items and sets, and support restrictions based on time, location or authorisation.
113. Enable library users to access unrestricted preview and view/listen copies through a web based delivery system.

114. Enable library users to access master copies for items that have no access restrictions.
115. Provide consistent interfaces for delivering digital collection objects, while providing rich delivery features appropriate to the type of object.
116. Enable library users to get full benefit of access to digital objects, through features such as:
 - Zoom and pan functionality;
 - Delivery of parallel versions or related content – for example, text transcripts linked to images or audio;
 - Delivery of synchronised content – for example, musical scores and sound; sound or video recordings and transcripts; and
 - Clear navigation within complex objects or series of objects – for example, ‘page flipping’ in series, folders and items; ‘stitching’ of adjacent maps images.
 - Print copies and download copies for offline use, subject to access and use conditions
117. Use structural metadata to deliver complex objects.
118. Support interactive representations of content views and dimensions, hierarchies, and timelines.
119. Support creation and delivery of themed collections, finding aids and additional descriptive lists at different aggregate levels to aid discovery and use of content.
120. Enable library users to navigate within various contexts for content – for example, within parts of an item; within hierarchical arrangements of content; within themed collections to which an item belongs; to related content.
121. Use published persistent identifiers in delivery systems to support external references to specific objects at various levels from series to tiles on a page.
122. Enable library users to interact with digital objects, through features such as text correction, tagging, annotations, easy linking or embedding in other sites, linking related objects in the collection.
123. Enable user contributed content to be collected and maintained with content.
124. Enable authorised staff to moderate user contributed content.
125. Enable library users to choose alternative delivery or download formats for content, where available.
126. Support e-lending models for controlled temporary distribution of content in accordance with rights conditions to both libraries and individuals.
127. Include descriptive and access rights metadata in content files when they are downloaded.
128. Authenticate library users, where required.
129. Accept secure online payment, where required.
130. Extensibly support incorporation of additional tools for search and delivery of content.

B2.4.3 Deliver – Scenarios

The following scenarios are intended to provide context and illustrate expected complexities to be addressed. They are not exhaustive.

B2.4.3.1 *Specific workflow aspects and variations examples*

The following scenarios highlight aspects of delivery workflows or exceptions to the usual workflows that imply certain system requirements.

Delivery scenario 1

Support restricted delivery: The Library has arrangements with rights holders to restrict delivery of some collection material. For example:

- Manuscripts branch is permitted to deliver an email archive, but only within the Manuscripts Reading Room and for research purposes (no copying for publication);
- Australian Collections is permitted to deliver a published multimedia publication onsite in the Main Reading Room;
- Oral History is permitted to provide copies of an interview to remote researchers who have received specific written permission from the interviewee; and
- Document Supply Service can supply a copy of a whole or part of a book to an individual user under document delivery provisions in the Copyright Act, but the library can only make this digital copy available under certain conditions until copyright expires.

Delivery scenario 2

Apply specific access controls: A newspaper website has been archived by the library. After providing access to the content, the library receives a complaint from a photographer whose images form part of the website content. The photographer had a contract with the newspaper to retain his own copyright in the images. As the material is still in copyright and the photographer does not want the material delivered publicly, access to these images needs to be prevented in the delivery system, while still providing access to articles and images from other photographers.

Delivery scenario 3

Deliver complex objects: Collection content may, by its nature, be complex in structure, relationships or interactions. Mechanisms for representing and delivering the complex nature of content will be required. Examples include:

- Audio recordings of oral history interviews or folk music may include multiple recording sessions and hours of content. To assist with content discovery and navigation, we would like to deliver audio recordings with time-coded TEI XML transcript and/or summary where available. The delivery interface should support searching for keywords in a transcript or summary, and link from text to time-points in the audio;
- Researchers have produced transcripts of digitised manuscript items, such as Cook's journal and Wills' diary. These may be text only (Word document or PDF), or marked-up text (TEI XML transcript). We would like to deliver the transcript with the page images of these items;
- An album in a manuscript collection contains photographs, and some are overlapped. Library users want to view this item as an album in the context of the manuscript collection, and also view the individual photographs. It should be possible to link the individual images and embed a view copy in other websites if the images do not have any rights restrictions;
- A library user in the Main Reading Room wants to view content from a CD-ROM which includes short videos and audio recordings as well as standard text and images. She wants

to be able to view and replay the video content without losing her place in the publication, zoom in on the photos and look at them in detail, and easily read the associated text without losing the context;

- Some of our rare books come in unusual shapes, for example with fold-out pages that display panoramic illustrations. We would like to display most books with a 'flipbook' option, but also need an alternative to display these images effectively. (Example content: <http://nla.gov.au/nla.gen-vn525928-s1-t> and <http://nla.gov.au/nla.aus-vn3945354-s11-y>); and
- Subject to rights conditions, library users should be able to examine e-book publications that include moving images, sound and links to other content.

Delivery scenario 4

Deliver related objects: Collection content can have multiple contextual relationships with other collection content. The delivery system needs to enable these contexts to be presented, navigated and explored by library users. An example is:

- The Library has a sound recording of a musical performance of a piece held in the sheet music collection. A Library photographer took photographs of the recording session. We would like to present the sound recording in the delivery system with an image of the sheet music and images of the performers.

