Learning Material Application Profile Scoping Study – final report

Author: Phil Barker, JISC CETIS & ICBL, Heriot-Watt University.
http://www.icbl.hw.ac.uk/~philb/

Date: 09 December 2008.

Project URL: http://www.icbl.hw.ac.uk/lmap/

Acknowledgements

This scoping study relied heavily on the expertise of others, especially the informants and expert working group listed in Appendix 3. I am grateful for their help. I would also like to thank everyone from the JISC, the CETIS Metadata SIG, the Dublin Core Education community and other organizations who commented on the ideas presented here and on drafts of this report; I apologise for not being able to list your names individually here. Special thanks goes to Lorna Campbell for her role in overseeing this project and the writing of this report.

Any errors in this report are my own responsibility and should not reflect on the assistance provided by those acknowledged above.

Funding for this project was provided by JISC.

Phil Barker, December 2008.

Table of Contents

1. Summary

This report details the findings of a scoping study carried out for the JISC to investigate a potential metadata application profile for learning materials. The objective of the study was to synthesize and analyse the advice that is currently available to managers of repositories containing educational materials who need to define a metadata element set to describe those materials. The hope was that this would help define the scope of a potential Learning Materials Application Profile. There was no intention to produce an application profile as part of this work, nor was the work limited to any particular metadata schema.

Our starting point was that "learning materials" were likely to be of many different resource types (images, videos, simulations, etc.), and that each resource type would have its own metadata requirements and conventions. Furthermore, metadata would be required to support a range of activities (resource discovery, collection management, supporting accessibility, etc.), again each activity would have its own metadata requirements and conventions. Representatives from expert groups and communities relevant to each metadata specialism were interviewed in order to ascertain the range of advice available. These interviews form the basis of the synthesis of metadata requirements given in this report. In order to analyse these requirements two domain models were investigated. One model concerns the functional units of a repository, the other model concerns the entities being described by the metadata and the relationships between them. Using these it is possible to see how the metadata requirements for various resource types and activities relate to each other, to the internal and external interactions of a repository, and to the resources being described.

The synthesis of advice regarding metadata for learning materials highlighted that some aspects of applied metadata are more well developed and better understood than others. Not surprisingly, metadata to aid discovery of scholarly works is well developed; unfortunately it seems to be those areas such as metadata describing educational use and accessibility, which are closest to the defining core of an application profile for learning materials, that are least well developed. The synthesis work also suggests that there may be areas of overlap between different applications of metadata that are worth exploring for solutions to common problems. For example technical metadata is an issue in preservation metadata, the solution of which might benefit resource discovery in general. More generally, the approaches taken to solving recalcitrant problems in resource description may be transferable between fields, for example the approach of providing a link to semi-structured free text descriptive information rather than codified metadata may work for several problems. The synthesis also served to bring out the many uses of metadata (and resource description in general) beyond resource discovery and to highlight that approaches based on social tagging and Google are not replacements for metadata.

The analysis helped to tease out what exactly we are wanting to describe, which in a large part are not characteristics of the learning material itself but rather of the users and

their context. Unfortunately it is not clear precisely what these characteristics are. The specific resource model that was examined (that is FRBR) has some promising features, especially if it can be implemented across a range of resource types. However this is a big "if", and the level of complexity introduced by the FRBR model will need to be handled carefully if it is adopted as the basis for an application profile. The conclusion is that metadata descriptions for learning materials would have many uses but there exist a number of problems that need to be addressed before any application profile for learning materials could be created.

The report's sections follow the aims of the project: section 2 provides an introduction to why and how the work was carried out; section 3 is a synthesis summarizing and collating information gathered from the various informants; section 4 analyses this information in terms of the models for repository functions and resources; section 5 provides a discussion of the implications of our findings for future work by JISC on an application profile for learning materials. Some supporting information and examples are in appendices to this report.

The target audience for the report as a whole consists of JISC programme managers who are responsible for, or who have an interest in the management, retrieval and use of learning materials, or interoperability between repositories that contain learning materials in order to facilitate their management, retrieval and use. This audience will be most interested in sections 4 (Analysis) and 5 (Discussion).

Others who may be interested in sections of the report are:—

- Services and individuals involved in providing specialist advice on metadata requirements for leaning materials, who will probably want to see how their area of speciality is treated in section 3 (Synthesis), and the summary in Appendix 1, and consider how their advice relates to that of specialists in other areas.
- Some repository managers may find this report of interest (although they should bear in mind that the purpose of this report is not to provide them with advice—it is to provide advice to others about a particular approach to metadata) especially the model of repository activities in section 4.2 and the approach taken in Appendix 1 to showing which metadata elements support which resource types and activities.
- People involved in Standards bodies who are developing metadata standards for learning materials will hopefully be interested in some sections of this report (e.g. section 4, the analysis).

2. Introduction

2.1. Rationale and scope

This work stems from a request in October 2006 from JISC for JISC CETIS to investigate the creation of a learning materials application profile to parallel the Scholarly Works Application Profile [SWAP] which had at that time just been completed. Other application profile studies were also commissioned in areas such as images and time dependant media, all linked to the JISC's support for the use of institutional repositories for managing resources (Bruce, 2008). The work was carried out between Oct 2007 and April 2008.

The agreed aims and objectives were for a scoping study looking at the application profile requirements for learning materials in relation to digital repositories. This study would synthesize and analyse advice that is currently available to managers of repositories containing educational materials who need to define a metadata element set to describe those materials. This will provide the managers of such repositories with an appreciation of the various domain issues they may need to consider and the range of metadata that they might be expected to accommodate. In addition, the study will inform advisory services from different domains about the range of competing demands on repository resources that may be made by others. It is not within the scope of this work to produce an application profile.

It seems useful to start off by defining what we mean by the terms “metadata” and “application profile”, and to give an overview of some existing work and recent initiatives in the area of metadata for education resources. We won't define what we mean by "learning materials", especially not in distinction to similar terms such as "learning object" or "educational resource", but we think that the phrase "anything used for teaching or learning" captures the essence of what we are interested in.

We define metadata as "structured information that describes, explains, locates, or otherwise makes it easier to retrieve, use, or manage an information resource" (NISO, 2004). The definition of an application profile is somewhat more difficult. Initially the concept was used to describe how metadata elements from different specifications could be combined to produce a description schema optimised for a particular application (see for example Heery and Patel, 2000). However more recently there has been a recognition, summarized by Nilsson (2008), that “mixing and matching” elements from metadata specifications with different underlying abstract models leads to semantic confusion. An alternative approach which avoids this confusion is to focus on the application of a single metadata specification to a specific domain for which functional requirements are known and for which the entities being described and their relationships to each other can be modelled (Nilsson *et al*, 2008). A better definition of an application profile is, therefore:

> "a specification of how a metadata standard is deployed to meet the requirements of some domain or community, typically including

> specification of how the structure of the metadata description is constrained, i.e. the set of metadata terms referenced and how they are to be used in combination"[1].

For this work we have not made any assumption about what standard the application profile might be based on. To understand why and to understand the general approach taken for this report it is necessary to understand a little about the current state of development of the two main metadata schema that might be profiled to use with learning materials.

The IEEE Standard for Learning Object Metadata (LOM) provides a conceptual data model for the description of many characteristics of resources that can be used in learning, education and training. It is based on work from around the turn of the century, and at the time of writing is in the process of being reaffirmed by the IEEE Standards Association so that it will remain in its present form until at least 2013. This standard has already formed the basis of a draft application profile for learning materials in UK educational contexts, namely the UK LOM Core, work on which came to a halt at the end of 2004. Reaffirmation of the LOM standard notwithstanding, there are questions about whether the LOM is the best basis for an application profile. On the one hand repository managers who have been using the LOM and the UK LOM Core for several years are now in a position to re-evaluate whether it provides the right metadata to support their activities; and, on the other hand, the underlying model of the LOM is being re-examined in the light of what has been learnt and what has changed in the last ten or so years about metadata, resource description and interoperability of information on the world wide web (see, for example, the record of a workshop held in Leuven on Metadata 2.0 (Duval, 2008) and ongoing discussions of the IEEE LOM Working Group).

