

MoPark Initiative

Metadata Options Appraisal

(Phase I)

**Report by Dennis Nicholson
Centre for Digital Library Research
Computer and Information Sciences Department
Strathclyde University
8th June 2004**

MoPark Initiative: Metadata Options Appraisal (Phase I)

Executive Summary

This *Metadata Options Appraisal* examines – and makes recommendations on - the needs of the Loch Lomond and Trossachs National Park as regards the metadata, metadata standards, and metadata management required for the competent handling of digital materials both now and in the future. Based on discussions with MoPark project and National Park staff, and on a detailed examination of project and Park documentation, three levels of requirement have been identified:

- MoPark project requirements (Level 1)
- Requirements if MoPark is extended to other topics and Park areas (Level 2)
- General requirements of the National Park beyond MoPark (Level 3)

Since early analysis suggests that it will not be feasible to determine all of the detail of even the Level 1 requirement without creating and field-testing at least one of the proposed 'interpretive journeys', and that fully determining the requirements in respect of the other two levels (and Level 3 in particular) will face similar (but longer term) difficulties, a phased approach to agreeing the full and detailed requirement is recommended. Phase 1, dealt with through the completion of this report :

- Sets out a framework within which the full requirement can safely develop
- Proposes a flexible forward development path that will progressively facilitate the specification of detailed metadata-related needs for Levels 1, 2, and 3 and ultimately lead to the determination of the full requirement

The second of these elements constitutes Phase II of the proposed approach and entails joint working between MoPark staff and consultants, Park staff, and CDLR to determine the detailed requirement progressively over the full lifetime of the project. The phased approach is necessary because MoPark and Park staff require further experience of the complex implications of managing digital objects for the purposes envisaged in project and Park documentation before they can provide some of the detail necessary to determine the full requirement.

The framework proposed has three elements:

- Adoption of the *Metadata Encoding and Transmission Standard* (METS) developed by the *Digital Library Federation*¹ to provide an XML document format for encoding metadata necessary for both management of digital library objects and their exchange between repositories or between repositories and their users. METS provides for all of

¹ The METS schema is a standard for encoding descriptive, administrative, and structural metadata regarding objects within a digital library, expressed using the XML schema language of the World Wide Web Consortium. The standard is maintained in the Network Development and MARC Standards Office of the Library of Congress, and is being developed as an initiative of the Digital Library Federation. The official METS website is at <http://www.loc.gov/standards/mets/>. The JISC web-site also has information on the standard.

the metadata types likely to be required within MoPark and the Park – *descriptive metadata* (MARC, Dublin Core etc) at both individual object and aggregate (i.e. Interpretive Journey) level, *administrative metadata* (technical metadata, rights metadata, analogue source information, digital object files provenance), *Files metadata* (for files containing content which comprise the electronic versions of the digital object), *Structural Map metadata* to outline the hierarchical structure of a digital library object such as an Interpretive Journey, *Structural Links metadata* to allow links between hierarchical levels to be described, and *Behaviour metadata* to allow metadata on 'executable behaviours' to be encoded (for example, the need to run a piece of software in order to present a particular part of an Interpretive Journey to a visitor).

- The adoption, where possible and appropriate, of national and international standards relevant to the field.
- Cooperation with other key players with similar needs and interests to harmonise approaches and ensure interoperability beyond Park systems.

Phase II of the Metadata Options Appraisal process – the proposed flexible forward development path for the specification of the full requirement - is set out in rough outline in Section 4 of this report. It allows for ongoing discussion and examination of the requirements in terms of metadata, metadata standards, and metadata management. Questions regarding full implementation of the requirements and associated issues such as the staffing and training requirements and costs of implementation are, of necessity, dealt with in Phase II of the approach.

A full draft of this report was presented to the MoPark Project, and discussed with project leaders and with Datavisibility, the DAMS consultants. A presentation on the report was given to key National Park staff at Balloch Castle on 6th May and the report was subsequently accepted by the National Park Senior Management Team on 10th May 2004. This, slightly amended version of the report was submitted on 9th June 2004 to MoPark and represents the agreed way forward as regards the MoPark Metadata Options Appraisal – that is, the next steps will be as specified in Section 4 of this report – the Phase II Outline.

MoPark Initiative: Metadata Options Appraisal

Contents

Section	Title	Page
	Executive Summary	2
	Contents	4
1	Essential Contextual Information and Background	5
2	Metadata Needs (I) – Initial Analysis	7
3	Metadata Needs (II) – The Need for a Phased Approach	12
4	The Way Forward – Phase II Outline	14
5	Conclusions and References	16
Appendices		
A	METS: An Overview & Tutorial	18
B	METS Extension Schemas	27
C	Phase II Metadata Detail Questionnaire	31

Section 1. Essential Contextual Information and Background

MoPark Overview

The MoPark project aims to encourage green tourism within the Loch Lomond and the Trossachs National Park. It focuses on the east side of Loch Lomond, including Balmaha; Inchcailloch; East Loch Lomond; Loch Katrine and the National Park Gateway Centre in Balloch.. By providing opportunities for sustainable travel and by interpreting its use, it will fulfil all four of the National Park's objectives. It will *promote the sustainable use of the National Park's resources*. It will use innovative interpretation to *help people understand and enjoy the area's special qualities*, which will in turn encourage positive behaviour that helps *to conserve and enhance its natural and cultural heritage*. Ultimately, it will *help the National Park's economic and social development* by encouraging visitors to come to the Park, to stay longer and by involving local communities in the development of the interpretation.

Digital Asset Management System (DAMS)

The creation and population of a digital multimedia repository and management system will underpin the initiative. The information stored in this system could include a variety of media including documents, photographs, film or audio archive, web-site links. This system needs to interface directly with spatial data (GIS), and is closely related to a Map-based orientation system.

MoPark will pilot a limited number of interpretive journeys (focused within a relatively small geographical area of the National Park). However the digital asset management system must be capable of supporting substantial growth to enable the development of numerous additional interpretive 'journeys' in the future. The system must facilitate interpretive journeys to be developed along the same themes and messages, but with varying levels of complexity or knowledge e.g. to allow the same interpretive journey to be available for different ages of school groups by tagging data (Note this is essentially a design problem that the system needs to be able to support).

Metadata Options Appraisal

Metadata describing the digital materials used in MoPark will play a key role in both the delivery of digital services to visitors and the management of the digital materials by staff and the *Metadata Options Appraisal* is therefore a key requirement. During the first half of the appraisal process, it became clear that, whilst the scope and complexity of the general requirement in this area was already evident, much of the information required to determine detailed requirements could only emerge over time as the MoPark project designed and piloted actual 'interpretive journeys' and Park staff gained insight into the requirements of managing digital objects. It was therefore agreed with the project leaders that a phased approach to the Metadata Options Appraisal was sensible – with Phase I, completed with the production of this report, setting out a general framework within which the full requirement could be met, and Phase II stretching out over the full lifetime of the project and adding detail as specifics of the requirement emerged.

In broad outline, Phase I of the appraisal process consisted of:

1. Participation in an initial briefing session with the client group and suppliers.
2. Initial analysis of the outcomes, followed by an in-depth consultation with project leaders on information needs, the amount and level of cataloguing required, the types of digital object likely to be stored and presented, the likely modes of presentation of these within interpretative journeys, and any requirements to access distributed information sources and integrate with other repositories, either now or in future.
3. In-depth examination of a range of project and Park documents on issues relating to, or having an impact on, metadata needs.
4. Creation of an outline, but detailed, sketch of the scope and complexity of the likely general needs of the project and the National Park in respect of metadata for digital objects.
5. An examination of prominent global digital library initiatives and the metadata issues and solutions adopted.
6. Identification of the *Metadata Encoding and Transmission Standard (METS)* maintained by the Library of Congress as providing a digital object metadata framework sufficiently complex and flexible to meet the needs of the MoPark project and the National Park.
7. Identification of a range of areas and issues requiring more detailed answers before the detailed metadata requirements of the project and the Park can be specified.
8. Creation of a set of detailed questions that need to be answered in Phase II of the Appraisal in order to draw out and agree the detailed requirement.
9. Creation of an outline plan to enable the detailed metadata requirements to be specified and implemented within the METS framework.
10. Production of a full draft of this report on Phase I of the Metadata Options Appraisal.
11. Discussion of the draft report, its implications, and of follow-up actions, with MoPark staff at a meeting in Glasgow, and by telephone and email.
12. Similar discussions of metadata and other issues at a meeting held at Balloch.
13. Discussions with Datavisibility on their proposed approach to the next phase of the project, on the implications of Phase II of the Metadata Options Proposal for their proposed approach, and on contractual matters relating to Phase II.
14. The creation of a presentation on the full draft of the Metadata Options Proposal and its delivery at a meeting of National Park Staff at Balloch Castle. Subsequent discussion of the report and its implications with the staff.
15. Alterations to the draft report to produce this final report on Phase I of the Metadata Options Proposal.

