Report of the METS Profile Development Project

Judith Pearce (National Library of Australia)
David Pearson (National Library of Australia)
Megan Williams (National Library of Australia)
Scott Yeadon (APSR)

http://www.apsr.edu.au/nla-mets/mets.pdf

November 2007

Table of Contents

Introduction

This document reports on the outcome of the APSR METS Profile Development Project. The APSR METS profile project aimed to develop:

- an open, extensible and standard way of packaging metadata for digital objects which could be relevant to both Australian and broader contexts; and
- a generic, repository independent metadata submission and exchange profile for use among APSR repositories.

The basis for this project was the draft METS exchange profile developed by the NLA as part of the 2006 APSR-funded PRESTA project. The project undertook to test this profile against actual implementations and then to register it with the Library of Congress.

Document structure

This document provides more detail on the background and context for the project and then looks at outcomes by answering the following questions:

- What did we want to do?
- How did we try to do it?
- What did we actually do?
- What changes did we make to the draft profile?
- What lessons did we learn?
- What needs to happen next?

Background and context

In 2006 a draft profile for exchange of digital objects between repositories was developed as part of the APSR/NLA PRESTA project (PREMIS Requirement Statement Project report). This profile used METS (Metadata Exchange Transmission Standard) as the packaging format.

METS is a framework standard that enables metadata describing an object and its structure to be recorded in a document that can be used as a Submission Information Package (SIP) or Dissemination Information Package (DIP) in digital object management and delivery scenarios. It is extensible by plugging in various other extension schemas such as MODS (Metadata Object Description Schema) for resource description, MIX (Metadata for Images in XML) for still image technical metadata and PREMIS (PREservation Metadata Implementation Strategies) for provenance and fixity. This makes it a very powerful packaging format but also difficult to implement because of its complexity.

A profile is a way of defining best practices for implementing a standard or set of standards (as in this case) within a given community by specifying the requirements that conformant implementations must satisfy. A good profile addresses the range of usage scenarios and content models requiring to be supported to ensure interoperability within that community and provides both entry level and extended options for conformance.

The PRESTA work broke new ground internationally by specifying how PREMIS metadata might be implemented as part of a METS document to enable the transfer of an object from one repository to another with its full history. A self-describing object is important in the Australian repository environment because of the need for

research data to be long-lived and able to be shared as part of an Australian “data commons”.

Based on this work, the Australian National University and the University of Queensland were able to implement demonstrators for exchanging content between a DSpace repository and a Fez-Fedora repository. This proved that METS could be used to support the transfer scenario but also exposed the need to test the profile against a range of data content models and also against submission and delivery scenarios.

What did we want to do?

The 2007 APSR/NLA METS Profile Development project was an APSR RIFF project (Repository Interoperability Framework Project - NLA RIFF P1) which aimed to facilitate the adoption of the Metadata Encoding and Transmission Standard (METS) as a common format for exchanging data between repositories and other workflow environments.

To do this we wanted to test and refine the PRESTA profile against real data and usage scenarios so that it could be registered as a core generic METS profile with the Library of Congress. We also wanted to assist the RIFF project partners to develop implementation-specific profiles for use in their specific repository workflows.

The RIFF project partners were the Australian National University, National Library of Australia, University of Queensland and University of Sydney. The specific repository workflows to be tested against the profile are documented in Appendix A.

APSR project managed all of the deliverables as part of the RIFF project.

How did we try to do it?

We started by incorporating the changes to the PRESTA draft profile based on the demonstrators created by ANU and UQ. Then we held meetings with various project partners in early February, followed by a workshop to gather requirements. The workshop used a diagrammatic representation of the METS schema to assist with a shared understanding of the data model and elements needing to be recorded (Appendix B). We asked project partners to identify usage scenarios, specific requirements for extension schemas and controlled vocabularies. We also asked them to provide a description of the structure of their content and sample METS documents using the enhanced PRESTA draft profile for guidance.

We appointed an editor to gather and incorporate changes to the draft profile based on the information provided by the partners through the workshop and subsequent exchanges. The editor also surveyed best practices across the METS community. We also identified a number of different models for how the generic profile could work in relation to the implementation profiles. These are documented in Appendixes C & D.

- Model 1 consisted of a core profile registered with the Library of Congress (LC) with related content model specific sub-profiles, also registered with LC that in fact would have been implementation profiles.
- Model 2 consisted of a core profile registered with the Library of Congress (LC) with related content model specific sub-profiles, also registered with LC and a third set of implementation profiles that in fact were more like contracts with suppliers and consumers to meet the needs of specific scenarios.

Model 1 would have resulted in a proliferation of implementation profiles dealing with similar content and usage scenarios in different ways. In a way, it treated each

implementation as a separate content model and did not recognise the ontological relationships between content of a similar type.

Model 2 required the development of generic content models that could then be inherited by implementation profiles. The task of doing the analysis to identify the range of content models needing to be supported across the Australian "data commons" appeared to be daunting and out of scope of the project as originally defined and resourced.

In both models we struggled with how to deal with specific scenarios. It seemed as though separate implementation profiles would be required to support submission and dissemination scenarios and that new implementation profiles might need to be registered whenever there was a change in provider or consumer or underlying technology. Therefore, a model was needed that could minimise the number of registered profiles while enabling repository managers to establish contracts with providers and consumers and to update their technologies.

What did we actually do?

What did we test against?

The draft profile ended up being tested against five different workflows: digitised newspapers (National Library of Australia) and digital journals, conferences, still images and fieldwork outputs (University of Sydney). The other RIFF project partners attended initial workshops but their implementations were not sufficiently advanced for testing within project timeframes.

The profile was also tested against a number of other content models within the National Library of Australia. These included digitised still image and audio content, websites and physical format digital publications. The Library's experience with these content models informed development of the profile but resources were not available to generate valid METS documents based on existing implementations. Examples therefore needed to be hand-coded.

A detailed study was also conducted of profiles already registered with the Library of Congress to compare and contrast approaches. This highlighted the need for a scalable and modular approach that could be applied to a range of workflows and content models.

What did we submit?

A generic Australian METS Profile was registered with the Library of Congress at http://www.loc.gov/mets/profiles/00000018.xml as the first step in implementing a three-layered model for using METS in a standards-based service-oriented way to describe objects in Australian repositories. A diagrammatic representation of the Australian METS Profile Version 1 is presented in Appendix E.

Figure 1 below represents the three-layered model in diagrammatic form:

- At the first level the Australian METS Profile 1.0 operates as a generic profile detailing the requirements to produce a self-describing document of any type, with the capability to support submission, delivery and transfer scenarios.
- At the second level an evolving set of content-model specific sub-profiles will detail the rules for describing content of a given genre based on test implementations, e.g. journals tested against an Open Journal System (OJS) implementation; newspapers tested against the NLA Newspaper Digitisation Project.

- At the third level a set of implementation profiles will detail exception extension schemas, controlled vocabularies (mainly for descriptive metadata) and technical requirements.