The Dublin Core Metadata Initiative (DCMI) develops metadata standards for the description of a broad range of resource types for a diverse variety of purposes. They are best known for the fifteen element "simple" Dublin Core Element Set [DCES], but more significant for this scoping study is the Dublin Core Abstract Model [DCAM] and other related work, e.g. the Draft Guidelines for Dublin Core Application Profiles [DCAM], creating an extensible and refineable framework for resource description metadata. Within the DCMI there is an ongoing effort to apply this to learning materials through the work of the DCMI Education Community [DCEd], two aspects of which are significant. First there is the work to create an application profile which focuses entirely on describing the educational characteristics of a resource. The aim is that terms for describing these educational characteristics can be used in combination with those that describe other characteristics (e.g. the title, subject, date of creation, accessibility requirements etc.) of the resource to create application profiles that adequately describe educational resources. The other relevant activity associated with the DCMI Education Community and the IEEE LOM Working Group is the Joint DCMI/IEEE LTSC Task

[1] Many thanks to Pete Johnston for providing this definition.

Force, which is aiming to create recommended best practice for the representation of LOM elements and instances in the Dublin Core Abstract Model.

At the time of commissioning this work the LOM, and particularly the UK LOM Core, was clearly the more widely used metadata standard for learning object repositories in UK Further and Higher Education. However, there is a widespread opinion that the LOM and its use should be reviewed, there are promising developments in the model used by Dublin Core and its application to learning materials, and Dublin Core is used for other related application profiles supporting The JISC's repositories work. For these reasons it was decided that this study should be neutral with respect to which metadata standard would form the basis of any application profile for learning materials.

Another initiative related to metadata for learning materials is the Metadata for Learning Resources (ISO MLR) work being undertaken under the auspices of the International Organization for Standardization (ISO/IEC JTC1 SC36). While development of ISO MLR initially followed an approach derived from the LOM, the standard is still under discussion, with the potential of it being based on an information model that has a higher degree of compatibility with Dublin Core. It is not yet clear when this work will be complete, or what impact it will have when it is complete.

As well as metadata encoded in formally defined schemas, there is also a wide spectrum of less formal metadata and resource descriptions. For example there is social bookmarking and tagging: the association of keywords of the user's own choosing with a resource and the sharing of these keywords with the aims of aiding resource discovery and access (by the user in the future or by others). User tagging can be accommodated within metadata if the inherent structure of the relationship between the tag string, the resource, the tagger, etc. is explicated and encoded (see, for example, Newman, 2005 and the Moat Project, no date). More generally users may create their own descriptions of a resource or how they have used it, for example as blog posts or more formal publications. While these are useful for resource discovery, they are probably better seen as distinct but related information resources in their own right rather than part of structured metadata pertaining to the described resource. There is clearly a potential role for metadata in sharing knowledge about the existence of such related resources.

We have tried to avoid any assumptions about how learning materials are stored or managed. Many permutations are possible: institutional repositories, national repositories, subject specific repositories, everything in one repository or separate repositories for each resource type: images, videos, scholarly works etc. The use of an application profile implies some thought is being given to the management of the resource descriptions this does not imply the use of a repository at all: the learning materials might be stored in a VLE (as is common) and the metadata held elsewhere. This report will only be relevant in situations where resource descriptions are being managed in some way, however we do not intend to advocate any one approach to resource management for learning material over any other. The relative merits of different approaches will depend on local requirements.

2.2. Method

Our approach is founded on the premise that a typical collection of learning materials is likely to include many different types of resources, for example text documents, images, assessment items, simulations. Since different characteristics are important for different resource types, each resource type will have its own resource description requirements. Furthermore the services provided by the repository, or activities undertaken that relate to the resources (e.g. preservation, ensuring accessibility, rights management, facilitating discovery and selection) will each result in additional metadata requirements. Expert and advisory groups for most object types or activity domains have developed their own specialized metadata and have their own perception of the minimum effort required for best practice. For this study we have created a list of the resource types and activities relevant to learning materials in an attempt to scope out what the complete set of requirements for resource description might look like. An initial list was created as part of the JISC CETIS proposal for this work, this initial list was posted to the CETIS Metadata and Repository Special Interest Group [MDRSIG] and the DCMI Education Community [DCEd]) for comment. The list was refined in the light of these comments and the final version is given in the Synthesis section of this report.

For each resource type or activity we identified one or more representative informant, for example from a JISC Advisory service or other expert group, and in an interview (normally by telephone, but occasionally face to face) asked their opinion on what metadata was required for their domain of expertise. The lead question for these interviews was:

> "what would be your recommendations to the manager of a digital repository of learning materials regarding metadata requirements for the adequate description of characteristics relevant to [the domain of the expert]".

That is a rather stark question, and one which in many cases is unanswerable without specific information about the repository in question—since the situation was hypothetical that specific information was not available. The question did however act as a prompt that was successful in soliciting the information required for this scoping study, either in general terms or through reference to some specific relevant work. This information is summarized in the Synthesis section of this report.

In order to help analyse the information gathered we have proposed two models: a functional model for the repository setting in which the metadata is held and an entity-relationship model for the resources being described. As with the list of domains these were first prepared as drafts which were posted for comment and discussion and then refined in the light of the comments. The refined models are presented in the Analysis section of this report.

This approach, with its emphasis on the totality of what may be described, contrasts with that of the DCMI Education Community with their application profile module focussing solely on describing those characteristics of resources that are relevant to education. This is deliberate, not because of any sense that the DC Education approach is wrong,

but rather to provide a complementary view, avoiding duplication of effort, showing the sort of resource description into which the DC Education module might plug. Throughout this project there has been continued communication with, and assistance from the DCMI Education Community.

2.3. Audience for this report

The primary audience for this report are those programme managers in JISC who are responsible for, or who have an interest in the management, retrieval and use of learning materials and interoperability between repositories which contain learning materials in order to facilitate their management, retrieval and use. We aim to provide this audience with the information necessary to inform their decisions on any future development of a metadata application profile for learning materials and on approaches to describing, locating, explaining and facilitating the use of learning materials in general.

A secondary audience for this report are advisory services and individuals who provide specialist advice to repository managers in UK Further and Higher Education on metadata requirements for leaning materials. We hope that this report will provide them with an overview of the general picture to which their advice contributes.

Some repository managers may also find this report of interest, however they should remember that it is a scoping study about an application profile developed in very generic terms, and it is not intended to be the direct basis of an application profile which they could implement. It is hoped that the relevant advisory services will be able to provide them with advice that is informed by this report, but which is more directly tailored to the needs of their specific situation.

Hopefully, this report will also be of interest to participants in Standards bodies who are developing metadata standards for learning materials, informing future versions of the standards mentioned above.

3. Synthesis

This section contains first a list of the domains which may drive the metadata requirements for learning materials and then a summary of the information scoping those requirements.

3.1. Range of metadata relevant to learning materials

The following is the list of domains of expertise relevant to learning materials for which information about metadata requirements was solicited.

Metadata for scholarly communication

Metadata specific to the citation of scholarly works, e.g. conference proceedings and journal publications, to facilitate the use of these resources in teaching and learning. Examples include the Scholarly Works Application Profile, PRISM and OpenURL (for citations journal citations).

Image metadata

Metadata specific to still images, for example photographs, line drawings, vector and bit-mapped graphics. Example metadata schema include the JISC sponsored images application profile and Z39.87/MIX.

Moving image metadata

Metadata specific to moving images, video and animations. Examples include JISC sponsored application profile on time based media.

Geo-spatial metadata

Metadata relating to geographic location e.g. GIS metadata.

Complex object metadata

Metadata specific to the description and management of complex or compound objects which comprise several resources of different types. Examples include elements of IMS Content Packaging [IMSCP], METS, DID, OAI-ORE.

Technical metadata

Metadata describing the technical features and requirements of a resource. Much of this is covered under metadata requirements of the specific resource types or activities (e.g. the pixel information of bit-mapped images) however it is useful to collate this under one heading and consider whether there are any gaps.

Personal / organizational data

Information about people and organizations relevant to their role in the creation and use of educational resources, for example as the author of a resource. Example schema include vCard and FOAF and elements of many other schema.

Metadata for general resource discovery

Covers general descriptive metadata applicable to any resource, e.g. title, description, subject. Typical formal metadata schemas are Dublin Core and MARC; folksonomy or tagging approaches are also examples.

Collection management metadata

Administrative metadata required for managing the collection of objects in a repository, for example reviewing the suitability of the content for dissemination.

Curation / preservation metadata

Metadata that is used to assist in ensuring that digital resources remain available and readable over a period of time. Examples include METS and PREMIS.