Section 2. Metadata Needs (I) – Initial Analysis

Introduction

The Metadata requirements of any system are primarily related to function. In order to determine the requirement for MoPark and for the National Park, it is necessary first to determine what the requirements are in respect of handling digital information and information products. In rough terms, to determine *which* things need to be done to or with *what* information and information products, by or for *whom*, in what circumstances (*when, how, where*), and for what purpose (*why*) in order to meet the management and service requirements of the system. Other issues affecting decisions on metadata include resources available for implementation and ongoing management, existing constraints in terms of the existence of legacy metadata repositories and legacy systems used by organisational partners, and the need to 'interoperate' with services outwith the group of organisational partners. A key issue in respect of the metadata needs for MoPark and the National Park is that there are three levels of requirement:

- MoPark project requirements (Level 1)
- Requirements if MoPark extended to other topics and Park areas (Level 2)
- General requirements of the National Park beyond MoPark (Level 3)

General Requirements and Detailed Requirements

By the completion of step 4 of the process described at the end of the last section of this report, two things were clear:

1. That there was sufficient detail available from initial discussions with key MoPark and Park personnel (step 1), from follow-up discussions with project leaders (step 2), and from project and Park documentation (step 3), to enable the general needs of the project and Park in respect of digital object metadata to be specified.
2. That specifying the need in detail would be more difficult. It would require – at minimum - further work with actual examples of the complex digital objects (interpretive journeys) likely to form the primary elements of the Level 1 and 2 requirements, more experience amongst project and Park personnel of the issues and problems associated with managing complex digital objects, and an in-depth survey of the likely range of other digital objects, their usage, and their life-cycles.

The remainder of this section of the report deals with the first of these two elements of the requirement. The approach to identifying the specifics of the detailed requirements is covered in the next section, section 3.

General Requirements

The following general requirements have been identified:

- The 'interpretive journeys' envisaged by MoPark are complex digital objects comprising coherent aggregates of a range of simpler digital objects in a variety of formats. They will require a metadata framework that encompasses (1) descriptive

metadata (e.g. a MARC record) for both the aggregate object and the constituent objects, (2) Administrative metadata for the two levels – including technical detail on format and use characteristics, intellectual property rights information, descriptive and administrative metadata on the analogue source of the digital object, digital provenance metadata, (3) Metadata on the files encompassed by the aggregate object, (4) Metadata on the hierarchical structure of the digital object and on hyperlinks between levels (structural metadata), and (5) Metadata on executable programs required to run some of the constituent digital objects.

- Different types of Park visitor will be presented with different versions of each interpretive journey and these versions will themselves vary according to visitor choice of things like preferred terrain, and mode of transport. It is not clear at this stage whether each journey will be one single aggregate object or many similar but different aggregates. Either way, the metadata framework will have to be sufficiently flexible to cope with the inter-relationships involved and the complex management issues entailed.
- The issues described in the two points above are made even more complex because of the fact that different presentations of the various interpretive journeys are likely to be required for a range of different output modes (PDAs, web-sites, mobile phones, and others)
- Although it is clear even at this stage that there is a need for at least one² professional with information management and digital asset management skills to oversee the metadata and the ensure adherence to standards, the creation and management of these complex objects is likely to involve a range of Park personnel with different skills and expertise and this will have implications in respect of the need for a range of record statuses and the ability to record staff involved in upkeep and creation in the metadata. Issues like information currency, accuracy, authority, access rights, and so on will make such involvement necessary and make workflow control and associated metadata needs essential elements of a working system
- Initial discussions suggest that the constituent digital objects that will make up the detail of the interpretive journeys are likely to be in a wide range of formats, including (but not necessarily limited to), textual materials, audio material, video materials, still images, animated sequences, 3D animations, 3D and 2D maps, interactive packages of various kinds. Any framework will have to be able to deal with the metadata requirements associated with each.
- Special software will be needed to handle some of the constituent digital objects involved and it will have to be possible to specify this in the metadata framework. Included in this requirement is the need to obtain and respond to GPS data and to interact with an external GIS database system and repository.
- It is likely that the required descriptive metadata for the aggregates and the constituent objects will in some cases be held on external systems (e.g. maps metadata and, possibly, MARC or Dublin Core metadata). In other cases, it may need to be embedded in the record for the object itself. The metadata framework will have to be able to handle either or both situations

² Whether one is enough will depend on the numbers of digital objects ultimately involved and on associated management issues. It will also – inevitably – be influenced by the resources available to the Park for such purposes

- The range and nature of digital objects likely to be managed when the needs of the Park as a whole are taken into account is at present unknown in detail. The framework adopted must therefore be flexible and wide-ranging to enable it to cope with future requirements
- Project documents suggest the probability that there will, in time, be a need to preserve some or all of the digital objects in the repository, so the metadata framework adopted needs to be able to deal with this likelihood
- It seems likely that there will be a need to present or otherwise disseminate the metadata in various forms and various formats for different purposes in different situations. Possible needs in this respect include the need to interoperate with Library, Museum and (possibly) Archives community systems in Scotland³, to exchange data with other Parks or project partners in other countries, to meet government requirements such as e-Gif, and so on. It would be wise to allow for the use of a variety of standards, of mapping to different standards, and output in various formats. Either storing the metadata in XML format or permitting its export and import in XML is advisable. Adopting national and international standards where possible is also advisable.

The METS Framework

The kinds of issues specified above have been tackled to a greater or lesser extent by a range of digital repository developers in Scotland, the UK, and the world generally. It would be possible to choose one or other of the approaches adopted by one of these and adapt it to the needs of the project and the Park. The CDLR itself has one such project – Victorian Times – which has a variety of complex requirements that echo at least some of those in the MoPark project and some initial work was done to examine the possibility of extending it to meet the needs of MoPark and the National Park. However, having researched work done at a variety of digital library repositories around the world and examined, in particular work done in this area by the Library of Congress and its partners, and documents on the topic on the web-site of the Joint Information Systems Committee (JISC) of the UK Higher and Further education communities, an approach more in line with both the short and longer-term needs of the Park has been identified. The METS: Metadata Encoding and Transmission Standard is a new standard that has been developed specifically to provide an overall framework within which metadata for all types of digital materials - digitized video, sound files still images, electronic texts, and others – can be integrated. Expressed in the World Wide Web Consortium's XML schema language and maintained by the Network Development and MARC Standards Office of the Library of Congress, it provides a flexible and coherent framework within which to meet MoPark and National Park digital materials metadata requirements.

METS aims to provide an XML document format for encoding metadata necessary for both management of digital library objects within a repository and exchange of such objects between repositories (or between repositories and their users). It caters for all of the types of metadata (descriptive, administrative, structural, executable behaviors etc) used to describe digital library objects. It permits both the embedding of the metadata within the document and

³ Required to allow visitors at Visitor Centre web-sites to explore cultural, historical or scientific themes on other Scottish systems or, indeed, potential visitors to find links from external web-sites to Park interpretive journeys

the referencing of externally held metadata. Hierarchical structures and links between levels can be encoded and pointers to software required to run particular constituent objects can be specified. Since the data is held in XML format, it will be platform and software independent and readily exchangeable with partners and cooperating organisations. The schema is a new one but has already been adopted by a number of digital library projects. It is understood that a European interest group (Gartner, Richard 2002) is being set up and that an international editorial board has been established to coordinate future developments.

The primary elements of a METS document are:

The METS Header. This contains metadata describing the METS document itself, including such information as creator, staff involved in creation and maintenance and their roles, and record status. This would enable MoPark requirements as regards staff roles and metadata and digital object maintenance workflows to be managed.

The descriptive metadata section. This section holds or points to pieces of metadata describing an aggregate digital object and its constituent parts. In the MoPark context, it might contain (say) a section pointing to descriptive metadata for Park area maps held on another server, another pointing to a MARC database containing records describing MoPark digital content at the aggregate or interpretive journey level, and several embedded sections containing actual descriptive metadata for the videos and sound files and images that are the parts of the interpretive journey.