Delivery scenario 5

Enhanced delivery: Digital delivery offers opportunities to enrich and enhance user experience, aid discovery and exploration of content, and provide context.

The delivery system needs to support and enhance user interaction so that researchers can readily get the information they need without visiting the library. Examples include:

- Library users can use 'flip book' representations of digitised books including atlases, journals and newspapers, or flip through publicly accessible digitised contents of a manuscript series online as easily as they would view the physical material. OCR transcripts of textual material are also available to support searching within content;
- Researchers can zoom and pan high resolution images to view detail of maps, aerial photographs, newspaper articles, books with small print or pencil annotations, sheet music, handwritten documents, etc;
- Library users can search all types of collection materials using map-based, bounded-area or place name based searches;
- Researchers can search across archived Web content using free text or URL based searching, can browse content via a range of collection views – for example, themed collections with finding aids; timeline views; 'topic map' based views; – or explore via combinations of searches and collection views;
- Library users can view and interact with a representation of a three dimensional object such as a globe;
- Map users can consult graphical map finding aids and series indexes and view high resolution copies of the relevant maps online, with 'on-the-fly' stitching of adjoining maps and overlays of geospatial data; and
- Where available, library users can choose to consult or download alternative delivery formats – for example: alternative image file formats or resolutions; streaming delivery or

download of content at various sampling rates for sound and video; a range of e-book publication formats; viewing or download of proprietary file formats in the original format or in alternative access copies.

Delivery scenario 6

User interaction: Enabling library users to interact with and contribute to content through correction, annotation or creation of relationships may enhance knowledge of the Library's collections and aid other library users with similar interests. An example is:

- A group of researchers is working on a project using material from the digitised newspapers and books collections. They would like to be able to correct text in the OCR transcripts to assist searching within the material. They also want to make annotations and create a collection of related articles and pages for their own research, but do not want this information to be publicly available.

B3 General Requirements

Should the Library proceed to stage two of the procurement as set out in A2.2, it is intended that the following information will be included within any statement of requirement.

B3.1 Interface Requirements

131. System components of any proposed solution must provide interfaces to the following Library systems:

- Acquisition management system – RefTracker;
- Archival management system for manuscript materials – Archivists' Toolkit;
- Authentication services – Active Directory;
- Equipment and services in the Library's audio conversion facility, for audio preservation, capture, analysis and processing – QUADRIGA AudioCube, DOBBIN;
- Integrated Library Management System (ILMS) – Voyager;
- Interlibrary loan management and document delivery system – Relais;
- Issue tracking system - Atlassian JIRA;
- Online ordering system, for clients to request copies of collection material requiring mediated access - Copies Direct (in-house system);
- Online Public Access Catalogue (OPAC) - based on VuFind;
- Payment services – eWay payment gateway;
- Records Management System – TRIM;
- Rights Management System – Sprightly (in-house system);
- Search and discovery service – Trove (in-house system);
- Storage system – Digital Object Storage System (DOSS) (in-house system); and
- User Management System – in-house system;

132. System components must be capable of using Web Services interfaces for interaction with third-party services.

B3.2 Design Constraints

133. Software components must be Unicode compliant.

134. We have a preference for versions of server-side software components that run on Unix/Linux operating systems.

135. Proposed solutions must offer flexible, extensible modular software architecture with integrated components, aimed at reducing future development costs and improving maintainability.

136. System components supporting delivery must have a web interface to deliver content to end users via the World Wide Web.

137. Systems should support programmatic access to data to allow integration with other systems and development of additional software tools or plugins.

B3.3 Performance Requirements

138. System components supporting searching must return the first page of search results within 2 to 5 seconds for 95% of searches.

139. As indicated in the following table, system components must support the following numbers of concurrent users, and be scalable to support increased user demand.

System components supporting:	Concurrent users supported
Digitisation workflows	50 Library staff users.
Web archiving	30 Library staff and partner agency users.
Deposit on physical format carriers	20 Library staff users.
Online deposit	15 external depositors and Library staff users.
Preservation	20 Library staff users.
Delivery	50 record accesses per second.

140. As indicated in the following table, system components must support the following baseline estimates for material throughput, and be scalable to support increased activity:

System components supporting:	Processing throughput capacity
Digitisation workflows	<p>3-6 million images per year through mass digitisation processes.</p> <p>50,000 to 100,000 images per year through curated digitisation processes.</p> <p>150,000 images per year of on-demand digitisation for client requests.</p> <p>Up to 2500 hours of publication quality audio per year.</p> <p>Up to 2500 hours of publication quality video per year.</p>
Web archiving	1 billion files per year (approx 40 TB)
Deposit on physical format carriers	<p>3000 carriers per year.</p> <p>May include large collections, for example a collection from a photographer containing 900 GB of images.</p>
Online deposit	3000 submissions per year.

141. The system should assume support for at least 1 gigabit per second network interface speed between infrastructure components.