Rights metadata

Information about the terms and conditions under which a resource may be used. Typified by ODRL, XrML.

Accessibility metadata

Metadata that can be used to describe a resource's accessibility and its ability to match a learner's preferences. Examples include IMS AccessForAll Meta-data [IMSAcMD]and ISO Access For All Digital Resource Description [ISOAcMD].

"Usage metadata" and ratings

Information directly or implicitly relating to the value of a resource to end users, which can be used for recommendation systems. This may be as simple as the number of times a resource has been downloaded, the number of links to a webpage; also includes reviews and ratings of the resource.

Assessment metadata

Metadata specific to the description of assessment items, questions and tests, for example the LOM application profile and other metadata in IMS QTI 2.x

Education metadata

Metadata to aid resource discovery and selection for educational use, e.g. description of intended learning outcome and level of difficulty, descriptions of the educational context for which a resource was designed or in which a resource has been used, and, in the case of resources such as learning designs, the activities and pedagogical approach described by the resource. Typical examples are IEEE LOM and course description schema.

3.2. Metadata for scholarly communication

> "Metadata specific to the citation of scholarly works, e.g. conference proceedings and journal publications, to facilitate the use of these resources in teaching and learning. Examples include the Scholarly Works Application Profile, PRISM and OpenURL (for citations journal citations)."

Scholarly works have been the subject of their own application profile [SWAP] funded by JISC. We would expect this to form the basis for any description (in UK HE) of scholarly works that are part of a collection of learning materials. The reasoning for this is that, firstly, the functional requirements that SWAP is designed to address are as applicable to the use of the scholarly work for educational purposes as they are to its use for research purposes. Secondly, we envisage that SWAP will be implemented in many UK Further and Higher Education Institutions for repositories of research output and it is sensible to maintain compatibility with this.

Only a fraction of the scholarly works used in teaching and learning at an institution will be available for deposit in that institution's repository for learning materials: most will be written by authors at other institutions (though even then there may be circumstances when copies are stored). There may be a case for storing references to scholarly works held elsewhere, i.e. the data required for a standard bibliographic reference to material accessed online, along with information about how those works had been used in education. SWAP contains the elements necessary for such a reference. Alternatively these references will be within the text of the learning materials for which they are used, and the scholarly work itself will not feature in the repository.

The properties described by SWAP are listed below for reference, under headings which refer to the entity in the SWAP model to which they pertain (this model is similar to the model proposed in the Analysis section of this report).

Scholarly Work

title, subject, abstract, grant number, has adaptation, identifier (URI)

Expression

title, description, date available, status, version number or string, language, genre/type, copyright holder, has version, has translation, bibliographic citation, references, identifier (URI)

Manifestation

format, date modified, identifier (URI)

Copy

date available, access rights, licence, is part of, identifier/locator (URI)

A few of the properties listed above would not need to be described in the context of providing access to the scholarly work as a learning material. For example it is doubtful that the grant number of the work would be significant.

3.3. Image metadata

> "Metadata specific to still images, for example photographs, line drawings, vector and bit-mapped graphics. Example metadata schema include the image application profile and Z39.87/MIX."

Image metadata transpired to be an interesting case. Unless the image happens to be embedded in a document in close proximity to text about the image, discovery of images by search services is almost entirely reliant on metadata. Also any given image is often available in different formats: an original painting with a digital surrogate available as a thumbnail for browsing, a larger format for use in presentations or web pages, and perhaps a high resolution scan—each of these might come with different usage conditions. For these reasons images provide good examples of many more generic issues relating to metadata.

JISC are funding a specialist application profile for images, which will hopefully be based on requirements that are compatible with those of learning materials. Generally, the important metadata for images include descriptive metadata for the image and technical metadata about the image file. We can expect that the metadata for images as learning materials will need to encompass the following:

General descriptive information

A set of descriptive metadata for the subject matter of the image to support resource discovery and selection. This descriptive information could act as a caption or alternative long description of the image when it is used, this can aid accessibility for the visually impaired.

This metadata should also describe any aspect of the image that might provoke sensitivities, e.g. depiction of violence, sexually explicit content—the exact nature of what is sensitive will depend on the audience of repository.

Relationships

In order to aid the discovery of images of particular subjects, it is useful to provide information on how the image relates to other images, objects, places, and whether there is information about the related resource available. For digitized images, it may be important to separate out information that relates to the original image-object from that which relates to the digitized image. For example, in order to find suitable quality digital surrogates of a portrait of a particular person it would be necessary to describe the properties of the digitized image, those of the original portrait and those of the person in the portrait, perhaps as separate but related objects.

It may also be helpful to identify whether the image is included as part of a larger resource or collection.

One could also show relationships between different versions of an image, e.g. in order to identify a high resolution version suitable for a large print copy for use on a poster or for a visually impaired learner.

Identifiers are important in representing all these relationships.

Scale

For maps, scientific images, and other technical images, specification of scale or magnification can aid discovery and selection.

Rights and Restrictions

Users and repository managers need to understand what they can and cannot do with the images, e.g. include in PowerPoint presentations, web pages, handouts, distribute to other image collections, manipulate to disseminate at other resolutions. Any conditions of use should also be recorded, for example credit required, limitation on use, e.g. for internal or non-commercial uses only. For images of people, especially in the case of medical images, it is necessary to record whether the subject has given their consent for the use and distribution of the image. There may also be confidentiality and data protection issues.

Knowing the provenance of the image will be important, i.e. where it has come from, so that users and repository managers can go back to clarify rights issues.

As well as general purpose licence languages, e.g. ODRL, there are image specific rights metadata languages, e.g. PLUS (picture licensing universal system) which relates to images from commercial stock image services. Metadata may also be required to support images created under the CLA scanning licence, which enables image creation, and DACS (design arts and copyright society) licence in negotiation, which aims to allow greater freedom of educational use of images by an institution than the CLA licence does.

Technical metadata

Prior knowledge of basic technical metadata, e.g. file format, file size and pixel dimensions, can be useful in order to aid selection of an appropriate copy. Maintaining a record of this information can also assist the management of the collection, since one might need to restrict access to high resolution images for, say, IPR reasons.

If other information is available in the source image when deposited in the repository (it may have been automatically generated by the camera or scanner that generated the digital image) and if it can be easily extracted automatically it may be worth storing and selectively exposing it. Such information may, for example, help with resource management, preservation or quality control of the images, but it is not of direct importance to their use as learning materials. Example formats for this technical metadata are Z39.87/MIX and PREMIS.

3.4. Moving image metadata

> "Metadata specific to moving images, video and animations. Examples include the JISC sponsored application profile on time based media."

The issues around metadata for moving images are complex. Part of the reason for this is that audio/visual material can be considered to be a complex collection of parts in several different ways. They can, for example, be treated as a sequence of stills or frames; equally they can be considered as comprising a moving image stream with (potentially a choice of) soundtrack and possibly with other information such as subtitles, credits and other metadata. Several of these parts may be useful independently of the others, so a lecturer may wish to use a short clip from a longer video in their choice of file format, audio and subtitle language.

Publicly shared material like that found on YouTube will normally have metadata describing the whole piece, and not be specifically designed to support the discovery and use of segments within the video. This may be sufficient if the video is short enough to be treated as a single segment and available in only one form.

The metadata used in professional production and content management systems takes into account the complex nature of video, giving a time dependant account of what happens in the video. Technical and rights issues will determine whether these can be accessed either by cutting that clip from the rest of the video or by sending a request which includes the desired start and end time to a streaming media server. However the metadata available in professional video tends to be highly specific to the type of content, e.g. football matches will have metadata identifying when goals are scored.

Both self-generated and professional video are likely to have some embedded technical metadata which has been automatically generated during shooting and editing the video. While this may not be of direct importance to the use of the video as a learning material there is an argument for keeping at least some of it, e.g. version information, which may be useful downstream when using the content, especially if further editing is required.

Most existing metadata is aimed at and suitable for professional broadcast producers and may not meet the needs of typical education users (excepting media studies teachers and learners). In many cases the requirements of typical education users would be similar to those of other resources, e.g. information on the subject of the video, versioning and relationships, licensing, but with the complication that this may be time dependant.

3.5. Geo-spatial metadata

> "Metadata relating to geographic location e.g. GIS metadata."