The administrative metadata section. This includes four sub-sections: *technical metadata* (information regarding files' creation, format, and use characteristics), *intellectual property rights metadata* (copyright and license information), *source metadata* (descriptive and administrative metadata regarding the analog source from which a digital library object derives), and *digital provenance metadata* (information regarding source/destination relationships between files, including master/derivative relationships between files and information regarding migrations/transformations employed on files between original digitization of an artifact and its current incarnation as a digital library object). MoPark and the National Park would definitely require the first two of these types, and would probably require the last type (for digital preservation purposes). Further investigation is required to determine whether it would require metadata on the analog source of digitised objects, but it is possible, and the section is, in any case, optional.

Files metadata. This is where details of the various constituent digital objects of an interpretive journey would be listed, referenced and given unique identification numbers.

The structural map section metadata. This would be required to store information about the complex structure of an interpretive journey – hierarchical levels and sub-levels and so on. It would be a key part of interpretive journey metadata – the main thing required to turn an otherwise loosely connected set of constituent digital objects into a coherent journey.

Structural links metadata. This would be required to store information about links between hierarchical levels of an interpretive journey. For example, a link from texts describing a native bird on one web-page to a lower web-page offering an image, a sound file of the bird's song, and an video of it nesting (and probably back again).

Behaviour section metadata. This will be required to store information about executable behaviours associated with constituent parts of an interpretive journey and about the programs required to run them – for example the name and location of a program required to run a video sequence or an animation.

The METS schema meets all of the general requirements of MoPark and the National Park and is sufficiently flexible to allow it to meet the detailed requirements drawn out in Phase II of this metadata options appraisal. Further information on METS and its component parts is included as Appendix A (a tutorial on METS copied from the METS web-site that will be a useful reference document as the project proceeds). Useful information on METS extension schemas developed or being developed by the Library of Congress to deal with a range of digital object types (text files, sound files, images, video files) is listed in Appendix B.

Other Considerations

Two other recommendations can be made on the basis of the analysis of general requirements itemised above.

- The framework should assume the adoption, where possible and appropriate, of national and international standards relevant to the field. For example, in respect of descriptive metadata, it is sensible to investigate whether the use of Library of Congress Subject Headings (LCSH) on the one hand, and of the Dewey Decimal Classification system on the other, provide the best basis for subject description and classification of MoPark and Park digital materials.
- The framework should allow for and facilitate cooperation with other key players with similar needs and interests to harmonise approaches and ensure interoperability beyond Park systems. An obvious partner in this area is the Scottish Cultural Portal Project which has similar aims in a number of areas and an overlap in terms of client groups served. An example of possible cooperation here would be in the area of subject terminology. Even if LCSH were used as the basis of MoPark subject description, it is likely that it would have to be adapted and important that adaptations be harmonised with those made by other Scottish players. This and other related matters should be explored during Phase II, at which time other possible partners may also be identified. For example, it would be sensible to determine whether there is a need to work with other bodies engaged in providing digital learning materials to schools.

Section 3. Metadata Needs (II) – The Need for a Phased Approach

Specifying the Detailed Requirement – The Need for a Phased Approach

The METS schema provides a suitable framework within which to meet the metadata requirements of MoPark and the Park, but there are problems in identifying the detailed requirement at this point in project development. These cannot be readily resolved through discussion and analysis as yet. There is a need to determine other aspects of project detail before it will be possible for project and Park staff to answer the questions that need to be answered before the detailed requirement can be mapped out. A set of questions (see Appendix C) has been compiled to guide this process, but providing answers to these questions is difficult at this stage. The following sample questions illustrate the problem:

- Given a map of an area covered by one or more interpretive journeys, would the journeys and the parts of the journeys all relate to the whole map or would different journeys or digital materials within journeys relate to different areas within the map?
- How would a particular interpretive journey be presented differently in different presentation circumstances (e.g. on a web-site, on a PDA, on a mobile phone, in a static multi-media display)?
- Can you provide examples of how a single interpretive journey might be presented in different ways to different school and other groups?
- Do you have a view on whether the short and long versions of the same text would be separate documents or parts of one bigger document?
- What information are visitors and staff likely to require when retrieving and/or managing digital materials?
- Is there a requirement to store information on the source and reliability of information presented to visitors in an interpretive journey?
- Will different 'versions' of an interpretive journey or a particular digital object (photograph, video file etc) be required for different presentation technologies (e.g. different resolution for a map or an image)?
- Thinking in terms of terrain and Park knowledge, historical knowledge, knowledge of flora and fauna, knowledge of Park objectives, intellectual property rights issues, compliance issues, list as many types of Park staff member as you can think of who may have to be involved in the creation or amendment of an interpretive journey and its release as being 'fit for use'.
- How often is there likely to be a requirement to record information updates related to interpretive journeys (changes – accidents, closure due to upgrades, new information, new events, safety and weather, topic updates etc)? Daily? Weekly? Monthly? Yearly?
- Other than interpretive journeys and their constituent materials, what further types of electronic materials are likely to be stored by the Park (reports, databases, etc)?
- Give a rough estimate of the numbers of interpretive journeys, individual digital objects from interpretive journeys, and digital objects from other Park management and administrative activities, likely to build up over (say) three or five years?
-
- What kinds of management information will the Park and its employees, agents, funders, and advisors require from the digital repository?

All of these questions can be answered now to a certain degree – but only to the level of detail required to determine a general framework as specified above. Neither the project and Park staff, nor the various project consultants (of which the CDLR is one), have sufficient information as yet to answer the questions with any certainty to the level of detail necessary to specify the full requirement within this framework. Doing so will require – at minimum - further work with actual examples of the complex digital objects (interpretive journeys) likely to form the primary elements of the Level 1 and 2 requirements, more experience amongst project and Park personnel of the issues and problems associated with managing complex digital objects, and an in-depth survey of the likely range of other digital objects, their usage, and their life-cycles.

A phased approach to determining the detailed requirements as regards metadata, metadata standards, and metadata management is therefore both necessary and sensible. This approach has been discussed and agreed with the project leaders. A sketch of the proposed forward development path envisaged for Phase II of the *Metadata Options Appraisal* is outlined in section 4 below.

Section 4. The Way Forward - Phase II Outline

Phase II Outline

Identification of the general requirements of MoPark and the Park as regards metadata, metadata standards and metadata management, together with the specification of a suitable framework within which they can be met, the specification of questions to guide the path towards the detailed specification, and the creation of a full draft of this report for discussion with MoPark, their consultants, and National Park staff, entailed the use of 6.5 of the 8.5 consultancy days allotted for the *Metadata Options Appraisal*. The remaining two days, plus one further half day, were utilised in the following ways:

- Meeting and email and telephone discussions with MoPark staff (LF) and consultant (FP).
- Attendance at Balloch project meeting with MoPark staff and Datavisibility staff at which metadata and other issues were discussed.
- Various discussions with Datavisibility on their proposed forward path for DAMS and the relation of this to metadata issues, and on Phase II contractual arrangement proposals, including various telephone calls and a meeting.
- The creation of a presentation on the full draft of the Metadata Options Proposal and its delivery at a meeting of National Park Staff at Balloch Castle. Subsequent discussion of the report and its implications with the staff.
- Completion of the final version of this report.

In addition to conducting the in-depth work required to determine the metadata requirement in detail, Phase II will entail examination of associated issues such as the staffing and training requirements likely to emerge as a result of the project and the long term aim to manage all digital objects used by Park staff both now and in future. Although it is already clear that there will be a requirement for at least one professional with information management and digital asset management skills to oversee the Park's metadata and ensure the adherence to standards needed for the level of management and interoperability likely to be required, more discussion is needed before the requirement can be specified in detail and costed. Whether one professional is enough will depend on the numbers of digital objects ultimately involved and on associated management issues. It will also – inevitably – be influenced by the resources available to the Park for such purposes. These, in turn, may affect decisions on the level of metadata – and, by implication, visitor and staff service level and complexity it would be sensible to maintain. Such questions can only be resolved in the context of the development of the detailed metadata requirement envisaged for Phase II. In all probability, they will also require a decision to be taken at management level within the Park as to whether the best approach would be to develop and maintain an in-house team to manage the digital repository and imetadata, or to contract out either the whole task or specific parts of it to external experts.