Figure: 1: Three-layered Model for using the Australian METS Profile

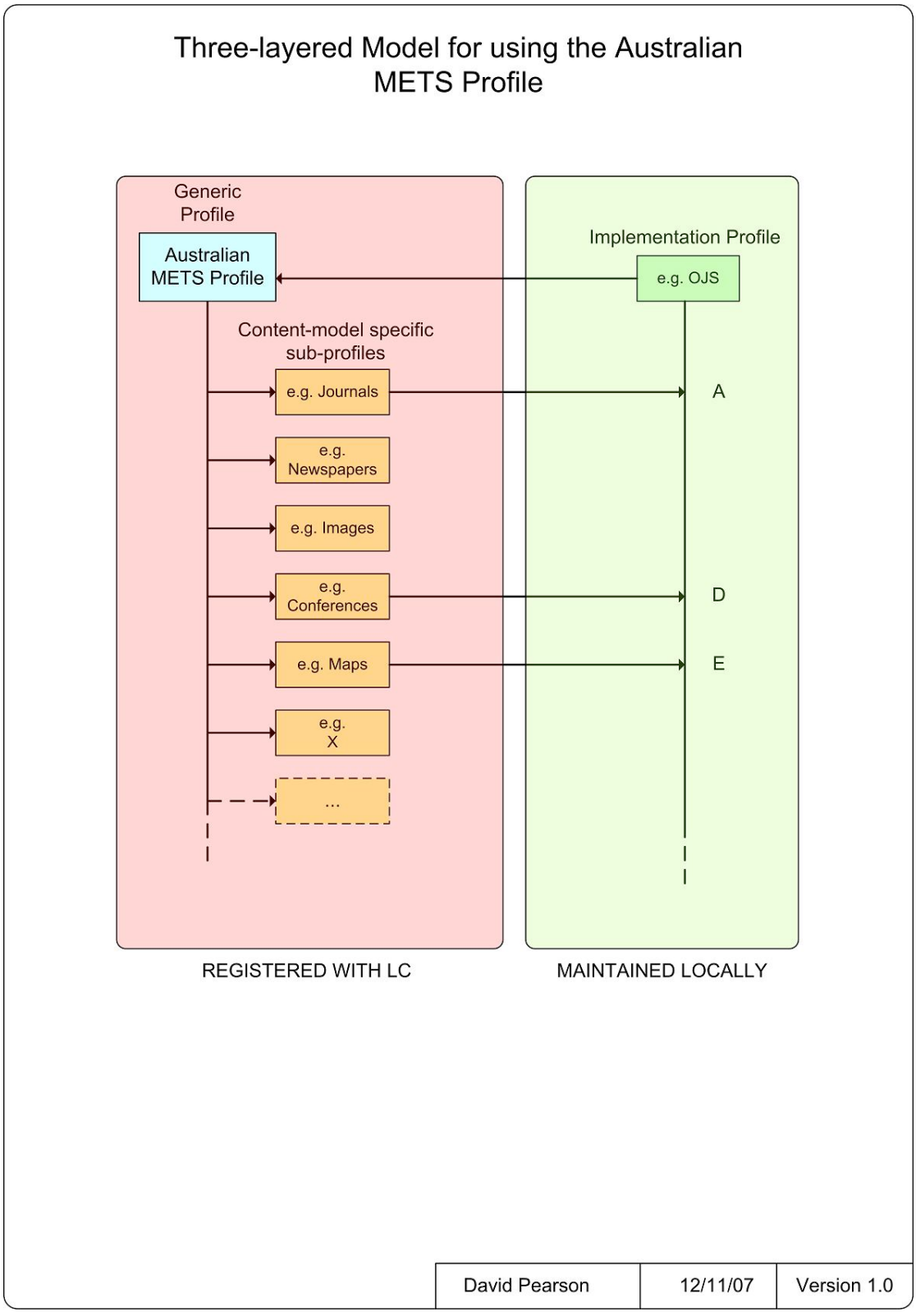
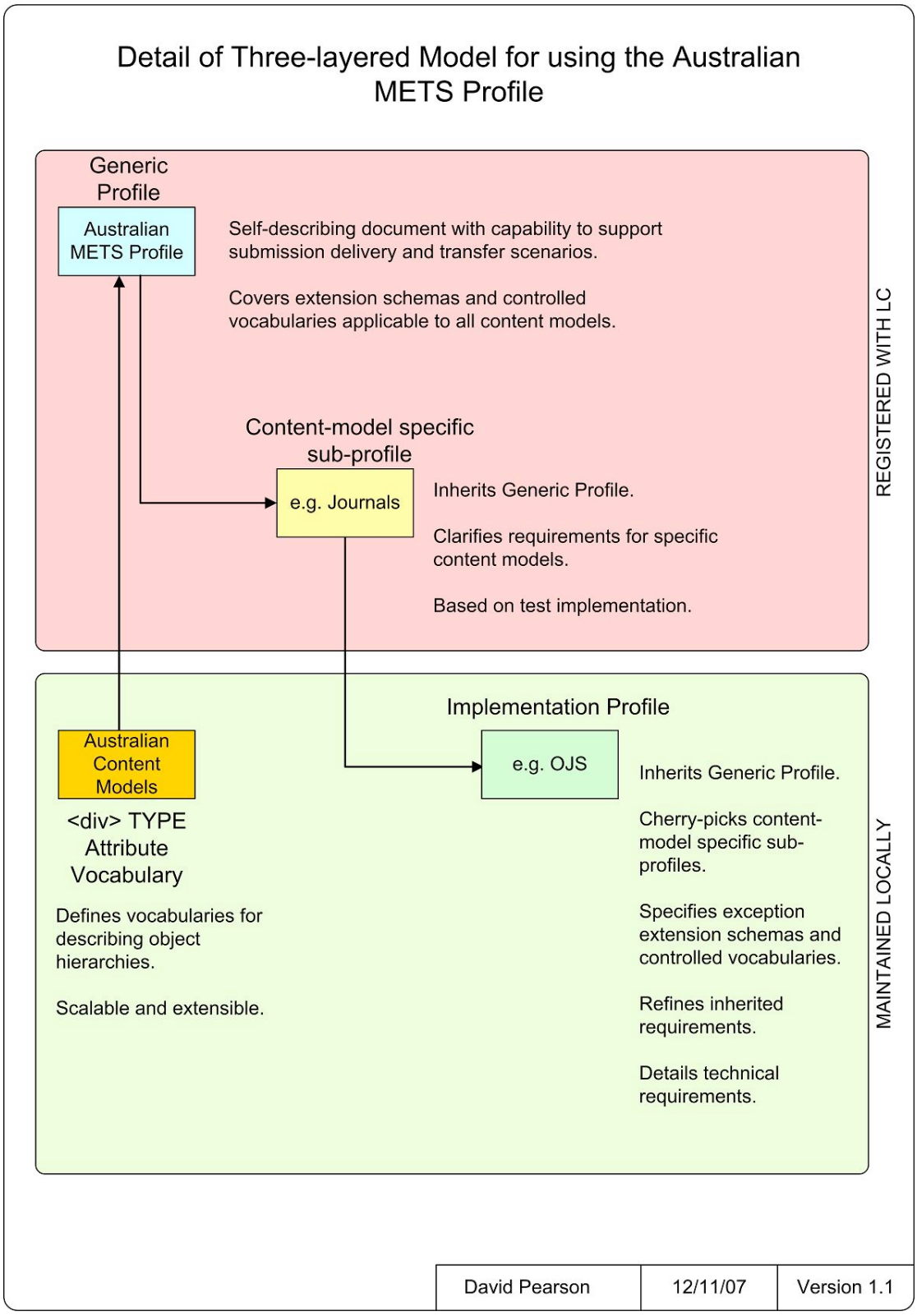


Figure 2 is a more detailed view of the model showing the relationship between the generic profile and instances of the content-specific sub-profiles and implementation profiles.