B3.4 Interoperability

142. As indicated in the following table, system components must provide facilities to interoperate via the following standards and protocols, where appropriate:

- Hypertext Transfer Protocol - HTTP 1.1 - RFC 2616;
- HTTP Extensions for Web Distributed Authoring and Versioning - WebDAV - RFC 4918;
- File Transfer Protocol - FTP - RFC 959;
- Web Services technologies as described in Web Services Architecture: W3C Working Group Note 11 February 2004;
- WARC File Format - ISO 28500:2009;
- Extensible Markup Language - XML 1.0;
- Metadata Encoding and Transmission Standard - METS;
- Metadata Object Description Schema – MODS;
- PREMIS Preservation Metadata Schema;
- MARC 21 / MARC XML;
- Encoded Archival Description 2002- EAD 2002;
- ALTO extension to METS for description of layout and content of page objects - ALTO;
- Text Encoding Initiative schemas – TEI;
- Web Content Accessibility Guidelines - WCAG 2.0;
- Lightweight Directory Access Protocol – LDAP - RFC 4510;
- Dublin Core metadata standard; and
- EAC-CPF schema.

B3.5 Usability

143. System components for Library staff shall allow 80% of users to successfully perform 80% of tasks after 8 hours of training.

144. Public users shall find system components supporting delivery intuitive and rewarding to use, and can use the system without mediation.

145. System components shall incorporate iterative design and development, including Usability Testing for user interfaces with representatives of the public, agencies and Library staff.

146. Software components should provide appropriate feedback on user actions, prevent users from error and allow recovery from error.

147. Software components for Library staff should incorporate user hints and help pages, and should provide user documentation and tutorials.

148. Software components supporting delivery for public users shall incorporate user hints and help pages, tutorials and examples of searches.

B3.6 Reliability

149.No system component is permitted to exhibit more than four critical outages each resulting in 100% service unavailability across 100% of the user base in any calendar year.

B3.7 Maintainability

150.Software components must be maintainable such that a competent programmer with the appropriate skills and tools should be able to:

- Diagnose and correct a complex defect without resorting to reverse engineering, inserting debugging code or running long or repetitive test routines;
- Add functionality without breaking the architectural model, or introducing new defects; and
- Redeploy the software without complex manual procedures.

151.Equipment and software components must remain under a maintenance agreement or warranty for the life of the equipment and software components.

152.Vendors must provide services under maintenance agreements for equipment and software components.

B3.8 Integrity

153.Transfer and storage of data must validate and preserve data integrity.

154.Data integrity check results must be logged.

155.Data integrity failures must be detected and reported.

B3.9 Quality

156.User acceptance testing of software components must succeed and detect no more than 3 defects per 1000 lines of code.

157.Software components and development, maintenance and support activities must adhere to a Quality Plan to be developed according to PRINCE2 or ITIL methodologies, the methodology selected being that most appropriate for the software component or activity.

158.Software components must have support for development, testing, integration and operational activities in separate development, test, staging and production environments.

159.For digitisation of image and textual materials, equipment and system components must be capable of producing International Colour Consortium (ICC) colour managed bi-tonal, greyscale and colour images at a minimum of 400 dpi resolution without interpolation as uncompressed or lossless compressed Lempel Ziv Welch (LZW) TIFF files, with de-skewing of page images to within plus or minus 1 degree from parallel.

160.For digitisation of image and textual materials acceptable alternative image formats are RAW, DNG, and potentially in future JPEG 2000.

161.For audio materials, equipment, digital audio output and metadata must comply with standards and quality criteria specified in:

- IASA-TC04: Guidelines on the Production and Preservation of Digital Audio Objects (Second Edition - 2009);

- AES 57 AES standard for audio metadata - Audio object structures for preservation and restoration; and
- Library of Congress - AudioMD – Technical metadata for audio.

162. For video materials, equipment, metadata must conform to standards and quality criteria specified in:

- Library of Congress - VideoMD – Technical metadata for video.

163. Optical Character Recognition (OCR) content must be supplied according to the ALTO XML schema

B3.10 Security

164. All hardware, network infrastructure, communications and software systems and processes for each proposed solution must comply with the requirements of the Australian Government Information and Communications Technology Security Manual (ACSI33) and the objectives of the Australian Government Protective Security Manual.

165. Security of personal data for system components supporting deposit on physical format carriers, online deposit, or delivery must, in addition to the general security requirement, comply with the Privacy Principles of the Privacy Act 1988 (Cth).

B3.11 Support

166. Final selected suppliers, if any, must provide timely services under acceptable maintenance agreements for equipment and software components.

167. Final supplier maintenance agreements must include staff training and certification, where applicable.

B3.12 Sustainable Solution

168. Final selected suppliers will be required to use sound environmental management practices and where applicable, confirm that they will comply with Australian Government environmental policies. These will be detailed within any subsequent tender documentation.

B4 Timeframe

The Library expects to have negotiations with the successful Respondent(s) completed by **mid-December 2011**. The following timetable indicates anticipated milestones and key dates which the Library will aim to achieve.

Milestones	Key Dates
Preferred Supplier(s) selected	December 2011
Contract(s) signed	December 2011/January 2012
Initial release implemented	December 2012
All remaining stages implemented	December 2015

Initial release is to consist of functionality to be agreed between the Library and the selected supplier(s), prior to contract signing, to be implemented in the production environment.