Phase II Elements

Taking all of these various considerations into account, it is possible to sketch out a very rough map of the likely work involved in Phase II of the *Metadata Options Appraisal*, as follows:

- Discussions with Project staff on likely shape and form of initial interpretive journey or journeys and on any differences in shape, form, format, and so on likely to be entailed in respect of presenting these to visitors through different output devices (PDAs, web-sites, and so on).
- Discussions with DAMS and content consultants of issues likely to effect metadata requirements for these journeys.
- Initial analysis and proposals in respect of how best to use the METS framework to implement the metadata requirements for these journeys.
- Re-visit each of the above processes on these various issues once the journeys actually exist and can be examined in detail.
- Re-visit each of the above processes on these various issues after the journeys have been field-tested.
- Prepare an initial report on the use of the METS framework for MoPark needs based on these initial investigations.
- In the light of experiences with initial interpretive journeys, conduct in-depth discussions with project staff, the various Park stakeholders and other relevant personnel as required. Aim to scope out likely numbers of interpretive journeys to be implemented in future, any variations likely in shape and form of these, and the likely timescales for their creation and implementation in the field.
- Re-examine and revise the initial report on the use of the METS framework for MoPark needs based on these further investigations.
- Identify Park staff likely to have knowledge of the wider needs of the Park in respect of managing digital objects.
- Discuss implications of MoPark and wider Park needs with appropriate members of Park management to determine the wider management and budgetary contexts within which digital repository developments will progress.
- On the basis of this, sketch out an implementation plan for the Park as a basis for discussion.
- Finalise the details of the probable way forward on this front.
- Finalise the report on the use of the METS framework for Park needs based on these final investigations.
- Utilise the information obtained from these processes to make recommendations on metadata requirements, staffing and training requirements, costs and implementation.
- Incorporate these into a final version of the metadata report.
- If required, compile a training manual and conduct training sessions with Park staff.

Phase II Costs

Costs are difficult to estimate at this stage, but the number of consultancy days required is likely to be significant. Further discussions with project staff and others will be required to pin down the details. One approach might be to operate on a 'pay as you go basis', agreeing and costing each step in turn before proceeding. This should ensure that only time contracted for is charged for. Another approach might be to agree a limit on the number of days and ask for an estimate of what can be done in that time. The advantage of this second approach is that it may make for a more focused approach on both sides. A rough estimate of the number of days required for all of the above processes to be successfully completed is 20-25 days, but this is only a rough estimate. Discussions are required to pin the requirement down and these may result in either a reduction or an expansion of the days likely to be required.

Section 5. Conclusions and References

Conclusions

Within the timescale allotted for the original *Metadata Options Appraisal* it has been possible to determine the general requirements of MoPark and the Loch Lomond and Trossachs National Park in respect of metadata, metadata standards, and metadata management, and to identify a framework within which these requirements can be met. Due to circumstances beyond the control of MoPark and Park personnel and the metadata consultants, it has not been possible to determine the specifics of the metadata requirement at this stage. A phased approach has therefore been agreed with the MoPark project leaders. Phase I – completed with the submission of this report – entails:

- An analysis of the problem in respect of establishing metadata requirements
- An explanation of why a phased approach is necessary
- A proposal in respect of the best framework within which to develop the full requirement
- A recommendation that the full requirement itself be determined in Phase II in the context of the proposed framework.
- A sketch of the elements likely to be entailed in Phase II
- A note on the likely costs of Phase II

The framework proposed has three elements:

- Adoption of the XML-based *Metadata Encoding and Transmission Standard* (METS) developed as an initiative of the Digital Library Federation and maintained in the Network Development and MARC Standards Office of the Library of Congress. METS provides an integrated and coherent infrastructure for all of the metadata types likely to be required within MoPark and the Park – *descriptive metadata*, *administrative metadata* (technical metadata, rights metadata, analogue source information, digital object files provenance), *Files metadata*, *Structural Map metadata*, *Structural Links metadata*, and *Behaviour metadata*. It also provides (see section 2 of this report) for the levels of complexity, structure, and interoperability likely to be required as the digital repository progresses.
- The adoption, where possible and appropriate, of national and international standards relevant to the field (for example, the METS format would allow the use of MARC and LCSH in the descriptive metadata section should Phase II establish the requirement).
- Cooperation with other key players with similar needs and interests to harmonise approaches and ensure interoperability beyond Park systems (For example, the Scottish Cultural Portal project which has needs and a client group that overlap with those of MoPark).

An outline of proposals for Phase II is provided in section 4 of this report.

References

In addition to drawing on lessons and documentation related to various digital library projects in which CDLR is involved (Victorian Times, Glasgow Digital Library and Scottish Cultural Portal in particular), a range of digital library and digital library metadata sites were visited and studied in the process of conducting this appraisal. Those found to be of particular value are listed here:

Digital Library Federation (2004) Digital Library Federation Web-site Retrieved March 17, 2004 from <http://www.diglib.org/>

Gartner, Richard (2002). METS: Metadata Encoding and Transmission Standard. *JISC Techwatch report* TSW 02-05, October 2002. Retrieved March 17, 2004 from http://www.jisc.ac.uk/index.cfm?name=techwatch_report_0205

Gartner, Richard (2003). MODS: Metadata Object Description Schema. *JISC Techwatch report* TSW 03-06, 2003. Retrieved March 17, 2004 from http://www.jisc.ac.uk/index.cfm?name=techwatch_report_0306

International Federation of Library Associations. IFLA DIGITAL LIBRARIES: Metadata Resources webpages. Retrieved March 17, 2004 from <http://www.ifla.org/ll/metadata.htm>

Library of Congress. AV Prototype Project Working Documents Extension Schemas for the Metadata Encoding and Transmission Standard. Retrieved March 17, 2004 from <http://lcweb.loc.gov/rr/mopic/avprot/metsmenu2.html>

Library of Congress (2004). METS Official Web-site. Retrieved March 17, 2004 from <http://www.loc.gov/standards/mets/>

Library of Congress Help Desk (2003). METS: An Overview & Tutorial. Retrieved March 17, 2004 from <http://www.loc.gov/standards/mets/METSOverview.html>

Library of Congress. VIDEOMD: Video Technical Metadata Extension Schema Initial Version: VMD_020322.xsd. Retrieved March 17, 2004 from <http://lcweb2.loc.gov/mets/Schemas/VMD.xsd>

Appendix A: METS: An Overview & Tutorial (Copied from METS web-site)

METS: An Overview & Tutorial

Introduction

Maintaining a library of digital objects of necessity requires maintaining metadata about those objects. The metadata necessary for successful management and use of digital objects is both more extensive than and different from the metadata used for managing collections of printed works and other physical materials. While a library may record descriptive metadata regarding a book in its collection, the book will not dissolve into a series of unconnected pages if the library fails to record structural metadata regarding the book's organization, nor will scholars be unable to evaluate the book's worth if the library fails to note that the book was produced using a Ryobi offset press. The same cannot be said for a digital version of the same book. Without structural metadata, the page image or text files comprising the digital work are of little use, and without technical metadata regarding the digitization process, scholars may be unsure of how accurate a reflection of the original the digital version provides. For internal management purposes, a library must have access to appropriate technical metadata in order to periodically refresh and migrate the data, ensuring the durability of valuable resources.

The [Making of America II](#) project (MOA2) attempted to address these issues in part by providing an encoding format for descriptive, administrative, and structural metadata for textual and image-based works. METS, a [Digital Library Federation](#) initiative, attempts to build upon the work of MOA2 and provide an XML document format for encoding metadata necessary for both management of digital library objects within a repository and exchange of such objects between repositories (or between repositories and their users). Depending on its use, a METS document could be used in the role of Submission Information Package (SIP), Archival Information Package (AIP), or Dissemination Information Package (DIP) within the [Open Archival Information System \(OAIS\) Reference Model](#).

A METS document consists of seven major sections:

1. [METS Header](#) - The METS Header contains metadata describing the METS document itself, including such information as creator, editor, etc.
2. [Descriptive Metadata](#) - The descriptive metadata section may point to descriptive metadata external to the METS document (e.g., a MARC record in an OPAC or an EAD finding aid maintained on a WWW server), or contain internally embedded descriptive metadata, or both. Multiple instances of both external and internal descriptive metadata may be included in the descriptive metadata section.
3. [Administrative Metadata](#) - The administrative metadata section provides information regarding how the files were created and stored, intellectual property rights, metadata regarding the original source

object from which the digital library object derives, and information regarding the provenance of the files comprising the digital library object (i.e., master/derivative file relationships, and migration/transformation information). As with descriptive metadata, administrative metadata may be either external to the METS document, or encoded internally.