Encoding information about geographical location can be useful in resource discovery when the subject or origin of a resource is a place. The data may be generated automatically (some cameras have this facility) or it may be added to a resource in the repository (sometimes called geotagging or geocoding). When combined with mapping software (e.g. Google Earth) geo-spatial metadata can form the basis of innovative resource discovery approaches.

Geo-spatial metadata comes in two forms: coordinate data or gazetteers. Coordinate data specifies the location of a point or boundaries of a box with reference to a grid. Gazetteers are essentially controlled vocabularies of place names or codes (e.g. postcodes) and are more appropriate for specifying locations in terms of extended, irregular areas e.g. Scotland. Crosswalks exist to convert between different coordinate systems (e.g. latitude/longitude to UK National Grid) and between coordinate and gazetteer approaches.

3.6. Complex object metadata

> "Metadata specific to the description and management of complex or compound objects which comprise several resources of different types. Examples include elements of IMS Content Packaging, METS, DID, OAI-ORE."

Many learning materials are complex objects comprising several individual files, for example a tutorial package with pages of text and animations or an assessment with several questions, some of which may in turn comprise objects such as images or video. There are two different approaches being taken to treating complex objects: the packaging approach and the metadata relationship approach. The packaging approach allows complex objects to be stored in repositories as single entities, e.g. IMS Content Packages. However the parts of a complex object may be useful in their own right and there is value in being able to discover them separately while still knowing that they are part of a potentially more useful aggregation. The metadata relationship approach is supported by the IEEE LOM and DC metadata with hasPart and isPartOf relation types that can be applied to show the relationship between simple assets and a complex object that incorporates them.

It is important that a suitable object model and an identifier framework exist to support the representation of simple assets, compound objects and the relationships between them.

3.7. Technical metadata

> "Metadata describing the technical features and requirements of a resource. Much of this is covered under metadata requirements of the specific resource types or activities (e.g. the pixel information of bit-mapped images) however it is useful to collate this under one heading and consider whether there are any gaps."

Technical metadata supports resource discovery and selection by allowing the user to find and choose resources compatible with the software they use to display and edit them. It also faciliates the selection of copies of a resource suitable for a specific use; see the discussion of technical metadata for images, above. As well as a general requirement for resource discovery, technical metadata was identified as particularly important for ensuring accessibility and preservation.

One approach to technical metadata is to describe characteristics of the resource rather specify the software required to run it. At a minimum the MIME type of the resource and file size (or bit rate) can be specified; for some resource types specialized schema are available to provide more detail. However this information is not always sufficient. A simple MIME type may not be enough to specify a format precisely (for example IMS Content Packages may have a MIME type of application/zip but not all software that can open a ZIP file will understand the contents of the package; similarly application/msword covers many versions MS Word). Even where more detailed technical information is available there needs to be a service that maps from this to the capabilities of various software that may be available (see for example the DCC representation information registry repository at http://registry.dcc.ac.uk/omar/.

The alternative approach is to provide a machine-processable list of compatible software in the metadata for each resource. This is the approach taken by IEEE LOM and it directly meets the requirement outlined above for resource discovery and selection. This approach works for specifying a few simple requirements relating to platform (e.g. resource for a specific operating system), however if taken any further the result is complex, very inefficient and difficult to maintain. A simpler method that is frequently adopted is to provide human-oriented free text information to provide an overview of the requirements and assist selection and use of materials.

In the absence of a usable approach that fulfils the requirements to describe the technical characteristics of a resource sufficiently to allow automatic matching of resource requirements to the user's available software capabilities, a combination of minimal structured metadata (MIME type and file size) supplemented by human readable text is enough to provide basic filtering of appropriate or desired resources and give users some advance "clue" about whether the resources they find will be useful to them.

3.8. Personal / organizational data

> "Information about people and organizations relevant to their role in the creation and use of educational resources, for example as the author of a resource. Example schema include vCard and FOAF and elements of many other schema."

Information about the authors, creators, publishers and users of a resource and its metadata can be useful in identifying and selecting learning materials (e.g. finding resources from a trusted source, or that have been used in a similar institution to one's own). The location of authorship and publication can be useful in assessing the cultural appropriateness of a resource. Information about rights holders can be helpful when using a resource (for example, to query or negotiate what is allowed under the); information about all the above can be useful in managing a repository.

The requirements are similar to those of scholarly works, so the information required to identify agents in the scholarly works application profile may be sufficient. This is:

Type (person or organization);

Name;

Family name;

Given name;

Workplace homepage;

Mailbox;

Homepage.

It may additionally be desirable to provide some information about the institution(s) that a person works in. Allowing the user to judge whether a resource was developed at, or a comment originates from an institution with a similar profile and mission to their own gives contextual information that can help with resource selection.

3.9. General resource discovery

> "Covers general descriptive metadata applicable to any resource, e.g. title, description, subject. Typical formal metadata schemas are Dublin Core and MARC; folksonomy or tagging approaches are also examples."

The precise information required to support resource discovery will depend on the type of material, what users want it for, the users' search strategies, the search facilities available (e.g. is a search over metadata the only search available or is full-text searching also available), and the information environment within which the repository sits (e.g. what resource discovery services does the repository interoperate with). It is important to have both a standards-compliant metadata schema and agreed cataloguing conventions, but the latter are largely out of scope for this study. Typically the following metadata elements will be important:

Title

Usually required for the display of search results to aid resource recognition and selection. Even where there is no widely known title for a resource (which is more often the case with some types of learning materials, e.g. computer simulations, that with some other types of resource) a title can give an immediate indication of what a resource is about, e.g. "How to truss a chicken". Titles may also support simple searches since special weight can be given in returning search results where the search terms occur in the title.

Keywords (and source)

As well as supporting simple searches, keywords are useful for onward browsing, e.g. the "find similar resources" type of functionality. While free keywords can be useful in supplementing what is provided in other descriptive fields, it is useful if keywords come from controlled vocabularies such as thesauri in which case additional functionality (e.g. offering to find similar resources based on broader or narrower terms) can be provided. For controlled

keywords the source of the term should be recorded. User provided keywords are dealt with under tagging, below.

Description

A description supports simple search since it elaborates on the keywords. Where no immediate preview of the resource can be shown a description may be important in helping resource selection. Consistency is less important here than in keyword and classification fields, so authors and users may provide useful content for this field.

Classification by subject (and source)

Classification by subject supports resource discovery by search and browse in ways similar to keywords. It can also be important in filtering metadata records when a repository that covers a range of subjects provides data for a subject-specific resource discovery service. The classification scheme used should be identified in the metadata.

Date of resource creation

The date of resource creation is probably the most important date for users who might want to select up-to-date resources or narrow their search to a relevant time period.

Resource type

It can be useful to separate resource type from the technical format and medium of a resource, e.g. is the resource a diagram, a tutorial or an assessment. For example a lecturer might want a diagram to illustrate a specific point; a learner might need an online tutorial to meet their learning needs. Users will express these needs in natural language which is where the distinction between type and format may become lost.

Provenance: author or publisher and institution

Many users will want to consider reputation and context when selecting resources, for example choosing resources that were created by a well-known expert or come from an institution with a similar profile to their own. It may also be useful to display information about other resources from the same source (e.g. subject classification, keywords, user tags) as a means of helping users form their own opinion about the authority of a resource creator or publisher.

Relationships between resources

Resource selection and discovery of related resources may be assisted if related resources are linked, for example different editions of a book or different parts of a course.

Warnings about sensitive content

> A point made specifically for images is probably general enough to repeat here. The metadata may include mention of any aspect of the resource that might provoke sensitivities, e.g. descriptions of violence, sexually explicit content—the exact nature of what is sensitive will depend on the audience of the repository.

3.9.1. User generated metadata and social tagging

While discovery metadata is frequently created by resource authors and cataloguers, there is a growing recognition that user created metadata can aid resource discovery. A good example of this is delicious.com, which allows users to tag resources with simple keywords and to add their own description; both keyword and description are stored on the delicious.com website, not the resource website. In addition to the functionality supported by conventional keywords and descriptions, tagging has a social function, making it possible for users to identify others with similar interests and to discover resources they have tagged. Whether this type of activity is best supported at the repository or at an independent site will probably depend on the nature of the repository, the resources it holds and the user group.