B5 Background information - National Library of Australia Digital Library Environment

B5.1 Collection materials to be managed

The Library's digital library environment manages the following range of materials:

Pictorial, Manuscript or Print Material/Formats:

- Materials digitised from physical material:
 - Individual pictures, photographs, sheets, maps;
 - Sided materials - e.g. letters; and multi-view items - e.g. realia, from different angles;
 - Multipart items - books, albums, multi-volume works;
 - 'Tiled' items - extensive maps, banners;
 - Items within other items - map inserts in books, illustrations, plates; and
 - Details within individual items - 'detail of left hand'.
- Born-digital images.

Audio Material/Formats:

- Audio materials:
 - Preservation quality transfer to WAV plus conversion to BWF, from a range of analogue media (tape reel, cassette, grooved disc);
 - Preservation quality transfer to BWF from digital audio tape (DAT) and compact disc;
 - Born-digital audio recordings (WAV) with conversion to BWF, from live studio and HDD field recordings; and
 - Delivery derivatives - MP3.
- Text equivalents:
 - Summaries (with or without time-points); and
 - Transcripts (with or without time-points).

Web archiving (Whole Domain Harvest) Formats:

- Masters:
 - (W)ARC containers - contain virtually every format ever published to the Australian web.
- Derivatives:
 - CDX (simple indexes of the WARC's by URL and timestamp);
 - DAT (extracted metadata for researchers); and
 - Nutch full text search indexes.

Web archiving (Selective) Formats:

- Web content accessible to the crawling software on the live Web - consolidated in ZIP, TAR and later ARC or WARC files.
- Web content received in a package from a supplier - received 'unpacked' (as files and directories) or consolidated in ZIP, TAR, ARC or WARC files.
- Documents retrieved or received for archiving - received as individual files.

Digitised Newspapers Formats:

- Scanned images: currently a pair of bi-tonal and greyscale TIF for each newspaper page; single colour TIF for Australian Women's Weekly.

- Derivatives for QA of scanned images: bi-tonal JPEGs for newspapers; colour JPEGs for Australian Women's Weekly.
- Derivatives for QA of content analysis/OCR: de-skewed bi-tonal JPEGs for newspapers; de-skewed colour JPEGs for Australian Women's Weekly.
- Derivatives for delivery: de-skewed greyscale JPEGs for newspapers; de-skewed colour JPEGs for Australian Women's Weekly.
- Article text extracted from deliverables from content analysis/OCR contractor. Currently this is held in a database, along with corrections added by users of the public search and delivery system (Trove).
- May also want to keep the original deliverables from the content analysis/OCR contractor, i.e. METS XML file for each issue, ALTO XML file for each page. Currently this is held in a working area after the data is extracted -long term storage has not been determined.

Physical Material Digital Formats:

- Content received on physical carriers - CD, DVD, BluRay, USB drives, HDD and other block devices, Zip, Jaz and SyQuest discs, 3.5 inch and 5.25 inch floppy discs.
- Media images of content - CUE/BIN, TOC/BIN, IMG, ISO files (+ DMF, DMG files).
- Single files.
- Metadata generated by JHOVE and New Zealand's Metadata Extractor.
- In future we will need to support archive files (TAR, ZIP, GZIP, BZIP2, WARC) and directories of files.

B5.2 Existing Digital Library Systems

The Library uses the following systems for various workflows:

- **Digital Collections Manager (DCM)** - developed in-house - provides integrated collection management and delivery facilities for digital pictorial and audio material;
- **Newspapers Content Manager (NCM)** - developed in-house – a new workflow and management system to support the mass digitisation of Australian newspapers and journals;
- **PANDAS/PANDORA** - developed in-house – management and delivery systems for selective web archiving using HTTrack crawler (open source);
- **Prometheus** - developed in-house – workflow management system to manage collection of digital materials on physical carriers, including both published materials and unpublished manuscripts in digital form. Processes include transferring data from physical carriers to preservation digital mass storage, media imaging, file identification and metadata extraction. Has no delivery service;
- **Whole Domain Harvest Service** – outsourced service – provided by Internet Archive using Heritrix crawler;
- **Voyager** - commercial software – Integrated Library Management System used as the source of bibliographic metadata;
- **VuFind** - open source – customised extensions – Catalogue Service;
- **Archivist Toolkit** – open source – customised extensions - provides integrated support for the archival workflow, including appraisal, accessioning, description, arrangement, publication of finding aids;
- **QUADRIGA Audio Archival System** – commercial software - QUADRIGA DOBBIN is used for audio file analysis, BWF creation (including metadata merge and output metadata) and MP3 derivative generation and QUADRIGA DAT module for audio file analysis and BWF creation from DAT format (including metadata merge and output metadata);
- **Trove** - developed in-house – Search and Discovery Service; and

- **Rights Management system** – developed in-house – stores rights related information for an item. Does not manage access or permissions.

The Library currently does not have any system for online deposit of born digital material, preservation planning and preservation risk management functions.

The outcome of the project will replace DCM, NCM, Pandas and Prometheus management systems, incorporate and streamline ingest of Whole Domain Harvest data/ metadata and fill the current gaps in functionality, including community deposit of born digital material and preservation functions.