4. **File Section** - The file section lists all files containing content which comprise the electronic versions of the digital object. <file> elements may be grouped within <fileGrp> elements, to provide for subdividing the files by object version.
5. **Structural Map** - The structural map is the heart of a METS document. It outlines a hierarchical structure for the digital library object, and links the elements of that structure to content files and metadata that pertain to each element.
6. **Structural Links** - The Structural Links section of METS allows METS creators to record the existence of hyperlinks between nodes in the hierarchy outlined in the Structural Map. This is of particular value in using METS to archive Websites.
7. **Behavior** - A behavior section can be used to associate executable behaviors with content in the METS object. Each behavior within a behavior section has an interface definition element that represents an abstract definition of the set of behaviors represented by a particular behavior section. Each behavior also has a mechanism element which identifies a module of executable code that implements and runs the behaviors defined abstractly by the interface definition.

A more detailed explanation of each section and their inter-relations follows.

METS Header

The METS Header element allows you to record minimal descriptive metadata about the METS object itself within the METS document. This metadata includes the date of creation for the METS document, the date of its last modification, and a status for the METS document. You may also record the names of one or more agents who have played some role with respect to the METS document, specify the role they have played, and add a small note regarding their activity. Finally, you may record a variety of alternative identifiers for the METS document to supplement the primary identifier for the METS document recorded in the OBJID attribute on the METS root element. A small example of a METS Header might look like the following:

```
<metsHdr CREATEDATE="2003-07-04T15:00:00" RECORDSTATUS="Complete">
  <agent ROLE="CREATOR" TYPE="INDIVIDUAL">
    <name>Jerome McDonough</name>
  </agent>
  <agent ROLE="ARCHIVIST" TYPE="INDIVIDUAL">
    <name>Ann Butler</name>
  </agent>
</metsHdr>
```

This example contains two attributes on the <metsHdr> element, CREATEDATE and RECORDSTATUS, which are used to indicate the date and time the METS record was created, and indicate the status of the record's processing. Two individual agents are listed who have worked on this METS record, the person responsible for creating the record and an archivist responsible for the original material. Both the ROLE and TYPE attributes on the <agent> element employ controlled vocabularies. Allowable values for ROLE include "ARCHIVIST," "CREATOR," "CUSTODIAN," "DISSEMINATOR," "EDITOR," "IPOWNER" and "OTHER." Allowable values for the TYPE attribute are "INDIVIDUAL," "ORGANIZATION" or "OTHER."

Descriptive Metadata

The descriptive metadata section of a METS document consists of one or more <dmdSec> (Descriptive Metadata Section) elements. Each <dmdSec> element may contain a pointer to external metadata (an <mdRef> element), internally embedded metadata (within an <mdWrap> element), or both.

External Descriptive Metadata (mdRef): an mdRef element provides a URI which may be used in retrieving the external metadata. For example, the following metadata reference points to the finding aid for a particular digital library object:

```
<dmdSec ID="dmd001">
  <mdRef LOCTYPE="URN" MIMETYPE="application/xml" MDTYPE="EAD"
    LABEL="Berol Collection Finding Aid">urn:x-nyu:fales1735</mdRef>
</dmdSec>
```

The <mdRef> element of this <dmdSec> contains four attributes. The LOCTYPE attribute specifies the type of locator contained in body of the element; valid values for LOCTYPE include 'URN,' 'URL,' 'PURL,' 'HANDLE,' 'DOI,' and 'OTHER.' The MIMETYPE attribute allows you to specify the MIME type for the external descriptive metadata, and the MDTYPE allows you to indicate what form of metadata is being referenced. Valid values for the MDTYPE element include MARC, MODS, EAD, VRA (VRA Core), DC (Dublin Core), NISOIMG (NISO Technical Metadata for Digital Still Images), LC-AV (Library of Congress Audiovisual Metadata), TEIHDR (TEI Header), DDI (Data Documentation Initiative), FGDC (Federal Geographic Data Committee Metadata Standard [FGDC-STD-001-1998]), and OTHER. LABEL provides a mechanism for describing this metadata to those viewing a METS document, in a 'Table of Contents' display of the METS document, for example.

Internal Descriptive Metadata (mdWrap): An mdWrap element provides a wrapper around metadata embedded within a METS document. Such metadata can be in one of two forms: 1. XML-encoded metadata, with the XML-encoding identifying itself as belonging to a namespace other than the METS document namespace, or 2. any arbitrary binary or textual form, PROVIDED that the metadata is Base64 encoded and wrapped in a <binData> element within the mdWrap element. The following examples demonstrate the use of the mdWrap element:

```

<dmdSec ID="dmd002">
  <mdWrap MIMETYPE="text/xml" MDTYPE="DC" LABEL="Dublin Core Metadata">
    <xmlData>
      <dc:title>Alice's Adventures in Wonderland</dc:title>
      <dc:creator>Lewis Carroll</dc:creator>
      <dc:date>between 1872 and 1890</dc:date>
      <dc:publisher>McCloughlin Brothers</dc:publisher>
      <dc:type>text</dc:type>
    </xmlData>
  </mdWrap>
</dmdSec>

<dmdSec ID="dmd003">
  <mdWrap MIMETYPE="application/marc" MDTYPE="MARC" LABEL="OPAC
Record">
    <binData>MDI0ODdjam0glDIyMDA1ODkgYSA0NU0wMDAxMDA...(etc.)
    </binData>
  </mdWrap>
</dmdSec>

```

Note that all <dmdSec> elements must possess an ID attribute. This attribute provides a unique, internal name for each <dmdSec> element which can be used in the structural map to link a particular division of the document hierarchy to a particular <dmdSec> element. This allows specific sections of descriptive metadata to be linked to specific parts of the digital object.

Administrative Metadata

<amdSec> elements contain the administrative metadata pertaining to the files comprising a digital library object, as well as that pertaining to the original source material used to create the object. There are four main forms of administrative metadata provided for in a METS document: 1. Technical Metadata (information regarding files' creation, format, and use characteristics), 2. Intellectual Property Rights Metadata (copyright and license information), 3. Source Metadata (descriptive and administrative metadata regarding the analog source from which a digital library object derives), and 4. Digital Provenance Metadata (information regarding source/destination relationships between files, including master/derivative relationships between files and information regarding migrations/transformations employed on files between original digitization of an artifact and its current incarnation as a digital library object). Each of these four different types of administrative metadata has a unique subelement within the <amdSec> portion of a METS document in which that form of metadata can be embedded: <techMD>, <rightsMD>, <sourceMD>, and <digiprovMD>. Each of these four elements may occur more than once in any METS document.

The <techMD>, <rightsMD>, <sourceMD> and <digiprovMD> elements employ the same content model as <dmdSec>: they may contain an <mdRef> element to point to external administrative metadata, an <mdWrap> element to use when embedding administrative metadata within a METS document, or

both. Multiple instances of these elements may occur within a METS document, and all of them must carry an ID attribute so that other elements within the METS document (such as divisions within the structural map or <file> elements) may be linked to the <amdSec> subelements which describe them. One might, for example, have an <techMD> element which includes technical metadata regarding a file's preparation:

```

    <techMD ID="AMD001">
      <mdWrap MIMETYPE="text/xml" MDTYPE="NISOIMG" LABEL="NISO Img.
Data">
        <xmlData>
          <niso:MIMETYPE>image/tiff</niso:MIMETYPE>
          <niso:Compression>LZW</niso:Compression>
          <niso:PhotometricInterpretation>8</niso:PhotometricInterpretation>
          <niso:Orientation>1</niso:Orientation>
          <niso:ScanningAgency>NYU Press</niso:ScanningAgency>
        </xmlData>
      </mdWrap>
    </techMD>

```

A <file> element within a <fileGrp> might then identify this administrative metadata as pertaining to the file it identifies by using an ADMID attribute to point to this <techMD> element:

```

<file ID="FILE001" ADMID="AMD001">
  <FLocat LOCTYPE="URL">http://dlib.nyu.edu/press/testimg.tif</FLocat>
</file>

```

File Section

The file section (<fileSec>) contains one or more <fileGrp> elements used to group together related files. A <fileGrp> lists all of the files which comprise a single electronic version of the digital library object. For example, there might be separate <fileGrp> elements for the thumbnails, the master archival images, the pdf versions, the TEI encoded text versions, etc.