To support tagging on independent sites all that is required of a repository is that it provides a public URL for each resource. Getting the information back from the sites where resources are tagged is, however, more problematic. Tagging within the repository is probably best supported with a specific field, separate from the keyword field; additional information about who tagged the resource (the user IDs) would be required to support functionality such as displaying resources tagged by other people who had tagged a resource the user is looking at. Further information on the relationships between users (e.g. user X is interested in resources tagged by / recommends resources to user Y) will allow users to garner and pass on recommendations more directly. There are then legal and technical issues about how much of this information should be shared with other sites and how.

3.10. Collection management metadata

> "Administrative metadata required for managing the collection of objects in a repository, for example reviewing the suitability of the content for dissemination."

Much of the metadata that is helpful for collection management is present in the metadata required for other activities such as resource discovery. There is, however, some information about the resource and about the relationship of the resource to the collection that needs to be recorded specifically to aid collection management. Additionally there is information that may be recorded to support other activities that is also important for specific reasons relating to collection management.

Dates relating to the resource in the collection

In order to make sure that the resource descriptions are current it is necessary to record the date when the metadata describing the resource was first created, when it was updated, and when it was last checked.

Metadata creator and validator

A record of who created and validated the resource description that is sufficient to allow them to be contacted in case of a query about the metadata may be necessary.

Record owner

Information about the owner of the metadata might be required in order to facilitate rights management relating to syndicating metadata.

Resource access

Information about how often a resource has been accessed by users can be useful in identifying which resources are no longer of sufficient interest to be included in the collection.

Comments field

Notes on the history of a resource, its reliability, special reasons for its inclusion in the collection etc. can help with collection management decisions.

Version

Information about the version of a resource is necessary where version control is important.

Resource owner

Information about the owner of a resource is necessary in order to allow queries about rights issues relating to its inclusion in the collection.

Relationships between resources

Information about, for example, the different manifestations of the same resource, different parts of a serial resource and different parts of a compound object, is useful because if action is required for one part of such a resource it might indicate that other parts also need attention.

Classification (subject, resource type etc.)

Classifications of the resource by themes such as subject and resource type may also be important in ensuring that the coverage of the collection matches the interests of the repository's users.

3.11. Curation / preservation metadata

> "Metadata that is used to assist in ensuring that digital resources remain available and readable over a period of time. Examples include METS and PREMIS."

The area of curation and preservation of learning materials is one where there is very little existing practice and so, at present the requirements for preservation of learning materials should be outside the scope of this work. (It is however being addressed elsewhere[2]). However, the general requirement for curation and preservation, that a reasoned decision be made on what approach should be taken the long term curation of resources, stands. The decision might be that the materials are ephemeral and long term curation is not necessary. If that is not the case, and some form of preservation activity is required, then the preferred metadata approach is based on PREMIS and the OAIS model for information packages in archival repositories.

In general there seem to be broad similarities between the metadata requirements and challenges for preservation and for sharing resources. For example, OAIS includes representation information, that is the information required to make the object in archive intelligible to the users of that archive. This representation information will include metadata about the technical requirements sufficient to display the material. So there is possibly a link between the requirements of preservation metadata and those of technical and accessibility metadata; requirements that are challenging to meet. Similarly there are preservation requirements for provenance and rights metadata. Further investigation of these similarities would be desirable in the hope they are close enough to allow for joint solutions to be developed, and that ultimately any extra effort required at the repository in order to address preservation issues would be useful in fulfilling other requirements.

3.12. Rights metadata

> "Information about the terms and conditions under which are resource may be used. Typified by ODRL, XrML."

Metadata relating to digital rights is necessary in order to inform users what they are and are not allowed to do with a resource that they obtain from a repository and under what conditions (e.g. whether payment is required). Rights metadata includes licence terms, information about the rights holder, and in some cases usage monitoring data.

The licence terms and conditions under which a resource is offered can be provided in machine readable form using a number specifications. These typically encode in XML a description of what someone is and is not allowed to do with a resource under specified conditions. The number of elements involved in these specifications can be large, however, if many resources in a repository are available under the same conditions then

---

[2] The JISC has issued an ITT for e-Learning materials preservation and curation studies, see http://www.jisc.ac.uk/Home/fundingopportunities/funding_calls/2008/01/elearningmaterials.aspx The attitudinal study is being carried out by Evidence Base, see http://www.ebase.bcu.ac.uk/projects.htm

there is no need to edit this metadata for each resource, a single record template can be applied to all resources. Alternatively some form of assistance can aid the depositor or repository manager in selecting from a choice of licences or selecting a set of terms and conditions in order to create a licence. The XML can either be embedded in the metadata record, for example as an extension to the LOM Rights category, or linked to from the resource description metadata. One aspect of rights metadata that may require further attention is vocabularies describing acceptable use in education-specific contexts, for example freely available for UK-HE only.

Recording the rights holder of a resource allows users of the resource and potential users of the resource and managers of the repository to contact the rights holder in order to check the licence conditions or to negotiate terms for use not covered in the licence.

In some cases it is necessary to record how a resource has been used. For example the CLA requires a high level of usage reporting including details of the duration of the specific courses that the resource has been used for and the number of students who have taken that course. This information can be provided using an extended profile of the XCRI (eXchanging course-related information) specification.

3.13. Accessibility metadata

> "Metadata that can be used to describe a resource's accessibility and its ability to match a learner's preferences. Examples include IMS AccessForAll Metadata and ISO Access For All Digital Resource Description."

There are two distinct approaches to ensuring that repositories support accessibility. The first is to describe the modalities of resources at an asset or individual file level, for example stating whether the asset has auditory, visual or textual content. Accessibility can then be enhanced if the repository can provide information about alternatives in different modalities, for example a transcript of the audio. This is the approach taken by existing metadata specifications and standards, for example the ISO Access for All standards. The ISO standard also provides a means for describing learner preferences which can be matched to resource descriptions to identify appropriate materials for each individual. At present, uptake of accessibility metadata is patchy.

The second approach is to focus on the intended learning outcome rather than the accessibility of the object *per se,* and to make appropriate adjustments to the teaching rather than try to make a technical substitution. The context of the learner is crucial here, since different adjustments would be necessary for, say, master-level distance learners compared to first year undergraduates on-campus even for the same learning material. For example, in some contexts pairing students might be appropriate, in others the best adjustment might be for the student not to spend time on this part of the course. There needs to be information available to support the choice of adjustment, perhaps based on sharing previous experience of using the same resource or resources of similar type in similar learner contexts. One possible approach is that this information, along with information about the modalities used by parts of the resource, could be held as a free-text record (see TechDis, no date). Access to this free text record could be open for users

to add their own supplementary information about, for example, what assistive technology or other measures they used to ensure accessibility of the resource in specific circumstances. The details of how this might be achieved in practice, however, is currently not well understood, nor is it clear whether this approach could be used to select resources automatically given knowledge of a learner's preferences.

3.14. "Usage metadata" and ratings

> "Information directly or implicitly relating to the value of a resource to end users, which can be used for recommendation systems. This may be as simple as the number of times a resource has been downloaded, the number of links to a webpage; also includes reviews and ratings of the resource."

Usage metadata has been extensively exploited in some commercial systems that have similarities to repositories, a notable example being Amazon recommendations. Typical usage metadata includes the number of times a resource has been downloaded (or bought), ratings, short reviews, inclusion of the resource in lists of favourites: all of which support the capability of the system to point people to other resources that were liked by other users with similar tastes to their own. A simplistic application of this to learning materials might be problematic: the teaching or learning context is critical in determining how valuable a resource is; the critical mass of information (about resources and users) required before any useful inferences can be drawn might not be available from the usage of a single repository; academics from disciplines where there is a culture of providing in-depth book reviews as part of the scholarly output might feel that providing short off-the-cuff reviews and ratings could damage their reputation. Pooling of usage metadata between repositories might help get the critical mass of information required. Reviews of learning materials and how they have been used might require longer reports which could capture richer, deeper, more nuanced information than can be accommodated in a simple metadata comment. If these reports exist as independent information resources there is however the problem of how to maintain links to them from the resource description in a host repository.

An interesting example of information that could be a part of usage metadata is provided by "track back" or "pings" in blog postings [TRACKBACK, PINGBACK]. These allow a sort of reverse link from a blog posting to others who are referencing that post. A similar system might help repositories keep track of reviews of materials they host. Furthermore, one can imagine the utility of a system whereby a learning environment sends information to a repository about how it is using resources sourced from that repository. This information can be used at the repository to find out which resources have been used in which courses, for example, when appraising which resources are most (or least) useful, when a new lecturer wants to know what their predecessor used, and in avoiding inadvertently giving students the same material to cover in two courses. In blogs the back link typically includes a snippet of text from around where the reference was made; for learning materials the back link could include information about the course on which the resource was used (c.f. the example of using XCRI to specify

which course copied materials have been used in section 3.12 Digital Rights Metadata above).