Current main functions of Digital Collections Manager (DCM)

- Create and manage records for intellectual works and copies, their hierarchies and relationships;
- Associating items to control records;
- Managing work and copy records, versions, access restrictions, process histories;
- Managing users, tools, projects;
- For audio material: generating XML metadata for BWF creation; Managing consignment, manifests and import;
- File management and ingest;
- Masters download; and
- Statistical reporting;

Current main functions of Newspapers Content Manager (NCM)

- Create and manage records for intellectual works (titles, issues, pages, articles, illustrations) and copies;
- Assign persistent identifiers;
- Create and manage relationships between intellectual entities: titles, issues, pages, articles, illustrations;
- Create and manage metadata for intellectual entities: titles, issues, pages, articles and illustrations;
- Manage original OCR text of articles and manage updates (corrections) to the text;
- Create and manage access restrictions for articles;
- Create and manage contributors, work orders (i.e. specified units of work associated with contracts for scanning and content analysis/OCR), scan jobs, OCR jobs, incoming and outgoing tapes;
- Manage consignment batching (for content analysis/OCR contractors), manifest creation, metadata import on return;
- Support QA of scan jobs and OCR jobs, including sampling based on job size and accuracy calculation based on errors recorded by QA staff;
- Ingest digital items and metadata to preservation storage; ingest derivatives to derivative storage;
- Manage derivative creation, import and ingest;
- Support multiple master types and functions (e.g. pairs of bi-tonal and greyscale TIF images; colour TIF; article text);
- Support search (within workflow application);
- Support indexing by public search and delivery system;
- Support delivery of derivatives and metadata - including navigation of hierarchies, contextual display with descriptive metadata, concurrent delivery of text and scanned pages.

- Support interactive delivery of articles e.g. text correction, tagging, comments;
- Master download;
- Manage queues for back-end processes e.g. derivative creation, restoration of masters to online; and
- Statistical reports.

Whole domain harvest functional requirements

- Material is harvested by a contractor and delivered as a large number of physical hard drives containing ARC files. They also provide Nutch indexes (Lucene segments files);
- Large batch ingest of delivered ARCs (20TB annually) which gets converted to WARC and fixity checks against checksums supplied by contractor are performed;
- Reports are generated on Mime types of WARC records, number of different types of domains (.gov.au, .edu.au, .com.au etc);
- Delivery to web using Wayback tool. It needs either file system level access to (W)ARC or HTTP; and
- The Library uses METS / MODS, ALTO and MARC to transfer metadata between various systems but metadata is extracted and stored in internal formats in databases.

B5.3 Metadata

A variety of metadata is kept, some supplied by digitisation contractors, some automatically generated and some hand generated.

- Descriptive (bibliographic) metadata;
- Relationship metadata: structural metadata e.g. parent-child relationships, and provenance relationships e.g. between masters and derivatives;
- Technical metadata about files e.g. bit rate, resolution, file size, file format, checksum, colour space;
- Provenance metadata e.g. scanning hardware and software, derivative creation processes, harvest logs; and
- Rights metadata e.g. rights holders, publishers, permissions, intellectual rights.

Not all types of metadata are stored in all Library systems. More metadata could be required to be kept in future, for instance it has been recognised that more technical metadata is required to inform preservation planning processes.

Standard metadata schemes are used mainly for transfer of metadata, for example importing to ingest systems and exporting to delivery systems. Some standards used are MARC, METS, MODS, ALTO, MIX (e.g. image metadata in JHOVE). Generally metadata is extracted from the standard formats and stored in an internal format in the database. However at least some of the transfer files will be kept in the repository e.g. METS and ALTO files supplied by the OCR contractor for digitised newspapers.

Section C – Evaluation and Selection

C1. Evaluation process

- C1.1 Following the Closing Time, the Library will conduct an initial review of the Proposals submitted. A Proposal **will not** be considered further if the Library determines, in its sole and absolute discretion, that the Proposal:
- (a) does not meet the Conditions for Participation set out in clause C2; or
 - (b) subject to clause E4 (Errors of form), does not provide sufficient information either in the form of Section D (Proforma Proposal) or in another format so as to allow the Library to evaluate it against the other evaluation criteria.
- C1.2 The Library may also exclude a Proposal, in its sole and absolute discretion, if:
- (a) it is submitted after the Closing Time;
 - (b) it contains alterations or erasures;
 - (c) it contains prices or other information which are not clearly and legibly stated;
 - (d) it contains false or misleading information or statements;
 - (e) the Respondent, or a director or officer of the Respondent, is insolvent or bankrupt;
 - (f) the Respondent has an actual, potential or perceived conflict of interest that cannot be managed to the satisfaction of the Library acting in its sole and absolute discretion;
 - (g) it does not address all evaluation criteria; or
 - (h) there has been a significant deficiency in the Respondent's performance of a substantive requirement or obligation under a prior contract.
- C1.3 The Library also reserves the right, in its sole and absolute discretion, to reject at any time any Proposal where in the opinion of the Library:
- (a) the preparation or lodgement of the Proposal involved collusive or anti-competitive behaviour or similar unlawful conduct with any other Respondent or any other person;
 - (b) the Respondent, or its officers, employees, agents or advisers, offered inducements in connection with the preparation of their Proposal;
 - (c) the Proposal was compiled with the improper assistance of employees, ex-employees, contractors or ex-contractors of the Commonwealth or the Library; or
 - (d) the Proposal utilised information that was unlawfully obtained from the Commonwealth or the Library.