Consider the following example of a file section from a digital library object for an oral history which has three different versions: a TEI-encoded transcript, a master audio file in WAV format, and a derivative audio file in MP3 format:

```

<fileSec>
  <fileGrp ID="VERS1">
    <file ID="FILE001" MIMETYPE="application/xml" SIZE="257537"
CREATED="2001-06-10">
      <FLocat LOCTYPE="URL">http://dlib.nyu.edu/tamwag/beame.xml</FLocat>
    </file>
  </fileGrp>
  <fileGrp ID="VERS2">
    <file ID="FILE002" MIMETYPE="audio/wav" SIZE="64232836"
      CREATED="2001-05-17" GROUPID="AUDIO1">
      <FLocat LOCTYPE="URL">http://dlib.nyu.edu/tamwag/beame.wav</FLocat>
    </file>
  </fileGrp>
  <fileGrp ID="VERS3" VERSDATE="2001-05-18">
    <file ID="FILE003" MIMETYPE="audio/mpeg" SIZE="8238866"
      CREATED="2001-05-18" GROUPID="AUDIO1">

```

```
<FLocat LOCTYPE="URL">http://dlib.nyu.edu/tamwag/beame.mp3</FLocat>
</file>
</fileGrp>
</fileSec>
```

In this case, the <fileSec> contains three subsidiary <fileGrp> elements, one for each different version of the object. The first is an XML-encoded transcription file, the second is a master audio file in WAV format, and the third is a derivative audio file in MP3 format. While such a basic example does not really seem to need the <fileGrp> elements to distinguish the different versions of the object, <fileGrp> becomes much more useful for objects consisting of large numbers of scanned page images, or indeed any case where a single version of the object consists of a large number of files. In those cases, being able to separate <file> elements into <fileGrp>s makes identifying the files belonging to a particular version of the document a simple task.

You may note the presence of the GROUPID attributes with identical values on the two audio <file> elements; these indicate that the two files, while belonging to different versions of the object, contain the same basic information (you can use the GROUPID for the same purpose to indicate equivalent page image files in digital library objects containing many scanned page images).

You should also note that all of the <file> elements have a unique ID attribute. This attribute provides a unique, internal name for this file which can be referenced by other portions of the document. You'll see this type of referencing in action when we look at the Structural Map Section.

It should be mentioned that <file> elements may possess an <FContent> element rather than an <FLocat> element. <FContent> elements are used to embed the actual contents of the file within the METS document; if this is done, the file contents must either be in XML format or be Base64-encoded. While embedding files is not something one would typically do when preparing a METS document for use in displaying a digital library objects to users, it can be a valuable feature for exchanging digital library objects between repositories, or for archiving versions of digital library objects for off-site storage.

Structural Map

The structural map section of a METS document defines a hierarchical structure which can be presented to users of the digital library object to allow them to navigate through it. The <structMap> element encodes this hierarchy as a nested series of <div> elements. Each <div> carries attribute information specifying what kind of division it is, and also may contain multiple METS pointer (<mptr>) and file pointer (<fptr>) elements to identify content corresponding with that <div>. METS pointers specify separate METS documents as containing the relevant file information for the <div> containing them. This can be useful when encoding large collections of material (e.g., an

entire journal run) to keep the size of each METS file in the set relatively small. File pointers specify files (or in some cases either groups of files or specific locations within a file) within the current METS document's <fileSec> section that correspond to the portion in the hierarchy represented by the current <div>.

The following provides an example of an extremely simple structural map:

```

<structMap TYPE="logical">
  <div ID="div1" LABEL="Oral History: Mayor Abraham Beame"
    TYPE="oral history">
    <div ID="div1.1" LABEL="Interviewer Introduction"
      ORDER="1">
      <fptr FILEID="FILE001">
        <area FILEID="FILE001" BEGIN="INTVWBG" END="INTVWND"
          BETYPE="IDREF" />
        </fptr>
      <fptr FILEID="FILE002">
        <area FILEID="FILE002" BEGIN="00:00:00" END="00:01:47"
          BETYPE="TIME" />
        </fptr>
      <fptr FILEID="FILE003">
        <area FILEID="FILE003" BEGIN="00:00:00" END="00:01:47"
          BETYPE="TIME" />
        </fptr>
      </div>
    <div ID="div1.2" LABEL="Family History" ORDER="2">
      <fptr FILEID="FILE001">
        <area FILEID="FILE001" BEGIN="FHBG" END="FHND"
          BETYPE="IDREF" />
        </fptr>
      <fptr FILEID="FILE002">
        <area FILEID="FILE002" BEGIN="00:01:48"END="00:06:17"
          BETYPE="TIME" />
        </fptr>
      <fptr FILEID="FILE003">
        <area FILEID="FILE003" BEGIN="00:01:48" END="00:06:17"
          BETYPE="TIME" />
        </fptr>
      </div>
    <div ID="div1.3" LABEL="Introduction to Teachers' Union"
      ORDER="3">
      <fptr FILEID="FILE001">
        <area FILEID="FILE001" BEGIN="TUBG" END="TUND"
          BETYPE="IDREF" />
        </fptr>
      <fptr FILEID="FILE002">
        <area FILEID="FILE002" BEGIN="00:06:18" END="00:10:03"
          BETYPE="TIME" />
        </fptr>
      <fptr FILEID="FILE003">
        <area FILEID="FILE003" BEGIN="00:06:18" END="00:10:03"
          BETYPE="TIME" />
        </fptr>
      </div>
    </div>
  </structMap>

```

This structural map shows that we have an oral history (with Mayor Abraham Beame of New York City) that includes three subsections: an opening introduction by the interviewer, some family history from Mayor Beame, and a discussion of how he came to be involved with the teachers' union in New York. Each of these subsections/divisions is linked to three files (taken from our earlier example of file groups): an XML transcription, and a master and derivative audio file. A subsidiary <area> element is used in each <fptr> to indicate that this division corresponds with only a portion of the linked file, and to identify the exact portion of each linked file. For example, the first division (the interviewer introduction) is linked to a portion of the XML transcription file (FILE001) which is found between the two tags in the transcription file with ID attribute values of "INTVWBG" and "INTVWND." It is also linked to the two different audio files; in these cases, rather than specifying ID attribute values within the linked files, the begin and end points of the linked material within the files is indicated by a simple time code value of the form HH:MM:SS. So, the interviewer introduction can be found in both audio files in the segment beginning at time 00:00:00 in the file and extending through time 00:01:47.

Structural Links

The structural links section of the METS format is the simplest in form of any of the major METS sections, containing only a single element, <smLink> (although that element may be repeated). The structural links section of METS is intended to allow you to record the existence of hyperlinks between items within the structural map, usually <div> elements. This is a useful facility if you wish to use METS to archive web sites, and wish to maintain a record of the hypertext structure of the sites separately from the HTML files of the site itself.

As an example, consider the case of a METS document for a web page containing an image which is hyperlinked to another page. The <structMap> element would probably contain <divs> like the following for the two pages:

```
<div ID="P1" TYPE="page" LABEL="Page 1">
  <fptr FILEID="HTMLF1"/>
    <div ID="IMG1" TYPE="image" LABEL="Image Hyperlink to
      Page 2">
      <fptr FILEID="JPGF1"/>
    </div>
</div>

<div ID="P2" TYPE="page" LABEL="Page 2">
  <fptr FILEID="HTMLF2"/>
</div>
```

If you wished to indicate that the image file in the <div> contained with the first page <div> is hyperlinked to the HTML file in the second page <div>, you would have a <smLink> element within the <structLink> section of the METS document as follows:

```
<smLink from="IMG1" to="P2" xlink:title="Hyperlink from
JPEG Image on Page 1 to Page 2" xlink:show="new"
```

```
xlink:actuate="onRequest" />
```

The <smLink> link element above uses a slightly modified form of the XLink syntax; all of the XLink attributes are used, but the "to" and "from" attributes are declared to be of type IDREF rather than NMTOKEN as in the original XLink specification. This allows you to indicate the existence of links between any two nodes in the structural map, and also use XML processing tools to confirm that the linked nodes actually exist.