Figure: 2: Detail of Three-layered Model for using the Australian METS Profile



To support the content-model specific sub-profiles a controlled vocabulary was developed for the <structMap> TYPE attribute. This vocabulary was submitted to the Library of Congress as a separate document with the intention of continuing to develop it as new content-model specific requirements are identified. To do this it needed to be given its own local namespace. A website was therefore set up at http://www.nla.gov.au/australianmetsprofile/ and the controlled vocabulary was located at http://www.nla.gov.au/australianmetsprofile/divtype/.

The content-model specific sub-profiles will also be registered with the Library of Congress as they are developed, while the implementation profiles will be registered locally, probably as part of a future Australian National Data Network (ANDS) Registry service.

At the time of writing the first content specific sub-profile is being finalised and will be submitted to LC shortly. This sub-profile will be called the Australian Journals METS Profile. Initially we had a concern (given that content models are built into the TYPE attribute vocabulary) that the sub-profiles would not add much to the generic profile. However, the first sub-profile has proved particularly useful in clarifying how to apply the requirement for separate descriptive metadata to journal content. It also gave us the opportunity to explore the relationship between the repository in which published content is archived (e.g., DSpace) and the system used to publish the content and provide current access (e.g., PKP's Open Journal Systems software).

What changes did we make to the draft profile

- We migrated to the METS Profile Schema.
- We distinguished between SIPs and DIPs.
- We clarified requirements.
- We specified controlled vocabularies for all mandatory elements.

Migrating to the METS Profile Schema

Converting the PRESTA table-driven draft profile into the required XML schema proved to be quite labour-intensive. It facilitated the use of a narrative approach and made us quantify requirements for extension schemas and controlled vocabularies. Some of the content in the PRESTA document was removed, being more appropriate to user guidelines or FAQs. The need for this level of support in implementing the profile was, however, noted and figures in our draft set of requirements for an Australian METS Profile Maintenance Agency (Appendix F).

Our decision to include two examples of a METS document describing the same object for submission and dissemination purposes confounded the profile schema, which expected unique identifiers for each example and for all internal references. This feedback has been provided to the Library of Congress.

Distinguishing between SIPs and DIPs

The draft profile had been designed specifically to deal with the transfer of an object from one repository to another. It proved to be exemplary in the way it incorporated PREMIS as an extension schema to address the full range of preservation management requirements. However, it did not provide guidance on how to generate METS documents needed for other usage scenarios such as support for submission or dissemination workflows.

Illinois University at Urbana-Champaign has developed a generic METS Profile (http://www.loc.gov/standards/mets/profiles/00000015.html) that, like the PRESTA work, pays particular attention to administrative and technical metadata, but also

differentiates between METS documents intended as submission information packages (SIPs) and those intended as Dissemination Information Packages (DIPs). We adopted this approach in the new version of the profile.

Distinguishing between submission and dissemination workflows also clarified the nature of the primary object being described by a METS document in a given content model. A focus just on submission scenarios for newspapers, for example, might lead implementers to think that the issue is the primary object needing to be represented in a METS document and to specify this in a profile, but in dissemination scenarios the primary object might equally be a page, an article or an illustration.

Similarly, it might lead implementers to think that the primary object and its components are all digital, whereas in a dissemination workflow, the object or some of its components may not yet have been digitised or ingested. Yet descriptive metadata and information about host-component relationships may have already been described to facilitate discovery, navigation and requesting.

Distinguishing between SIPs and DIPs also clarified that different metadata will be available for inclusion in a METS document depending on the usage scenario and the purpose of the repository. Therefore, mandatory requirements have been kept to a minimum in the profile and there still are areas where a repository requiring more information will need to specify this in implementation profiles to achieve interoperability at this level.

Figure 3: Australian METS Profile SIP DIP Interaction Model

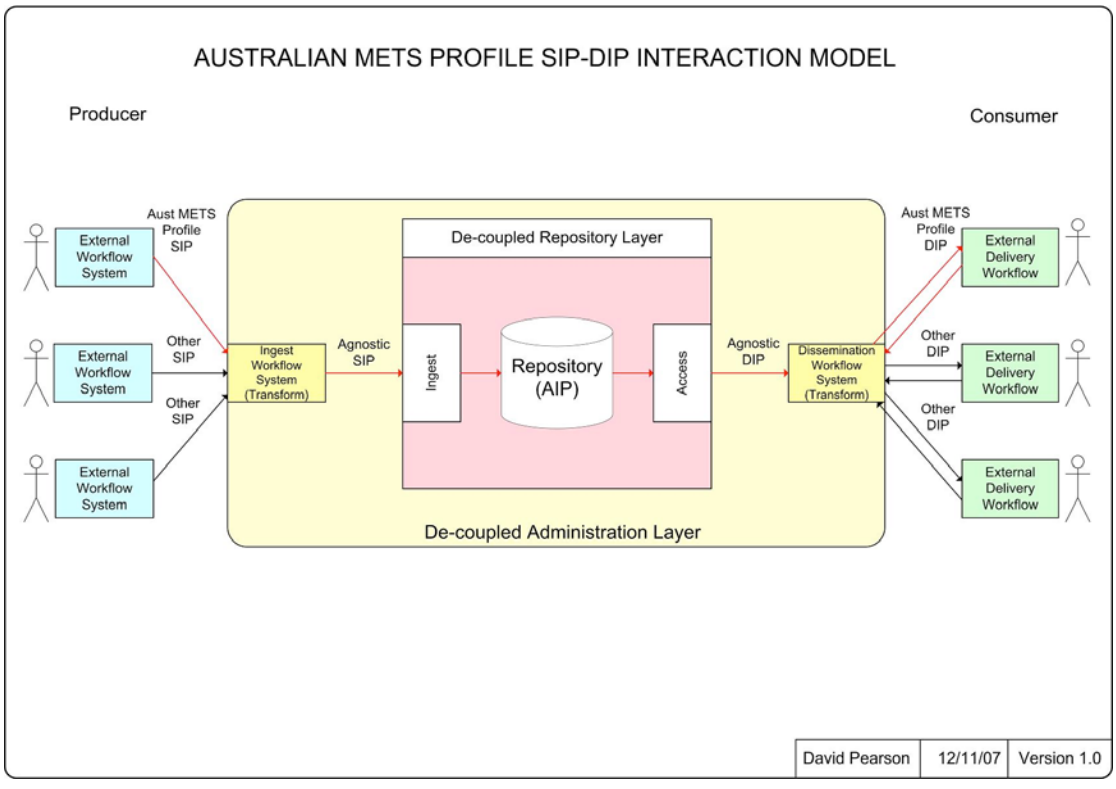


Figure 3 shows the role of the Australian METS Profile in both submission and dissemination workflows. The Australian METS Profile specifies the requirements for a producer to generate a METS document for submission to an Australian repository or for the repository to generate a METS document for delivery to a consumer. A repository might need to support more than one profile for these purposes and choose to convert incoming METS documents to an agnostic submission information package that is recognised by its ingest service. Similarly, its delivery service might

generate a METS document in the form of an agnostic dissemination information package that needs to be converted to the output expected by the consumer.