C1.4 The Library will evaluate the remaining Proposals on the basis of the evaluation criteria below. This evaluation process may include, in the Library's sole and absolute discretion, further clarifications from or discussions with Respondents.

C2. Conditions for Participation

C2.1 Respondents will be excluded from evaluation if they do not meet, or no longer meet, the following Conditions for Participation:

- (a) the Respondent and its nominated subcontractors must not be named as non-compliant with the *Equal Opportunity for Women in the Workplace Act 1999*; and
- (b) the Respondent and its proposed subcontractors must not have unpaid claims against them in respect of a judicial decision made relating to employee entitlements, other than those for which the Respondent or subcontractor has lodged a formal appeal that has not yet been determined.

C3. Evaluation criteria assessment

C3.1 In assessing Proposals that have met the conditions for participation, the Library will have regard to indicative prices (see clause C3.3) and the following evaluation criteria (which are listed in order of priority):

- (a) extent of the Respondent's understanding of the matters involved and whether the Respondent offers a sound approach to the Services required taking into account any risks;
- (b) proven capacity, in particular the Respondent's knowledge, proven expertise, demonstrated responsiveness and reliability in the Services, including past experience and the experience and qualifications of the Respondent's key personnel, including specialist qualifications if appropriate; and
- (c) insurance held (public liability, workers compensation and professional indemnity).

C3.2 Weightings will be determined by the Library in its sole and absolute discretion.

C3.3 Following the evaluation of Proposals against the evaluation criteria in clause C3.1, the Library will then assess the Proposals taking into account the indicative prices and the Proposal's score against the weighted criteria. Notwithstanding circumstances described in clause A2.3, the Library's intention is to short list based on how well Respondents meet the criteria.

C4. Debrief

C4.1 All Respondents will be offered the opportunity for a debrief. This debrief will provide feedback on the relative merits of their Proposal and will occur at the conclusion of the RFP and RFT processes (ie signing of a contract with the preferred Respondent, or termination of the RFP process, or a decision not to proceed with a RFT).

Section D – Proforma Proposal

Respondents are to ensure that they provide sufficient information in their Proposals to enable the Library to undertake a full evaluation of their Proposal.

The following comprises a Proforma Proposal to assist Respondents in providing a satisfactory Proposal. It is not, however, intended to be exhaustive, and Respondents must ensure that their Proposal addresses their capacity to meet the Conditions for Participation set out in **clause C2** and all other evaluation criteria.

Respondents are encouraged to propose innovative ways of delivering the outcomes which the Library wishes to achieve. In particular, Respondents may:

- (a) include in their Response more than one option for the Library to consider; or
- (b) include alternative Proposals (in addition to any Proposal that complies with the requirements set out in this RFP).

Proforma Proposal

Name of Respondent.....**Respondent's ABN/ACN**.....

Respondent's Address.....

Respondent's Representative:

(Name).....**(Telephone)**.....**(Fax)**

(Email).....

Note: Any questions that may arise during the evaluation of Proposals will be directed to the Respondent's Representative named above.

Conditions for Participation

Respondents must confirm that neither they nor their proposed subcontractors:

- (a) are named as non compliant with the *Equal Opportunity for Women in the Workplace Act 1999* (Cth); or
- (b) have any unpaid claims against them in respect of judicial decisions relating to employee entitlements, other than those for which the Respondent or subcontractor has lodged a formal appeal that has not yet been determined.

Criterion (a) - Extent of the Respondent's understanding of the matters involved and whether the Respondent offers a sound approach to the Services required taking into account any risks

Respondents are to provide here a solution to the Service requirements set out in the Statement of Outcomes at **Section B**, including an indication of how the Respondent will deliver the outcomes, a risk assessment of the proposed solution and whether any facilities or assistance will be expected from the Library, within the proposed timeframe.

Criterion (b) – Proven capacity, in particular the Respondent's knowledge, proven expertise, demonstrated responsiveness and reliability in the delivery of the Services, including past experience and experience and qualifications of the Respondent's key personnel, including specialist qualifications if appropriate

Respondents must provide here details of similar projects or engagements of similar size or complexity to the stated requirement. The following information must be provided at a minimum for each project:

- (a) project description;
- (b) client organisation;
- (c) referee contact;
- (d) date of commencement and period of association;
- (e) whether the services were delivered within the required timeframe; and
- (f) brief description, including services provided, number of personnel involved in the delivery of the services, geographical location, skills that were required, lessons learnt and how these experiences will be applied to the Library.

All relevant experience will be considered, whether or not that experience is with the Australian Government.

Respondents must also provide here details of key personnel to be involved in the project, including their relevant experience. Personal references must also be provided for proposed individuals.

Criterion (c) - Insurance held (public liability, workers compensation and professional indemnity)

Respondents must provide here details of insurance held (public liability, workers compensation and professional indemnity).

Pricing

Respondents must provide here their indicative pricing in relation to the Proposal. Pricing or price ranges should be such that it enables an understanding of the elements that constitute any total amount. Pricing is to be inclusive of GST but should also indicate the amount of GST payable as a separate amount.