Behavior Section

A behavior section can be used to associate executable behaviors with content in the METS object. A behavior section contains one or more <behavior> elements, each of which has an interface definition element that represents an abstract definition of the set of behaviors represented by a particular behavior section. A <behavior> also has a <mechanism> element which is used to point to a module of executable code that implements and runs the behavior defined abstractly by the interface definition.

Digital object behaviors can be implemented as linkages to distributed web services as in the following example from the Mellon [Fedora](#) project.

```
<METS:behavior ID="DISS1.1" STRUCTID="S1.1" BTYPE="uva-bdef:stdImage"
  CREATED="2002-05-25T08:32:00" LABEL="UVA Std Image Disseminator"
  GROUPLD="DISS1" ADMID="AUDREC1">
  <METS:interfaceDef LABEL="UVA Standard Image Behavior Definition"
    LOCTYPE="URN" xlink:href="uva-bdef:stdImage"/>
  <METS:mechanism LABEL="A NEW AND IMPROVED Image Mechanism"
    LOCTYPE="URN" xlink:href="uva-bmech:BETTER-imageMech"/>
</METS:behavior>
```

See also:

- [The Fedora Technical Specification](#) (pdf)
- [Sample Digital Object](#) (encoded using METS)
- [Sample Behavior Definition Object](#) (encoded using METS)
- [Sample Behavior Mechanism Object](#) (encoded using METS)

Conclusion

The METS schema provides a flexible mechanism for encoding descriptive, administrative, and structural metadata for a digital library object, and for expressing the complex links between these various forms of metadata. It can therefore provide a useful standard for the exchange of digital library objects between repositories. In addition, METS provides the ability to associate a digital object with behaviours or services. The above discussion highlights the major features of the schema, but a thorough examination of the schema and its included documentation is necessary to understand the full range of its capabilities.

[METS Home Page](#) - [Library of Congress Standards](#) - [Library of Congress Home](#)



The Library of Congress
[Library of Congress Help Desk](#) (July 18, 2003)

Appendix B: METS Extensions [Copied from Library of Congress Web-site]

METS Extension Schemas In Development at the Library of Congress

AV Prototype Project Working Documents

Extension Schemas for the Metadata Encoding and Transmission Standard Main Revision February 2003, Added Information January 2004

[Introduction](#)

[Table Showing Arrangement of Schemas](#)

Introduction. The Library of Congress Audio-Visual Prototyping Project (hereafter *AV Project*) will use the emerging [Metadata Encoding and Transmission Standard](#) (METS) to encode the metadata for digital objects. The METS website includes an [Overview and Tutorial](#) about the *primary schema* that is intended to serve as a framework document and explains that users may select *extension schemas* in order to provide additional metadata. The [table](#) below indicates the arrangement of the primary and extension schemas proposed for use in the AV Project as of February 2003.

In the [table](#), there are links to schema *xsd* files and to a data dictionary for the various attributes and elements. Please note that the data dictionary is *really* for the relational database in which the metadata is initially captured and from which the XML instances are generated.

The audio-file technical information ([schema](#), [data dictionary](#)) owes a great debt to work being carried out as a part of the Harvard University Library Digital Initiative and as an activity of the Audio Engineering Society. The same source provided the impetus for the *digiprov schema* (see also [data dictionary](#)) presented here, although this version is somewhat simplified from the Harvard-AES example (see [digiprov explanation](#)). The text technical metadata ([schema](#), [documentation](#), and [data dictionary](#)) is the work of Jerome McDonough, Elmer Bobst Library, New York University. The video-file technical information ([schema](#), [data dictionary](#)) is very tentative and awaits improvement by video engineering specialists.

Table Showing Arrangement of METS Primary Schema and Extension Schemas Proposed for AV Project.

Primary Schema		Extension schema and comment	
<mets>			For more information, visit the METS website.

	<dmdSec> Descriptive Metadata Section		MODS Schema	For more information, visit the MODS website.
	<amdSec> Administrative Metadata Section	<techMD> Technical Metadata (about "files")	AMD Schema (same schema as under sourceMD) DB data dict (file data only)	Information specific to audio files, e.g., sampling frequency.
			MIX Schema DB data dict	Information specific to image files, e.g., bits per pixel, color space. For more information, visit the MIX website. Note: for the time being, the AV Project is using a variant of MIX, with some added enumerated values and other slight changes.
			Text technical metadata (same schema as under sourceMD) Documentation DB data dict	Information specific to text files, e.g., character set, encoding used.
			VMD Schema (same schema as under sourceMD) DB data dict (file data only)	Information specific to video files, e.g., bit rate, compression codec.
		<rightsMD> Rights and Access Management	RMD Schema DB data dict	Rights, restrictions, and/or other categorizing information that can be used to support rights-

		Metadata		management and/or access-management systems.
		<sourceMD> Metadata about the Source for the File in this METS object	AMD Schema (same schema as under techMD, above) DB data dict (file and physical) Descriptive terms from MAVIS for reference.	Information specific to audio items presented for file-format migration or to be digitized, e.g., channel or track specifications, sampling frequency.
			IMD Schema (for use when source is physical item) DB data dict (IMD, file and physical) <hr/> MIX Schema (same schema as under techMD above; for use when source is a digital file) DB data dict (MIX)	<p>Information specific to items scanned to produce images (IMD), e.g., type or condition, of items presented for file-format migration (MIX).</p> <p>Note: for the time being, the AV Project is using a variant of MIX, with some added enumerated values and other slight changes.</p>
			Text technical metadata (same schema as under techMD, above) Documentation DB data dict	<p>Information specific to text files presented for file-format migration, e.g., character set, encoding.</p> <p>Note: the project assumes that scanning a paper items with typography produces an <i>image</i>, and that text conversion methodology is</p>

				covered in digiProvMD (below).
			VMD Schema (same schema as under techMD, above) DB data dict (file and physical) Descriptive terms from MAVIS for reference.	Information specific to video items presented for file-format migration or to be digitized, e.g., bit rate, tape stock information.
		<digiProvMD> Metadata about the Process Used to Create the File	PMD schema Explanation DB data dict	About the events/steps/processes that occurred in reformatting or migrating entities.
	<behaviorSec> Behavior Section			Section designed to identify digital object behaviors; for example as may be used in the FEDORA repository architecture.
	<fileGrp> File Groups Section			The inventory of files that are part of this digital object, with location pointers.
	<structMap> Structural Map Section			The metadata required to present and navigate this digital object.
</mets>				

Go to [top](#)
 Go to [AV Prototype Project Documents](#)
 Go to [AV Prototype Project Home](#)

(1/5/04)

Appendix C: Phase II Metadata Detail Questionnaire

This questionnaire will be used to guide discussion and analysis on detailed metadata requirements in Phase II of the Metadata Options Appraisal.

Metadata Options Appraisal: Questionnaire to Guide Phase II Discussion and Analysis

Introductory Remarks

The Metadata requirements of any system are primarily related to function. In order to determine the requirement for MoPark and for the National Park, it is necessary first to determine what the requirements are in respect of handling digital information and information products. In rough terms, to determine *which* things need to be done to or with *what* information and information products, by or for *whom*, in what circumstances (*when, how, where*), and for what purpose (*why*) in order to meet the management and service requirements of the system. The purpose of the stakeholder consultations is to determine reasonably detailed answers to these questions in respect of the needs of MoPark and the National Park. Other issues affecting decisions on metadata include resources available for implementation and ongoing management, and existing constraints in terms of the existence of legacy metadata repositories and legacy or systems used by organisational partners. It is likely that different types of metadata will be required for different purposes – item-level metadata for individual objects such as images or text files, ‘article-level’ metadata for groups of objects related under a single theme, collection level metadata for ‘journeys’ that encompass groups of items and ‘articles’.

Note that, based on earlier consultations on this matter there is an assumption that there are three levels of requirement as regards metadata:

- MoPark project requirements (Level 1)
- Requirements if MoPark extended to other topics and Park areas (Level 2)
- General requirements of the National Park beyond MoPark (Level 3)

Questions

Introductory Question

Do you have any general comments to make on the paragraph above under ‘Introductory Remarks’?

Interpretive Journeys and Any Similar Collections of Materials Planned

Is it likely that the various parts of an interpretive journey will have to be presented in a particular order to be understandable?