Clarifying requirements

Developing the profile in conjunction with the implementation of specific workflows proved to be a very effective and agile technique. It exposed requirements that were not clearly expressed or that would have been hard to implement. It also highlighted the need to support the initial processing of a METS document without having to parse content encoded in extension schemas. As a result, the profile requires elements such as file format, file size and checksum to be encoded as METS <file> attributes rather than as PREMIS <object> elements. It also requires each component of an object described in a METS document to have its own self-describing descriptive metadata rather than using the capability in MODS to embed child metadata.

The decision was also made to mandate each discrete component of the packaged object to have a <dmdSec> containing MODS descriptive metadata. While using the MODS part and relatedItem element to describe the package in its entirety within a single <dmdSec> is possible, relationships between components are more effectively reflected using the <structMap>. This allows each component of a packaged object to be fully described as a separate entity within the package.

The major advantage besides an arguably cleaner design, is that processing can locate and process all relevant information about each component solely via the packaging structure. If a single instance of a <dmdSec> were used to describe all components of a package, one would have to use ID/IDREF pairings. This would cross the boundary of packaging and payload markup. This would seem to reflect poor design and would make abstraction and separation of concerns on the processing front more difficult.

Specifying controlled vocabularies

The need to specify values for all mandatory data elements in the profile that required controlled vocabularies proved to be one of the major tasks undertaken by the project team. It would have been easy to leave these for implementation profiles, but then interoperability would have been compromised. Within the three-level model, therefore, controlled vocabularies for all mandatory data elements are prescribed in the generic profile.

Within METS the two most significant controlled vocabularies developed for the generic profile were the <fileGrp> USE attribute and the <structMap> TYPE attribute.

Some of the other profiles had specified <fileGrp> USE attribute values specific to a particular content model. We tried to identify a set that could be applied across content models. This was a particularly useful exercise as made us ensure that archival copies were clearly differentiated from derivatives that could be recreated from an archival copy. All derivatives are bundled under a single value to keep the payload light, with specific derivative behaviours detailed in an implementation profile. Based as it is on experience at the National Library of Australia across a range of content models, this set of values may well be able to form the basis of a standard controlled vocabulary for the <fileGrp> USE attribute.

While a number of members of the international METS community had expressed an interest in developing a standard controlled vocabulary for the <structMap> TYPE attribute, this work had been slow to progress because of the size of the task. During 2004 and 2005 a working group looked at this problem for book-like materials and separated the need for a controlled vocabulary to dictate specific delivery system

behaviours from a controlled vocabulary that might be used to describe the relationship between components for bibliographic citation and display. We made the <structMap> TYPE attribute mandatory to support delivery system behaviours and left the <structMap> LABEL and ORDERLABEL elements optional in the generic profile. We also abandoned the task of developing a comprehensive controlled vocabulary and focused on the values needed to support our particular content models and scenarios.

Within PREMIS, there was a range of mandatory elements requiring controlled vocabularies. Where possible we based these on PREMIS starter lists such as the one for Event Type. The recent addition of “creation” to this starter list proved to be extremely useful for digitisation scenarios. In order to distinguish between preservation intention and preservation capability, we separated Preservation Level into two controlled vocabularies, one to be applied to representations and one to files. The profile requires at least one PREMIS object with Object Category set to Representation and it is here that the preservation intention is recorded on a scale of 1 to x, with the behaviours associated with each level to be specified by repositories outside the profile.

Our treatment of Preservation Level clarified for us the difference between Representation and File in PREMIS. This thinking was fed back into the PREMIS Implementation Group.

What lessons did we learn?

The right team environment

Working in an agile way in a team with theoreticians and practitioners gives the best results. The practitioners tell the theoreticians what’s possible and what’s not. The theoreticians can generalise from implementation examples to identify what is common and can be specified at a generic level and what needs to be detailed at an implementation level.

Being able to generate examples based on real data is essential and also being able to make changes to the code generating the examples in an agile way to test requirements and their implementation. Handcrafting examples is one way of testing requirements but wastes resources and fails to involve the implementers. In order to do this for National Library examples, we should have included an implementer from the Collection Infrastructure Branch on the project team and worked hand-in-hand with them.

Project approach

In the first stage of the project we asked the project participants to provide usage scenarios and sample outputs as a way of gathering requirements. It would have been better to convert the PRESTA draft profile into the METS Profile schema much earlier on, as this identified a range of issues that really provided the agenda for the second phase of the project. Once we started to look at this in detail we were able to generate a template for gathering information from the RIFF partners that focused on the things they needed to tell us to start testing content models and developing specific implementation profiles.

Even so, a number of the participants were unable to provide the information required by the template as they were simply not ready to do so.

Knowledge and environmental requirements

METS is really complicated, because of what a METS document is trying to describe and all the extension schemas needed to do this. But in the end its purpose is just to

be used as a submission information package or a dissemination information package. A receiving repository needs to set standards for a SIP, be able to validate that it meets those standards and unpack it for further processing and ingest. Similarly, a consumer needs to know what it can expect to receive from a repository in the form of a DIP and to be able to request the kind of DIP needed for the particular use to be made of it and to unpack it for further processing and use.

The organisational infrastructures and agreements needing to be in place to enable this to happen require an in-depth understanding of usage scenarios, service usage models, future practice, etc. METS has to be seen at an organisational level and across organisations as a primary interface object in any submission or delivery request and response.

There is a risk until this vision is shared that some stakeholders will react adversely to the inherent complexity of METS without realising how flexible it is. The same profile can be used to define requirements for the use of METS as both a SIP and DIP. There is no reason why it cannot be used for objects not yet digitised or ingested or for recording the relationships between objects in more than one repository. METS may seem to have been around for a long time but it is still a relatively new standard and still in the process of being deployed in ways that will establish best practice.

Reliance on external dependencies

Having external dependencies delayed the project and limited the outputs. On the other hand, it provided us with potentially a richer body of examples because we had to draw on other resources. It also made us adopt a scalable, modular approach in the 3 level model to achieve the outcomes of the project in spite of the RIFF partners not being ready to fully contribute. In practice, the task could never have been completed by depending entirely on the partners to build their own content models. We are now positioned to add new content models if these are needed as these projects come on line. Handover and continuity will, however, be essential for this to happen.