3.15. Assessment metadata

> "Metadata specific to the description of assessment items, questions and tests, for example the LOM application profile and other metadata in IMS QTI 2.x"

To a large extent the characteristics one would want to describe for both assessment items (i.e. the information required for individual questions) and tests are the same as for any other educational resources. This is why the IMS specification for question and test interoperability uses profiles of the IEEE LOM for metadata. There is of course a need for specialist vocabularies to express assessment-specific terms for characteristics such as learning resource type, and relation type. There are also some properties where assessment has exceptional requirements, these are:

Title

Tends not to be relevant to assessment items.

Version and status

Highly important in some scenarios, for example knowing which questions are embargoed pending their use in high-stakes examinations.

Duration, typical learning time

Need to be careful about the interpretation of these in the context of time-limited exams

Difficulty, age range

Difficulty is handled by usage data; age range is often not relevant since ability is more important.

Interaction

Interaction has a different meaning in computer aided assessment than in other educational contexts, specifically it relates to how the question is answered, for example by text entry, choice or matching.

There is also a class of metadata that is entirely specific to assessment items:

Assessment usage data

Assessment usage data is context specific information about the performance of a given cohort on a specific question, for example how many students in the cohort got the question right, which wrong answers were provided, how good was the question at discriminating between able and less able students. Usage data provides information on the difficulty of a question and other measures that can be used to evaluate the appropriateness of the question to a given context as part of the quality assurance processes after a test is administered or

trialled. Subsequently this data can be used to support selection criteria for future use of the question. It is conceivable that the data could be shared between repositories holding the same question items, so long as the context of the usage were maintained—one would have to make suitable adjustments if usage data from high school physics students were to be used to select items for use with University Engineering students.

The IMS QTI specification provides a framework for creating a record describing various statistical measures based on assessment usage data; the minimal requirements for a repository wishing to utilise usage data is that there is a pointer in the metadata to this record, and that there is a description of the context in which this record was generated.

3.16. Education metadata

> "Metadata to aid resource discovery and selection for educational use, e.g. description of intended learning outcome and level of difficulty, descriptions of the educational context for which a resource was designed or in which a resource has been used, and, in the case of resources such as learning designs, the educational use that the resource describes. Typical examples are IEEE LOM, course description schema and pedagogic vocabularies."

Generalizing greatly, the overarching aim of the type of metadata considered in this section is to help people find, use and manage resources with reference to how they meet a general or specific learning, educational or training need. Unfortunately it is difficult either to articulate this aim more precisely or to specify the requirements for meeting it, though the DC Education working group [DCEd] is engaged on work that promises to do so with more rigour and in more detail than has hitherto been the case.

This seems to be an area where the expectation of what should be supported by metadata goes beyond resource discovery to include resource use. A related observation has been made in relation to learning designs (descriptions of educational activities): one view of learning designs is that they should offer a "script" that can be run by a VLE or learning management system (this is the intent of the IMS Learning Design specification and its machine-readable XML binding); another view is that they should provide a human readable description that teachers can use as inspiration for their own classes. Similarly for education metadata, there is a case that the description of the resource could be seen as providing inspiration for how and why one would use it in teaching and learning.

A complicating factor for education metadata is that many of the resource types and activities of interest are experimental or emergent, and consequently there is little agreement on how to support them. Learning design is one example of this, where there is debate over the nature and purpose of the resource. Another example is where central policies such as moves to support life-long learning, work-place learning, student

centred learning etc. may require changes in assumptions about the nature of the activities that the metadata is trying to support.

Some examples of the information that might be relevant to resource description in relation to educational scenarios are: the subject or concept being addressed, the educational level, the educational setting, and teaching and learning styles.

Subject or concept

To a large extent the requirement that one should be able to ascertain whether a resource is relevant to the subject being learnt or taught is met by the subject classification, which is also required for more generic resource description. However this may not be specific enough in some circumstances, for example where a resource is required to support learning towards some goal, objective or competency specified in a set curriculum. Of course, this brings in elements of educational level and discipline: the same concept may be addressed in different ways for first year undergraduate physics and for masters level chemistry. Also any mapping to a specific curriculum embodies an assumption that the educational setting in which the resource will be used is one where that curriculum is current.

Educational level

It seems intuitively clear that some form of indication of the level is helpful in selecting appropriate resources, hence the abundance of "*X* for beginners" or "Advanced *Y*", "Introduction to *Z*" and similar formulae for titles. However, defining the educational level of resources (as opposed to qualifications or stages of a curriculum) comprehensively in a transferable manner is difficult. A common approach seems to be to define (sometimes tacitly) the usefulness of a resource in relation to a qualification or curriculum stage.

There are clear relationships between educational level and other attributes which have been used to describe learning materials, such as the age of the intended audience, the educational setting (e.g. kindergarten or University), the difficulty of the resource, and in part, the topic itself (introductory quantum mechanics and introductory arithmetic are understood to relate to different educational levels). However, none of these relationships are always one-to-one, so for example people of (almost) any age, in any setting could be learning introductory French (but probably not using the same resources).

Educational setting

A common requirement of those looking for educational materials is that they wish to find something that is appropriate for their particular educational context or environment. Ideally they might want something that has worked for someone in similar circumstances to their own, be it a ten minute demonstration in a lecture or a complete module for studying independently. Examples of

differing settings are distance education versus face to face, learner or tutor directed, small group versus large group, formal or informal learning.

Time required to use the resource

The time required to use a resource is important: a tutor may be looking for something that can be used in a single class with limited time, or a learner may be looking for a complete course on some topic with the expectation that they will spend several hours studying over a period of time. The time required to use a resource is not always the same as the technical run time of a media file such as a video since it applies to resources such as simulations and texts that have no inherent run time and it may include time for other activities that are a necessary part of using the resource to learn. It is not generally possible, nor is it necessary, to provide this time with precision: it is normally sufficient to say that the resource will take less than about 10 minutes, or a single (1 hour) lesson, or an approximate number of hours to complete.

Teaching and learning style

Finding a match between the teaching approach and learning preferences is important in ensuring high quality education. Some educational materials are designed with a specific pedagogic approach in mind, some support certain approaches or make others difficult. For example, the microworlds type of resources pioneered by Seymour Papert's LOGO and "Mindstorms" (Papert, 1993) was designed to support constructivist (more specifically constructionist) teaching. Similarly learning materials based on concept or mind maps may favour visual learners. Finally some educational materials, for example learning designs, explicitly describe a teaching approach.

Teaching and learning style is related directly or indirectly to elements in current metadata schema for describing the resource's instructional method, interactivity type and level, and resource type.

The first two items above are of considerable importance for resource description relating to a specific learning goal. Where a set curriculum exists, relating learning materials to set points on that curriculum goes a long way to meeting these needs. Thus, for example COLEG, a charity that provides learning materials for Scotland's colleges, classifies learning materials against descriptors from the SQA qualifications catalogues (for example the Catalogue of National Qualifications, SQA, no date). Within an institution or department the same can be said about course syllabuses. This perhaps is one reason behind the practice of using VLEs as *de facto* repositories, since VLEs are structured entirely around course syllabuses and, where appropriate, these will relate to set curricula. A similar approach can be seen in many open courseware or open educational resource sites, where the material is available under browse headings that correspond to the originating institution's course offerings, see for example the MIT OpenCourseWare website (MIT, no date). However the draw back of mapping to curricula and syllabuses is that these tend to be local and to change over time. One

alternative approach which is being tested (Van Assche, 2007) is to map the curricula, syllabuses and resources to competences which would hopefully be more stable and of wider applicability.

A final indicator of the problems inherent in metadata for education is the problem of agreeing controlled vocabularies to describe educational characteristics. To pick two examples: first, even something as seemingly simple as producing a set of educational levels for all sectors throughout UK (MEG, 2003) proved to be deeply problematic; second, a report by Currier *at al* (2005, p.5) on pedagogic vocabularies found that the requirements of the community for such vocabularies "vary enormously and, to date, have not been clearly articulated in a coherent fashion". A conclusion that may be equally applied to education metadata in general: we believe that there are characteristics that it would be useful to describe, but it is very hard to specify precisely how this should be done.