Will visitor choice of mode of transport, preferred terrain, language, visitor type affect which aspects of the interpretive journey will be presented? What other visitor choices might effect this?

Will the digital objects that make up (say) an interpretive journey need to be linked to the National Park plan and to specific objectives?

What other types of electronic materials are likely to be stored by the Park (reports, databases, etc)?

Area Maps and GPS Issues

Will all spatial and map data reside at a central location and be accessible to mobile and 'on location' units via a network connection? Or will some be downloaded and stored on a local device? Give examples if you can.

In what ways will interpretive journey materials interact with spatial and map data and metadata? Please be as specific as your level of expertise in this area allows.

Given a map of an area covered by one or more interpretive journeys, would the journeys and the parts of the journeys all relate to the whole map or would different journeys or digital materials within journeys relate to different areas within the map?

Presentation Methods

The Park may use PDAs, mobile phones, multimedia presentations on solar boats, website(s), fixed 'posts', leaflets, posters, and promotional videos as methods of presenting materials held digitally. Is anything missing from the list?

How would a particular interpretive journey be presented differently in different presentation circumstances (e.g. on a web-site, on a PDA, on a mobile phone, in a static multi-media display)?

User Interface and Response Issues

If possible, provide examples of how a single interpretive journey might be presented in different ways to different school and other groups. What parts of the sample journey would differ, what parts would stay the same?

What different visitor groups would you cater for? In particular, are schoolchildren sub-divided into only 'primary' and 'secondary' or on some other basis (please specify)?

Would there be a requirement for Park interpretive journeys and similar to be compatible with the requirements of the Scottish curriculum and the English curriculum?

Would there be a requirement to permit school visitors to retrieve other curriculum based electronic educational materials associated with Park educational materials?

What other things would affect how the journeys are presented to visitors? Would these include type of terrain, transport mode, time available, visitor search topics or browse lists, highlights for the current location, educational material, recreational opportunities, time or season of year, time of day? What is missing from this list?

Would essentially the same information be presented in simple and brief form to some visitors and complex and longer form to others?

Would essentially the same information be presented in simple and brief form to visitors choosing one type of terrain or one route and complex and longer form to visitors choosing another type of terrain or a different route?

Do you have a view on whether the short and long versions of the same text would be separate documents or parts of one bigger document?

Is it correct that variations of the kind just described might be presented to visitors based both on their own choices and on an automated mechanism associated with GPS-based location?

Would/should the Park provide facilities to permit visitors to annotate journeys or parts of journeys or make further use of the digital content (e.g. in school work, on their own web-sites, and so on)?

The Park plans to offer web-based services to 'visitors' pre and post visit. How would this work in practice? Give at least one example.

Taking into account all of the functionality described above, what special needs issues need to be taken account of (e.g. visually/audio impaired visitors)?

Staff Interface and Response Issues

What are the different staff groups that would use the Park's digital repository?

If you are a member of staff at the Park, or work with the Park in some professional and/or advisory capacity, please give examples of the kind of electronic information you are likely to use from the Park digital repository and of some way or ways in which you would use them.

What internal skills and expertise exist in Park staffing complement as regards managing web-sites, digital objects, cataloguing, subject-matter expertise?

Thinking in terms of terrain and Park knowledge, historical knowledge, knowledge of flora and fauna, knowledge of Park objectives, intellectual property rights issues, compliance issues, list as many types of Park staff member as you can think of who may have to be involved in the creation or amendment of an interpretive journey and its release as being 'fit for use'.

What additional facilities and information might be needed for a ranger-led approach to particular digital presentation of an interpretive journey (as opposed to a visitor DIY approach)?

Do you have a role in managing digital or electronic information services in the Park? If so, please describe the role briefly and say something about the expertise you employ in carrying it out (brief notes or bullet points are fine).

Should the record for an interpretive journey store details of the links between it Park aims and objectives? Should the records for the individual objects that make up the journey store such information?

If Park aims and objectives change, or someone alters the aims of the journey, is an automatic notification to one or more staff members a requirement?

How often is there likely to be a requirement to record information updates related to interpretive journeys (changes – accidents, closure due to upgrades, new information, new events, safety and weather, topic updates etc)? Daily? Weekly? Monthly? Yearly?

Descriptive Metadata Issues

What information are visitors and staff likely to require when retrieving and/or managing digital materials? Examples are Author or Creator, Title, Subject, Location, Language, Rights. Try to think of something that you or a visitor group might need that is less than obvious to an outsider.

Can you give some examples of indexing terms/themes/topics likely to be used by tourists and other visitors to the park? Is it likely that you will have to offer different subject descriptions for different groups of visitors? Is Social Inclusion and issue here?

Can you give some examples of indexing terms/themes/topics likely to be used by different types of park staff?

Thinking about both staff and users, how wide-ranging are the subjects likely to be covered in the digital repository? To what extent do they cover the following broad headings: Literature, History, Geography, Science, Technology, Philosophy, Languages, Social Issues, Education, Scottish culture in general, Tourist issues, Economics. Give some examples of topics you would personally expect to find in the repository.

Is there a requirement to allow visitors to search or browse digital materials by subjects, topics or themes expressed in alternative languages? If so, which languages? Gaelic? Spanish? French? Chinese?

Does the Park have source lists for bird, plant, animal, fish names etc?

Administrative Metadata Issues

Is there a requirement to store information on the source and reliability of information presented to visitors in an interpretive journey? Give examples of how this would be used.

Describe some common situations likely to be encountered in respect of intellectual property rights of materials used in journeys.

What management and access issues are thrown up by the Freedom of Information Act – who will have the right to use MoPark/Park material? How will staff manage the implications? What problems will arise?

Will locational aspects of interpretive journeys provide safety information? What kinds? How frequently might the information change?

What kinds of management information will the Park and its employees, agents, funders, and advisors require from the digital repository?

Will the Park aim to preserve materials in the longer term, and will this include a requirement to be able to offer access to the materials even when software and hardware technologies change?

Technical/Structural Metadata Issues

What different media types are likely to be used by the Park (Images, Video clips, audio clips, text, and so on). What particular types of information might especially need to be recorded for each type (e.g. video or other format, resolution, file-size).

Is an interpretive journey made up of a number of different but *linked* digital objects (a video, some photographs, some text, audio description and orientation, maps, web-site links, information from a database, perhaps even external information such as weather reports)?

Would the *sense* of an interpretive journey be evident if all you had was a collection of these digital objects but no information about how they were linked in the journey? In other words, are the links and the structure vital to an understanding of the purpose and sense of the interpretive journey.

Integration and Standards Issues

Links between (e.g.) PDAs and web-site access at different Discovery Centres
Links to related materials elsewhere for further research (e.g. Scottish Film Archive, NLS)
Any requirement to harmonise with approaches to metadata taken by others in the organisation, project partners,

If you can, please say how the central repository of digital materials will 'feed' the various presentation technologies specified above (PDAs, fixed posts, solar boats etc) – live over a network, daily downloads, what?

Will different 'versions' of an interpretive journey or a particular digital object (photograph, video file etc) be required for different presentation technologies (e.g. different resolution for a map or an image) ?

Is the Park required to follow particular standards such as the e-GIF e-Government Interoperability Framework, e-GMS, the e-Government Metadata Standard, the system and metadata requirements of the Public Records Office?

What other requirements are laid on the Park in this respect (e.g. Disability Discrimination Act, requirements, W3C accessibility standards)?

If you can, give further details of the 'Aberfoyle system', 'Sheila's Photograph Library', 'Loch Lomond Inventory' web-site. How many of these systems store metadata or catalogue records? Roughly how many records does each store? Can you provide examples of records held?

What are the various 'existing systems' the 'Interpretive Warehouse' will have to inter-operate with? Are any of these outwith the Park's organisational structures (e.g. Scottish Film Archive, NLS)? Please give details or say who can give details.

Will MoPark have to interoperate with project partners in other countries? If so, please give details.

Is there likely to be a requirement to provide visitors with access to digital materials held by other services (The Scottish Cultural Portal, SCRAN, The National Library of Scotland, Museums repositories)?

Additional Questions

Are there similar projects elsewhere whose experience on metadata needs might prove valuable?

Give a rough estimate of the numbers of interpretive journeys, individual digital objects from interpretive journeys, and digital objects from other Park management and administrative activities, likely to build up over (say) three or five years?

Can you think of a question we should have but did not ask? If so, say what it is and, if possible, how you would answer it.