Standards and the community

Developing standards is hard. There is a steep learning curve. Mentorship is crucial. A range of skills is required, particularly in this case, where the profile needed to be expressed in an XML schema and fully-validated examples provided that illustrated the use of extension schemas and controlled vocabularies. To do this required xml skills and in-depth knowledge of the METS Profile Schema itself as well as METS, MODS, PREMIS, MIX, etc. There are tools that facilitate the validation of a complex package like a METS document. These would have made validating our examples easier.

Limitations of the LC profiling process

Profiles are a way for a community to agree on how a particular standard or set of standards will be implemented. They define levels of conformance, controlled vocabularies, extension schemas, behaviours, etc. to which all implementations within that community will strive to conform. These kinds of profiles tend to have maintenance agencies that continue to explore community needs and fine-tune the profiles against them.

The extent to which an implementation conforms to a profile then becomes something needing to be recorded in service or protocol registries such as defined in ISO 2146 (*Registry Services for Libraries and Related Organisations*). ISO 2146 assumes that, as part of registering and describing an online service, the protocols

and profiles supported are recorded and also any exception behaviours. This is the approach we have adopted for the Australian METS Profile.

The current LC profiling process encourages the registration of individual implementation profiles rather than making a clear separation between defining best practice and documenting levels of conformance.

Profiling work is important to the achievement of consensus and common practice within a community. This really has to happen before one can start talking about machine-actionable METS documents and conformance testing, although the two go hand-in-hand; i.e. there is a clear benefit in doing profiling work by testing usage scenarios against actual implementations.

Once we realised this we were able to move relatively quickly to a 3-level model and this then enabled us to be agile in articulating the requirements at each level.

Corporate memory

As a project progresses it is easy to forget the issues raised, problems solved and lessons learnt from previous iterations. It would have been really useful when we embarked on the project for the whole team to read and discuss the full APSR PRESTA Project report. We didn't do this and spent some time re-inventing the wheel. Luckily, we did take time out at the end of the project to review the report and identify any outstanding issues.

As an example, the PRESTA report had tested the audioMD and videoMD extension schemas against the National Library of Australia's implementation experience with audio and video content models and made some recommendations for additional data elements to support new digital formats. We only "remembered" this at a late stage in the project.

The Australian METS Profile 1.0 still just references audioMD and videoMD as extension schemas because it would have been too large a task to develop local extension schemas to cover the new requirements. However, we did take a policy decision to try and include the new data elements as revisions to the two schemas rather than as local extensions. This will be another task for the proposed maintenance agency as it develops the content-model specific profiles for audio and video.

What needs to happen next?

Handover to a maintenance agency

Handover of the work done, the lessons learnt and the things still needing to be done to implement the 3-level model will be essential:

- It is very easy to lose the facility that one has gained by working daily with standards as complex as METS, PREMIS, MODS and MIX, let alone other extension schemas.
- We need to make it easy for repository managers to register implementation profiles using a Wizard / forms approach so that they can enter the data without having to do a crash course in all of the above.
- We need to continue the process of testing requirements, registering new content-model specific profiles and extending the TYPE controlled vocabulary to support new content models.

- We particularly need to make sure that the National Library of Australia tests the profile against its own content models as part of moving to a service-oriented architecture for its digital library services.
- We need to continue development of the website and prepare FAQs, guidelines and training materials (taking a 'train the trainer' approach) to support implementation of the profile.
- We have achieved some credibility internationally in the METS community with the work done so far and we need to continue to play an active role in this space.

Draft requirements for a maintenance agency to advance these activities are included in Appendix F.

ANDS

The proposed new Australian National Data Service (ANDS) is a service that will facilitate the curation, discovery and use of data created by Australian research. A full description of the scope of this service may be found in the document *ANDS - Towards an Australian Data Commons* produced by the ANDS Technical Working Group to inform development of the service.

We need to make sure that ANDS embraces the Australian METS Profile and its ongoing development and promotion as a core part of the Utilities Program and that the outreach aspects of implementing the Australian METS Profile are fully incorporated into the ANDS Repositories Program.

ORCA –implementation registration

The Online Research Collections Australia (ORCA) Collection Service Registry will be a core service of the ANDS Utilities Program.

We need to make sure that registration of implementation profiles is done in the context of the ORCA Network and Collection Service Registry. We think that users should be able to find how a given repository or collection requires SIPs to be formatted for submission and also what kinds of DIPs it delivers through the registry service. Without a standard such as METS the interoperability promised by ORCA will be limited.

National Persistent Identifier Service

A National Persistent Identifier Service will be a core service of the ANDS Utilities Program. We need to make sure that the need to use globally persistent identifiers for objects described by Dissemination Information Packages (DIPs) conforming to Australian METS Profile is supported by and promotes the development of a National Persistent Identifier service.

OpenURL METS profile

Looking at dissemination workflows as part of this project has exposed the need for a protocol enabling a consumer to request a METS document for an object and also to specify certain kinds of behaviours for that METS document.

A consumer may know the persistent identifier of an object and want to resolve to a METS representation of that object rather than to the default resolution service. In addition the consumer might also just want, say, thumbnail images or the structural map rather than a full set of files.

OpenURL seems an appropriate protocol for this purpose. It is already being used to solve the "appropriate copy" problem in other spaces. The controlled vocabularies developed for the METS <fileGrp> USE attribute and the METS <structMap>

TYPE attribute (see below) could also be deployed effectively within OpenURL to request a METS document describing, for instance, only files of a given use or only components of a given type.

We need to make sure that this requirement is pursued as a national objective and that the Australian METS Profile Agency is actively involved at an international level with the development of an OpenURL METS Profile should this be perceived as the most appropriate solution.