Further Information (If any)

Respondents are to provide here any further information pertinent to their Proposal, including any further information required by this RFP (if applicable). In particular, Respondents should:

- (a) identify any information they will be providing to the Library, either during this RFP process, during contract negotiations or during the currency of any contract that is in operation between the parties, that they consider should be protected as confidential information, and provide reasons why this information should be protected as confidential information. The Library will consider these reasons in the context of its accountability and reporting obligations but does not guarantee to keep information confidential, notwithstanding that confidentiality has been claimed by a Respondent;
- (b) identify and declare in their Proposal any conflict of interest or potential conflict of interest that may apply to their responding to this RFP or providing the Services; and
- (c) describe the procedures they intend to implement for dealing with any known or potential conflict of interest that may arise in the provision of the Services.

.....
Signed for and on behalf of []

.....
Dated

Section E – Other Terms and Conditions

E1. Definitions

Business Days means days that are not a Saturday, Sunday or public holiday in Canberra.

Contract means the contract proposed to be entered into between the Library and the preferred Respondent or Respondents, pending negotiations.

Government Agency means the National and State Libraries Australasia (ie apart from the National Library of Australia also the Australian Capital Territory Library and Information Services; Library Board of Victoria; Library Board of Western Australia; Northern Territory Library; Library Council of New South Wales; Library Board of Queensland; Libraries Board of South Australia; The Crown in Right of the State of Tasmania; and National Library of New Zealand) and other Australian government bodies including departments and agencies subject to the Financial Management and Accountability Act 1997 (Cth) and Commonwealth authorities and Commonwealth companies subject to the Commonwealth Authorities and Companies Act 1997 (Cth).

Proposal means the Proposal to be submitted by Respondents in response to this RFP.

Request for Proposal or **RFP** means this document.

Request for Tender or **RFT** means the document the Library intends to issue if it proceeds to a second stage.

Services means the services set out in **Section B** (Statement of Outcomes).

E2. Clarifications

The Library reserves the right to forward any clarifications issued by it of the meaning of the content of this RFP to all known Respondents on a non-attributable basis.

E3. Addenda

The Library may issue addenda to this RFP in its sole and absolute discretion. All conditions of this RFP will apply to any addenda unless modified in the addenda. The Library will issue the addenda:

- (a) to all the potential suppliers that are participating at the time the information is amended, if known, and in all other cases, in the same manner as the original information; and

- (b) in adequate time to allow potential suppliers to modify and, if required, to re-lodge their initial Proposals.

E4. Errors of form

- (a) Without limiting the Library's other rights in this RFP, if, between opening a Proposal and making a decision on the outcome of this RFP process, there is discovered what appears to be an error of form in relation to one of the documents submitted as part of the Proposal, the Library may, in its sole and absolute discretion, allow the Respondent to correct that error of form.
- (b) If the Library provides any Respondent with the opportunity to correct errors of form after the Closing Time, it will provide the same opportunity to other Respondents who are in the same position.

E5. Intellectual property rights in the RFP

- (a) All intellectual property rights that exist in the information contained in this RFP or any related or attached material remain the property of the Library.
- (b) Each Respondent is permitted to use this RFP for the purpose only of compiling its Proposal and, in the case of the preferred Respondent(s), for negotiating the Contract with the Library.

E6. Intellectual property rights in Proposals

Any intellectual property rights that exist in a Proposal will remain the property of the Respondent. However, the Respondent licenses the Library, its officers, employees, agents and advisers to copy, adapt, modify, disclose or do anything else necessary to all material (including material in which the Respondent or other persons have intellectual property rights) contained in the Proposal for the purpose of:

- (a) evaluating/clarifying Proposals;
- (b) negotiating any Contract with a Respondent;
- (c) including the material in any RFT;
- (d) managing a Contract with the successful Respondent (if any);
- (e) audit, government and Parliamentary reporting requirements; and

- (f) responding to any disputes about this RFP process or in response to any requests from Parliament or a Parliamentary Committee.

E7. Right to conduct security, probity and financial checks

The Library reserves the right to perform such security, probity and/or financial checks and procedures as the Library, in its sole and absolute discretion, may determine are necessary in relation to the Respondent, its partners, associates or related entities, including consortium members and their officers or employees. The Respondent agrees to provide, at its cost, all reasonable assistance to the Library in this regard.

E8. Conflict of interest

The Library reserves the right to deal with a conflict of interest as it sees appropriate, and may cease further consideration of a Proposal in such circumstances.

E9. Late Proposal policy

- (a) Late Proposals (including corrections or additional information lodged after the Closing Time) may be admitted or excluded from evaluation in the Library's sole and absolute discretion.
- (b) The Library may ask for an explanation and reasons for the lateness of the Proposal, including whether there has been any mishandling by the Library or postal services.
- (c) Without limiting the Library's discretion to admit late Proposals in other circumstances, the Library will not penalise any Respondent whose Proposal is received after the Closing Time if the delay is due solely to mishandling by the Library. Mishandling by a courier or mail service engaged by a Respondent to deliver a Proposal does not constitute mishandling by the Library.
- (d) The Library is not required to give reasons for its decision whether or not to consider a late Proposal.