4. Analysis

4.1. Introduction

Two models were developed to aid the analysis of the information collated above. One is focussed on the repository holding the learning materials and helps to put the creation, management and use of metadata into context. The second is focussed on the objects being described and their inter-relationships, and will be of more direct importance to any application profile development work.

4.2. A model for repository activities

In order to draw some distinctions between some of the types of metadata discussed above we have used a simplified version of the OAIS Reference Model [OAIS], with specific adaptations, to provide a picture of the activities that are centred on a repository. The use of OAIS in this way follows the evaluation finding of the JISC Digital Repositories Support Team (Allinson, 2006) that a lightweight version of the OAIS model, with less emphasis on preservation, would provide a reasonable basis for a generic model of repositories. The high level functional entities of the model are depicted in figure 4.1 below:

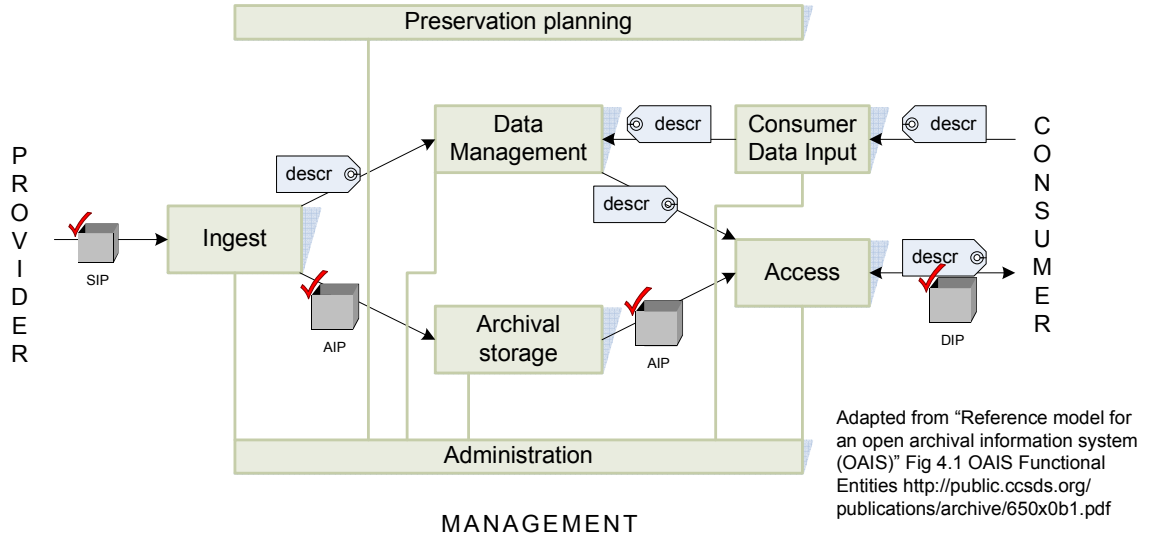

Figure 4.1: an adaptation of the "Reference model for an open archival information system (OAIS)" showing the communities, information packages (SIP, AIP, DIP, see below for a full explanation), functional entities (in boxes) and data flow relevant to a repository.

The repository system is shown in the centre interacting with three groups or communities: the providers of resources, the consumers or users of resources, and a group responsible for the management of the system. Resources in OAIS are modelled as three types of Information Package: the Submission Information Package (SIP) is the resource as received, the Archival Information Package (AIP) is the resource as stored,

and the Dissemination Information Package (DIP) is the resource as delivered to the consumer. An easy to understand example of how these may differ in format is the case of an image received in proprietary high resolution format, stored in an open high resolution format, and disseminated in lower resolution or using a lossy compression algorithm. The repository system itself comprises the following entities:

Ingest function

Involves receiving the resource from the provider and creating the archival information package, including extracting metadata from the submission package and/or creating metadata on submission.

Data management function

Responsible for managing descriptions of the stored resources, ensuring data integrity, performing dataset updates including schema updates, and handling queries and the generation of result sets.

Archival storage function

Responsible for long term storage and maintenance of the archived information packages and providing them for dissemination.

Access function

Supports the discovery to delivery process, allowing consumers to appraise the contents of the archive and managing access to copies of resources of interest to them.

Preservation planning function

Provides services and functions with the aim of ensuring that the resources in the archive remain available for as long as required even though the surrounding computing environment changes.

Administration function

Supports the management of the archive, including negotiation with providers and consumers, auditing, system maintenance.

Consumer data input function

This is an additional function not in the OAIS model, which has been included to highlight the concept that consumers may also be providers of information by, for example, reviewing and rating resources, providing contextual usage information, and commenting on and annotating resources or adding their own descriptions or tags. This does not include resubmission of re-purposed objects or contribution of user generated content, since in these cases the user is a content provider rather than consumer.

While the modelling and language of OAIS is formal, Allinson found that the model itself is applicable to formal and informal repositories at all levels, allowing that for

some informal or personal repositories the functions might be reduced to trivial tasks much simpler than is implied in description above. The model also shows a repository as a stand-alone system in direct contact with its provider and consumer communities. In a federated network of repositories and other services the situation is likely to be more complicated than this: for example, the repository might be a data provider for services that have diverse consumer communities. The consequences of this will depend on the nature and policies of the federated network.

4.2.1. Observations drawn from the model

The three information packages in the model are created through different means and serve different ends, not surprisingly they are likely to have different metadata. It is important to ensure that the metadata requirements of different information packages are not confused. So for example, the metadata required in an archival information package to support resource management and storage need not be part of the dissemination information package—unless the repository is part of a federation in which how an object is managed in that repository might impinge on its use elsewhere. Conversely, the metadata available for some information packages may be contingent on what is available in others. An example of this is the case of technical metadata for images (e.g. details of the camera settings when a photograph was taken), discussed in section 3.3 above, much of which is unlikely to be available if it is not created when the photograph is taken and present as part of the submission information package. Another example is information about the provenance and ownership of an object, the lack of which on submission creates extra work at ingest. In general the workflow question of what metadata is best created where is an important one in determining the quality of metadata (see for example Barton *et el* 2003).

It is probable that metadata generated by resource users (i.e. consumers in this model) will increasingly play an important role in resource description. User generated metadata encompasses usage data as described in section 3.14 above, but also includes classifications (tags) and descriptions generated by users of the resource. One of the complications to be considered in any federated architecture or other system based on separate data providers and service providers is whether and how the user generated metadata is aggregated in order to make it available to other services involved in managing and disseminating the resource.

Understanding what metadata is best produced at what stage (i.e. the workflow for metadata creation) and how metadata can be passed between repositories managing different copies of the same resource (i.e. repository interoperability infrastructure) will be important in determining what is practicably achievable for a metadata application profile for learning materials.

4.3. A model for resources

As part of this scoping study a model for resources has been discussed. The aim of the model is to show the different entities being described, the relationships between them, and to define what properties are described. In the DCMI Singapore Framework

(Nilsson *et al*, 2008) such a model is called the domain model and is a step in the process of defining the scope of the application profile. The model is presented here as a starting point for further discussion and serves as much to highlight problems as to solve them. It is by no means finalized, and the version below is just one of several variants on a theme considered during the time of writing.

The model follows the scholarly works application profile (SWAP) entity-relationship model, which is based on the Functional Requirements for Bibliographic Records (FRBR) entity model. The rationale behind this is that, firstly, scholarly works may be considered learning materials in higher education, so any model for learning materials should also be capable of describing scholarly works. Secondly, the FRBR model is well-tested and seems generic enough to describe many other types of resource (e.g. musical scores and performances, images, online resources).

The model is shown in figure 4.2. The "group 1" entities from FRBR in the centre of the figure (Work, Expression, Manifestation and Item) model the learning material at various degrees of abstraction; the related entities may also require description in order to support requirements relating to the learning material.