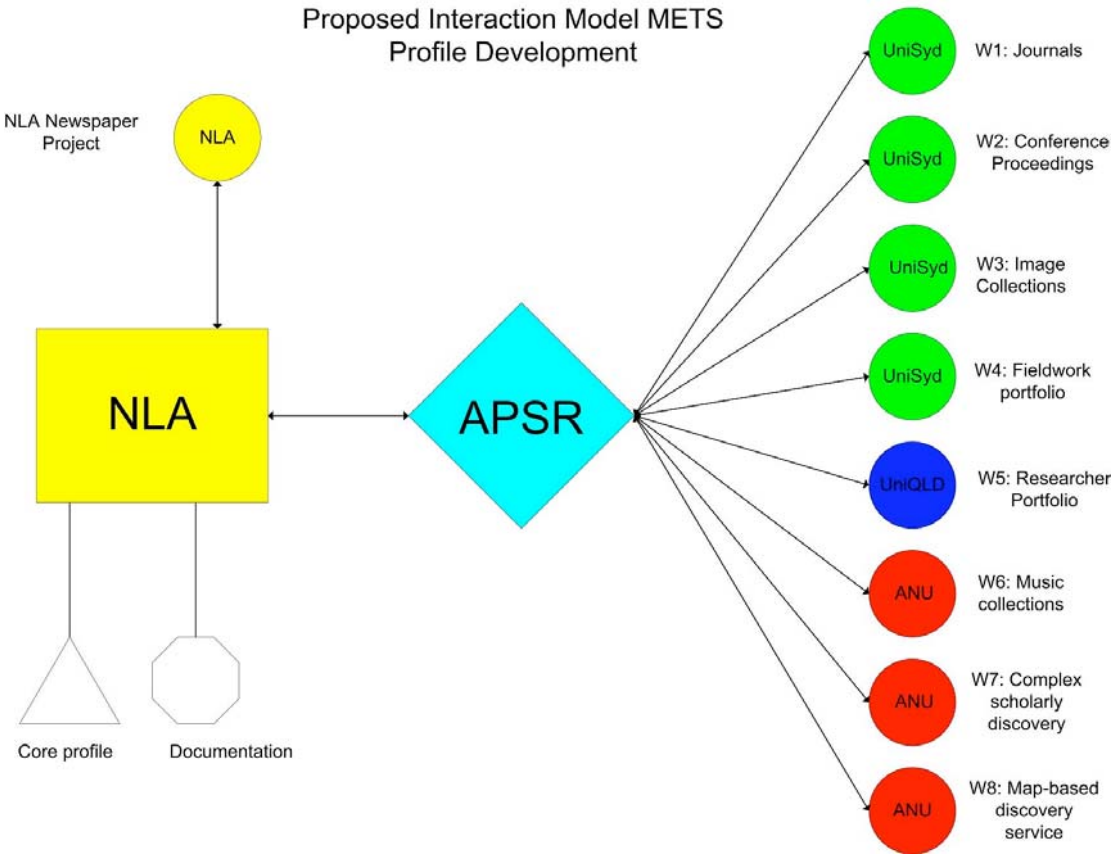
Conclusion

The APSR METS Profile Development Project has achieved a very successful outcome which exceeds the original scope of the project. While external dependencies delayed the project and we did not end up registering METS profiles for each of the stakeholder implementations, this turned out to be a good thing. Instead of adding to a proliferation of implementation and scenario-specific profiles, we were able to develop a 3-level model for setting best practice and registering conformance and exception behaviours. This model is scalable and extensible. It will allow for the gradual registration of content model specific profiles. Assuming that this work is continued, it will contribute significantly to an interoperable Australian data commons, as envisaged in the ANDS scoping paper. The major risk to achieving this is that the impetus is lost and repository managers have no standards or guidance towards achieving interoperability.

Acknowledgements

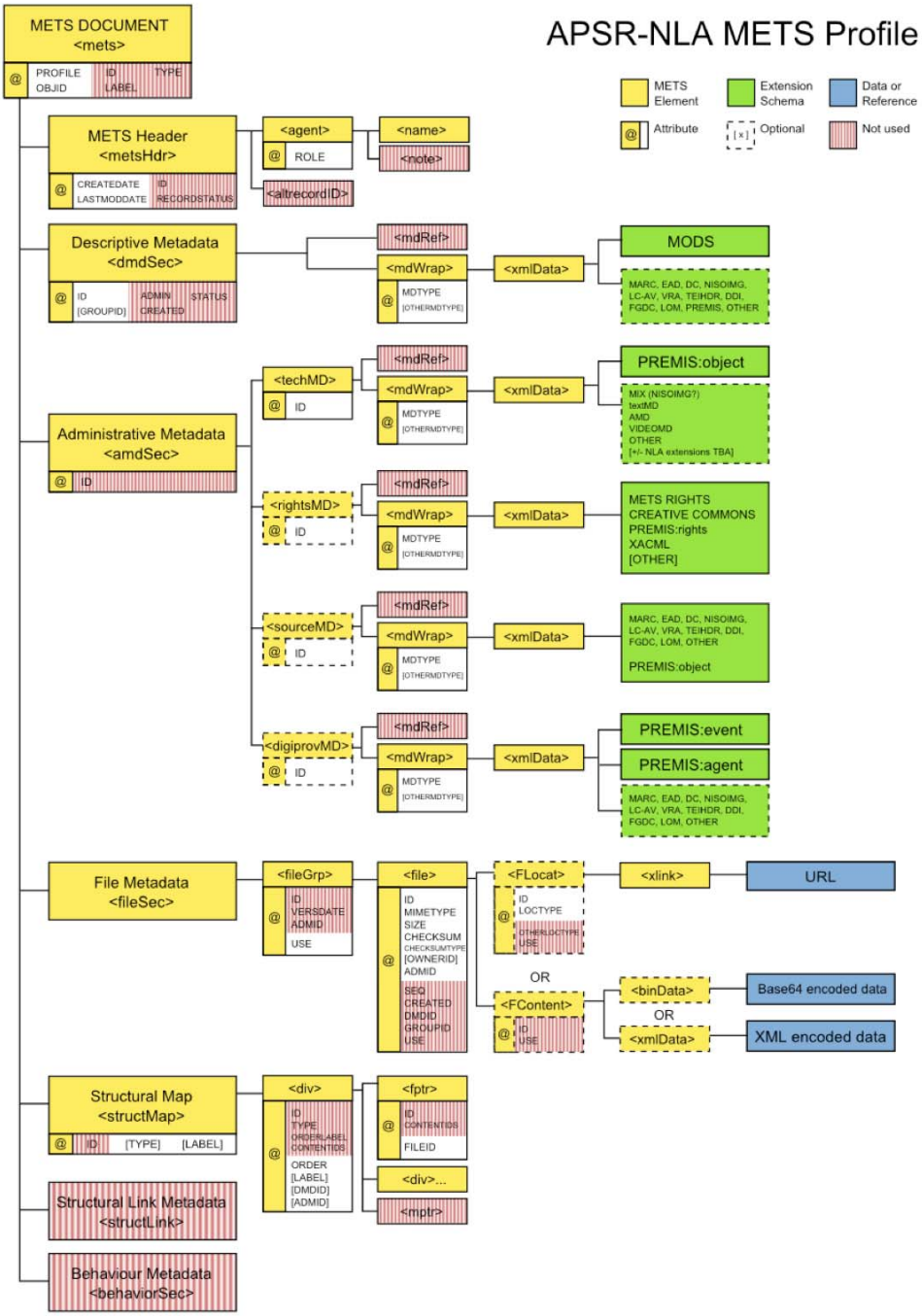
The Australian METS Profile was developed by a team comprising Judith Pearce, David Pearson, Megan Williams and Scott Yeadon. Colin Webb was the project sponsor, David Pearson managed the project, Chris Blackall managed the APSR RIFF projects for the first half of the year and Scott Yeadon for the second. Megan Williams was the profile editor and Judith Pearce was the technical lead. The APSR RIFF project partners provided input on content models. Thanks also to the ARROW project for their input and to Gerard Clifton, Douglas Elford, Nicholas Del Pozo, Bronwyn Lee for their ongoing support and feedback. Particular thanks to Gerard, Bronwyn and Somaya Langley for the PRESTA report which formed the basis for our work.