E10. Negotiations

Without limiting its other rights under this RFP, in the event that the Library concludes that during final negotiations the preferred Respondent has retracted, or attempts to retract, any part of its proposal, the Library, in its sole and absolute discretion, reserves the right to:

- (a) accept the Respondent's Proposal;
- (b) re-enter negotiations with one or more Respondents (including or excluding the preferred Respondent);

- (c) terminate this RFP process; or
- (d) exercise any other right reserved to the Library under law or elsewhere in this RFP.

E11. Other Library Rights

Without limiting its rights at law or otherwise, the Library reserves the right in its sole and absolute discretion at any time to:

- (a) change the structure and timing of the RFP process;
- (b) vary or extend any time or date in this RFP (including the Closing Time), at any time and for such period, as the Library considers appropriate, provided that if the Library extends the Closing Time, or permits the lodgement of new submissions, the new time limit will apply equitably to all participating potential suppliers;
- (c) otherwise change the timing of this RFP process;
- (d) terminate, suspend or vary this RFP process or any part of it;
- (e) terminate any negotiations being conducted at any time with any Respondent;
- (f) require additional information or clarification from any Respondent or anyone else or provide additional information or clarification, provided that the Library treats Respondents who are in a similar position in the same way;
- (g) seek presentations by some or all of the Respondents;
- (h) visit some or all Respondents' sites;
- (i) have discussions following the Closing Time with some or all Respondents;
- (j) have discussions with, and make visits to, customers or former customers of the Respondent and its subcontractors, whether or not those customers are listed as referees in the Respondent's Proposal;
- (k) negotiate with any one or more Respondents and allow any Respondent to change its Proposal or make a supplementary offer in respect of any changes to the originally stated requirements. The Library reserves the right to enter into any such discussions and negotiations at its absolute discretion (which includes dealing with any Respondent as it deems fit without the need to correspond with other Respondents during this post-Proposal period);
- (l) call for new Proposals, provided that any new time limit for lodging new Proposals applies equitably to all participating potential suppliers;

- (m) after it has terminated this RFP process or any part of it, negotiate with any person who is not a Respondent and enter into a Contract in relation to this RFP process with that person on such terms as the Library accepts without prior notice to any other Respondent;
- (n) add to, alter, delete or exclude any Services to be acquired by the Library, provided that the Library complies with the procedure in **clause E3** for issuing addenda;
- (o) publish or disclose the names of Respondent(s) (whether successful or unsuccessful);
- (p) allow or not allow a related body corporate to take over a Proposal in substitution for the original Respondent;
- (q) not proceed to an RFT stage with one or more Respondents; or
- (r) decide to sole source with one or more Respondents for all or part of the Services without proceeding to a formal RFT.

E12. Respondent acknowledgements

Every Proposal is made on the basis that the Respondent acknowledges that:

- (a) it examined this RFP and any other documents referenced or referred to in this RFP, and any other information made available in writing by the Library to Respondents for the purposes of submitting a Proposal;
- (b) it sought and examined all necessary information which is obtainable by making reasonable enquiries relevant to the risks and other circumstances having effect on its Proposal;
- (c) in lodging its Proposal it did not rely on any express or implied statement, warranty or representation, whether oral, written or otherwise made by or on behalf of the Library, its officers, employees, agents or advisers other than any statement, warranty or representation contained in this RFP;
- (d) it satisfied itself as to the correctness and sufficiency of its Proposal; and
- (e) it is responsible for all costs and expenses related to:
 - (i) the preparation and lodgement of its Proposal;
 - (ii) any subsequent negotiation; and
 - (iii) any other action or response in relation to this RFP.

E13. No contract or undertaking

- (a) This RFP, and any conduct or statement made prior to or subsequent to the issuance of this RFP, is not, and this RFP is not and must not be deemed to be:
 - (i) an offer to contract; or
 - (ii) a binding undertaking of any kind by the Library.
- (b) The Library is not liable to any Respondent on the basis of any contractual, quasi-contractual or restitutionary grounds whatsoever or in negligence as a consequence of any matter or thing relating or incidental to a Respondent's participation in this RFP, including, without limitation, instances where:
 - (i) a Respondent is not engaged to provide the Services;
 - (ii) the Library varies or terminates this RFP process or any negotiations with a Respondent; or
 - (iii) the Library exercises or fails to exercise any of its other rights under or in relation to this RFP.

E14. Government requirements

Respondents should note that the Library is a Commonwealth statutory authority and that it seeks to comply with Commonwealth requirements. In particular, Respondents should be aware that:

- (a) rights to access information in, and related to, documents in the possession of the Library exist under a range of legislation, including the *Freedom of Information Act 1982* and the *Ombudsman Act 1976*. Parliament, parliamentary committees and the courts also have legal rights to access a wide range of information;
- (b) information submitted in relation to this RFP process may be subject to review by the Auditor-General; and
- (c) the Library may provide information to and seek advice from Commonwealth departments, agencies, the Minister and the Library's advisers in connection with this RFP. Information may also be provided to Parliament or Parliamentary committees.

E15. Notices to Respondents

The Library may deliver any written notification or response to a Respondent by forwarding it to the address specified in their Proposal.

E16. Applicable law

- (a) The law applying in the Australian Capital Territory applies to this RFP.
- (b) Each Respondent must comply with all relevant laws in preparing and lodging its Proposal and taking part in this RFP process.
- (c) Respondents should obtain their own advice on the impact of all relevant legislation on their participation in this RFP process.