Appendix A



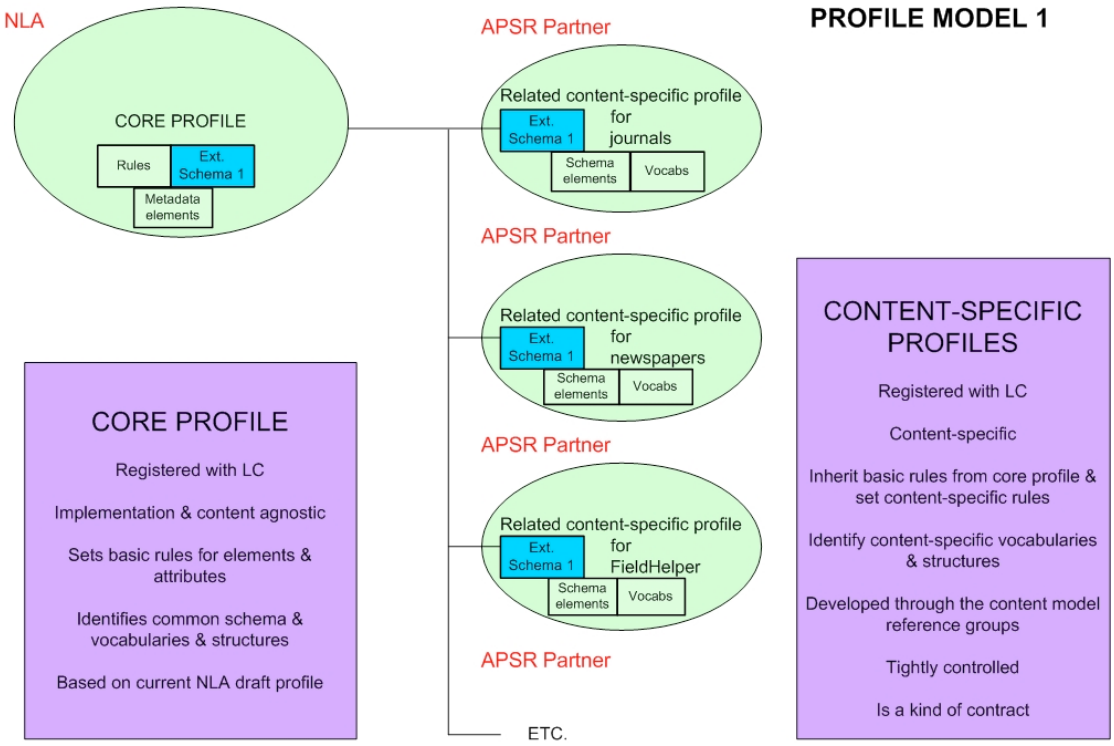
Drawn by David Pearson

Appendix B



Drawn by Gerard Clifton

Appendix C



Drawn by David Pearson

Model 1 is essentially the original model developed from the earlier PRESTA work, that is, a core submission and exchange profile based on the current NLA draft exchange profile, but with each of the APSR content-models expressed as separate, content-specific profiles.

In this model, the core profile would:

- be a core generic profile;
- not be governed by a particular implementation or type of content;
- express a core set of rules and requirements for submission and exchange packages – e.g. requirement that there should be at least one <dmdSec> for the entire intellectual entity represented by the METS document, and that the metadata in this <dmdSec> must conform to the MODS schema; and
- identify common schema and vocabularies – e.g. MIX schema for technical metadata about image content files, PREMIS event type terms for <digiprovMD>.

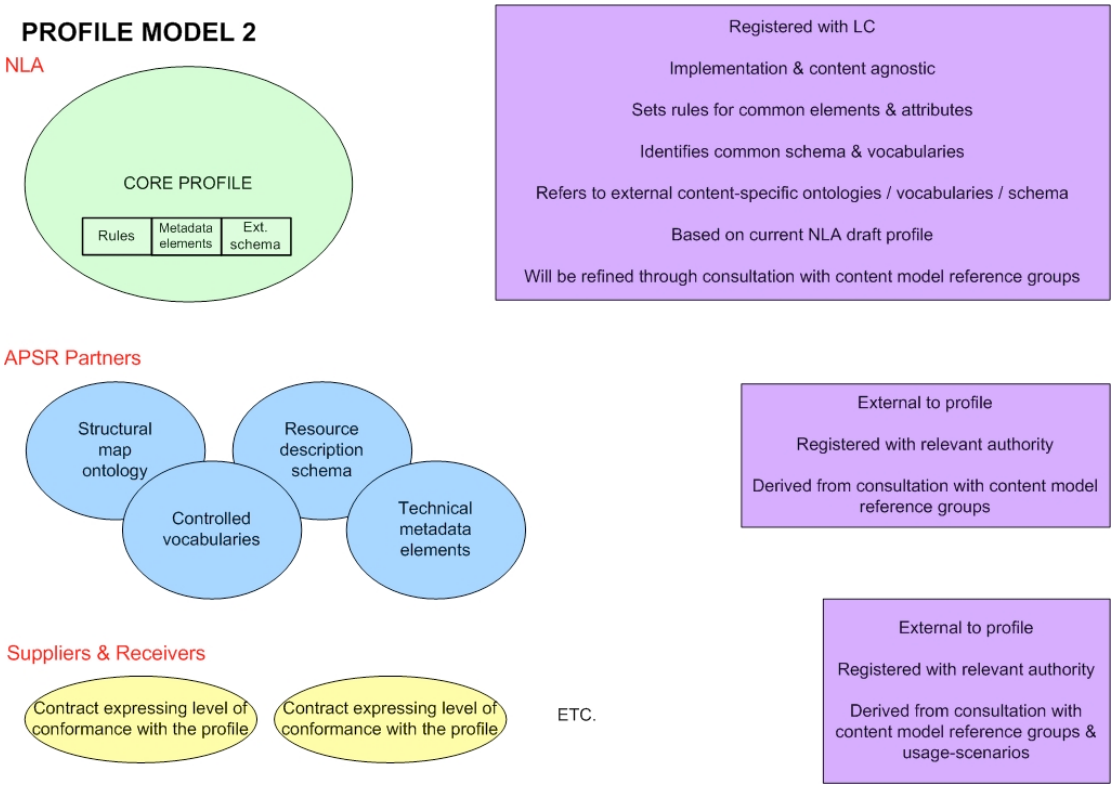
The content-specific profiles would:

- be developed from the content model reference groups – e.g. a profile for journals (OJS workflow), a profile for conferences (OCS workflow), a profile for FieldHelper (Fieldwork workflow);
- inherit and adhere to the rules of the core profile; and
- further express any content-specific requirements, such as attribute vocabularies and file structures – e.g. content-specific vocabularies for the <fileGrp> USE attribute

These multiple profiles would all be registered with the Library of Congress.

Positives	*Negatives*
Achieve ASPR objectives within current resources and timeframes	Content-specific profiles may be narrow in focus
Meet business needs of project partners	Would be greater number of profiles to update / maintain
Cater for diverse nature of project partners	
Profiles would be tightly-controlled	
But profiles would also be adaptable according to feedback from partners - e.g. could allow multiple structMaps if commonly required	

Appendix D



Drawn by David Pearson

Model 2 was developed from various discussions with stakeholders within the NLA, and aims to facilitate a more standards-based approach to developing the profile in a way that is perhaps more generic and more widely applicable. There was a realisation that there is some overlap in the content models, and a concern that the content-specific profiles might be too instance-specific, too narrow in focus. For example, would a new profile need to be developed to suit every new usage scenario, or any variations to the content that is being modelled?

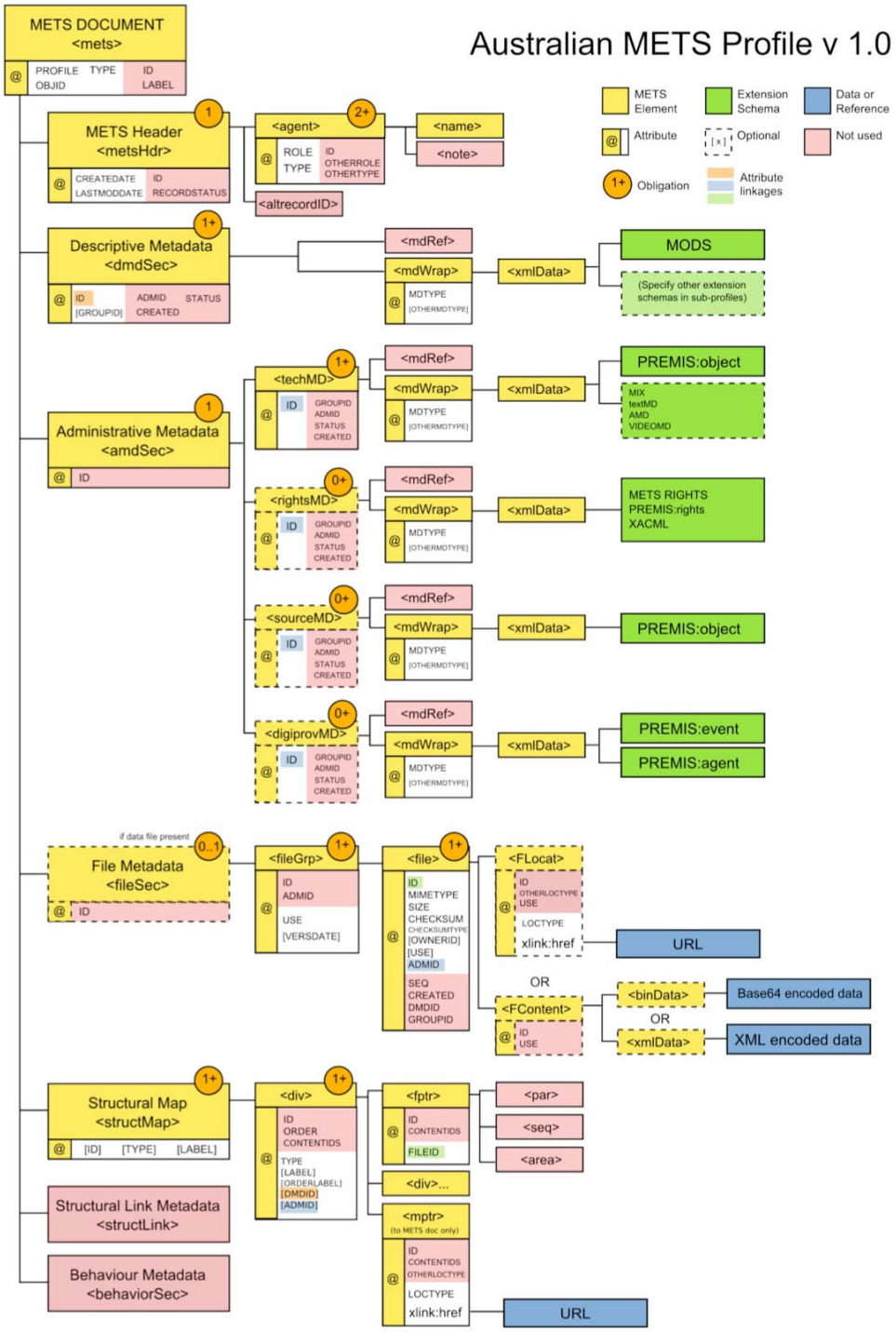
This second profile model features:

- one generic profile, which expresses a core set of rules and requirements for submission and exchange packages - e.g. requirement - there should be at least one <dmdSec> for the entire intellectual entity represented by the METS document, and the metadata in this <dmdSec> must conform to the MODS schema; and
- refers to externally registered and maintained sources for any content-specific information – e.g. <structMap> ontologies, attribute vocabularies, resource description schema, etc.; while
- a variety of usage scenarios could be catered for through externally registered contracts between suppliers and receivers, expressing levels of conformance with the profile – e.g. NLA contract with supplier of OCR text for Newspapers Digitisation Project.

In this model, the core profile would be registered with the Library of Congress. Any schemas or vocabularies or contracts referenced by the profile would be registered with the relevant authority.

Positives	*Negatives*
Standards-based	May be too generic and flexible
Promotes interoperability	May be more difficult to achieve within the timeframe and resources of the project
Flexible	Perhaps too ambitious for this year's APSR work
Single profile to update/maintain	

Appendix E



Drawn by Gerard Clifton

Appendix F

Proposed role and responsibility of Australian METS Maintenance Agency

Ongoing – (frequently). The maintenance agency will:

1. As part of the NLA SOA, work with the Collection Infrastructure Branch to specify requirements for using METS as a SIP and DIP in DCM and PANDAS usage scenarios;
2. Develop new content specific sub-profiles and/or refine existing profiles to incorporate the NLA requirements;
3. Provide advice and training to the ANDS (Australian National Data Service) Repository Program about how to use the three-tiered Australian METS Profile model to facilitate interoperability and the development of implementation profiles.
4. Assist Australian Repository Managers in other sectors (Government, Cultural) to create implementation profiles and develop new content specific sub-profiles and/or refine existing profiles to incorporate new requirements;
5. Register new or updated content model specific sub-profiles with the Library of Congress.
6. Ensure that new implementation profiles are registered on the NLA METS Website.
7. Add new content models to the Structural Map DIV Type controlled vocabulary as use cases occur. This will be maintained at http://www.nla.gov.au/australianmetsprofile/divtype/.
8. Participate in the ongoing development of the PREMIS data dictionary and best practice for preservation metadata.
9. Maintain the Australian METS Maintenance Agency Website http://www.nla.gov.au/australianmetsprofile/.
10. As part of establishing a best practice standard for SIP & DIP in Australian Repositories, produce documentation in the form of Web content, templates, how-to and frequently asked questions. Refine and update older documentation based on implementation experience.
11. Establish a rapport with Library of Congress METS office and other best practice practitioners.
12. Run training seminars as required.

Ongoing – (Less frequently). The maintenance agency will:

1. Issue new versions of the registered Australian METS Profile based on implementation experience to Library of Congress.

David Pearson & Judith Pearce
25 October 2007