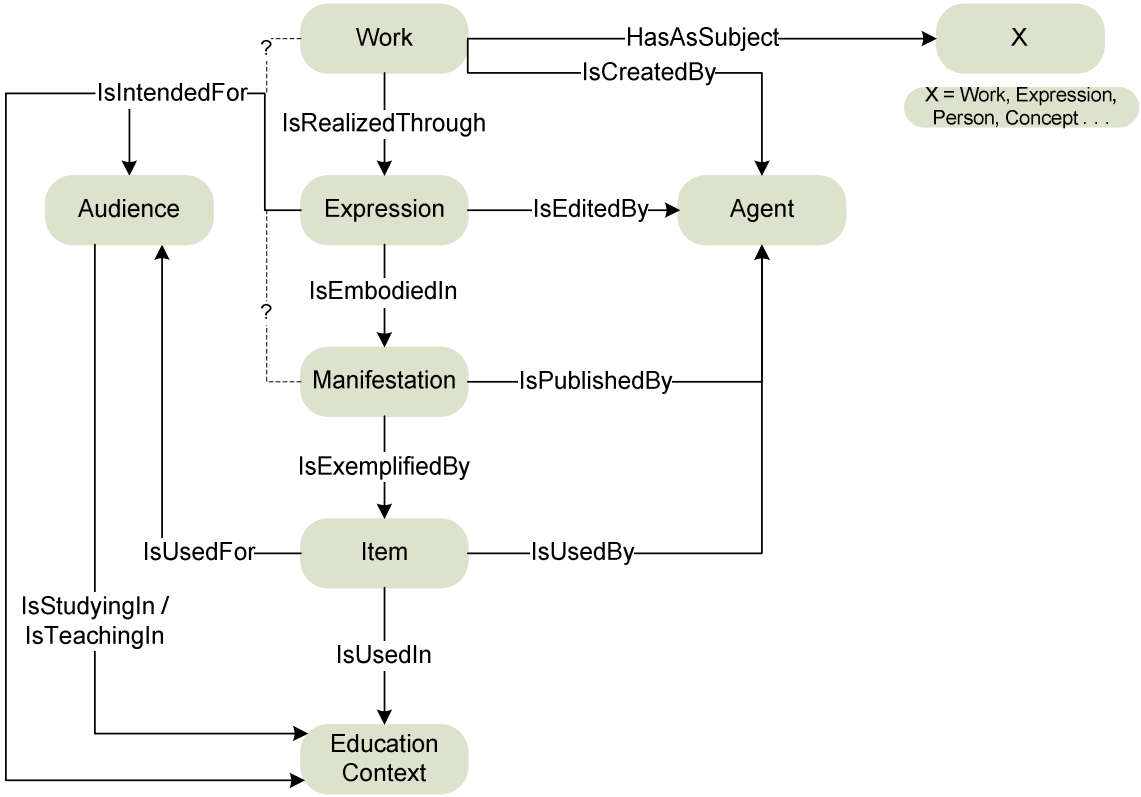

Figure 4.2: the entity-relationship model discussed for defining what is being described by metadata for learning materials.

4.3.1. The primary resource: the learning material

The FRBR group 1 entities, which model the learning material *per se*, and which are shown in the centre of the figure, are:

Work

Taken from FRBR, the work is the most abstract of the group 1 entities, modelling the learning material as "a distinct intellectual or artistic creation". A work is what we mean when we talk of Homer's *Iliad* or Shakespeare's *Romeo and Juliet* as an entity, without specifying a version of the text or performance of the play. A work is identifiable, is associated with a subject and we can talk about the creator of a work. Different versions of a work, i.e. variations which do not represent significantly distinct intellectual or artistic creations, such as updates, corrections, verbatim transcriptions or translations, arrangements of a musical piece, are modelled as expressions of the work. A work is a conceptual entity that allows us to talk about what these expressions have in common. By contrast, adaptation of a work for a different artistic medium (e.g. writing a film script based on a novel) or for a new audience (e.g. children) and writing summaries or abstracts are considered to involve the creation of a new work.

Expression

A FRBR Expression is a realization of the intellectual work in a specific form, for example as a text, a performance or a presentation. Different revisions or editions of the material are different expressions, as are translations, and different modalities (e.g. printed text and spoken word for book and audio book respectively). Expressions can be identified, as can the related editor (the individual or organization responsible for the production of that expression). Differences in physical format alone, e.g. page layout, hard or soft covering, disc versus tape or file download etc., do not constitute different expressions but are considered different manifestations of the same expression.

Expressions are often produced with a specific intended audience or for use in a specific educational context—e.g. speakers of a specific language[3], people with a preference for listening rather than reading, University-level students, to support a specific course.

Manifestation

A FRBR Manifestation is the physical embodiment of an expression a particular medium or format: for example, the text as a PDF, HTML or MS Word format file, or printed on paper; a performance or presentation as an

[3] If it weren't for the issue of audience language and translations being considered by FRBR to be expressions of the same work we could identify this relationship at the level of the work; conversely, when the accessibility needs of the audience are considered it may be that this relationship cannot be identified until we have a specific manifestation.

audio/video recording. Manifestations can be identified and related to an agent responsible for publishing them. A manifestation may be exemplified by many physical items.

Item

An item is a single exemplar or instance of a manifestation, and is the most concrete of the primary entities. For physical resources, the item is a copy of a book, disc or object that can be held and used; for digital resources an item is less concrete but may be taken to be the instantiation of a file or files on a specific system, e.g. the copy of a pdf file on my computer with my comments, rather than the copy on the server (Floyd and Renear, 2007).

At the level of the specific item we can identify the user of the resource, either as an individual (agent) or as a class (audience).

For complex/compound objects or aggregates any of these entities may comprise several parts, which may be expressed in the model by the FRBR "whole/part" ("has part" and "is part of") relationships. So, for example, it may sometimes be useful to describe a work as comprising several parts (e.g. parts or sections of a book, a story told as a trilogy of books). In the case of an anthology a single manifestation may aggregate manifestations embodying expressions of several works; conversely, in multivolume or multimedia editions the manifestation of a single work may have several parts, e.g. printed volumes, book and CD-ROM / website, text and image files etc.

An example of breaking down a typical learning resource into FRBR Group 1 entities is provided in Appendix 2.

4.3.2. Secondary entities

Secondary entities, which are related to the learning material and which have properties that need to be specified in order to provide a full description of it, are

Agent

Following SWAP, the agent entity is a merging of the FRBR person and corporate body entities and models a specific individual or organization that has an important connection with the resource. The relevant agents, about whom it might be useful to record details (name, affiliation, contact details etc.), are the resource creators, editors, publishers and users.

Two examples of queries which require information about agents are: find me materials created by William Shakespeare, and, find me the resources that were used by my predecessor teaching this course.

Audience

The audience represents a class of users for which a resource is intended or which actually uses the resource. For learning materials the audience will be studying or teaching in a specific educational context. Properties of the

audience include the language spoken, the educational level, and preferences for accessing the resource and for teaching and learning approach.

Education Context

Agents and audiences directly involved in education (that is teachers and learners) will be in a specific context. Context includes modalities such as distance learning or face-to-face, formal or informal learning, the type of institution, cultural issues etc. For formal learning we will likely be able to characterize some of the following: the qualification being studied for, the intended learning outcomes, a course code or identifier.

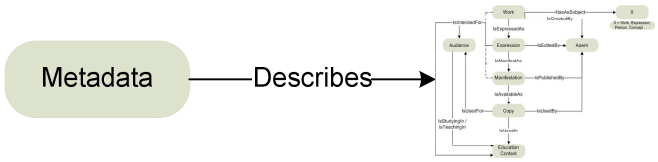

Figure 4.3: the entity-relationship model discussed for defining the relationship between the metadata description and the learning materials.

Metadata

Figure 4.3 shows the simplest model possible for the relationship between the metadata and the resource being described. The metadata entity has the properties necessary for collection management (e.g. creator, owner and date of creation of the metadata, annotation relevant to the metadata etc.). Some of this information could also be useful in helping users establish the basis for claims made in a metadata record, e.g. who says this resource is intended for the educational level stated? However this approach may be too simple in assuming that the metadata is a single record describing the whole resource rather than a set of statements from diverse sources, each describing specific entities or properties (Hillmann *et al,* 2004; Downes, 2004).

4.3.3. Properties of the entities

Appendix 1 presents a draft overview of suggested properties for each of the entities listed in sections 4.3.1 and 4.3.2, and also attempts to indicate the metadata range for each property, i.e. the relevancy of each property to the metadata types listed in section 3.

4.3.4. Observations

It should be stressed again that the model presented is one variant that has been considered. Possible variations mostly centre on the properties and relationships between entities, and the version presented here probably includes a fair amount of redundancy. For example, is it necessary to have both a direct link from the primary resource to the intended educational context and a relationship via the audience? Is educational level a property of audience or context? Indeed, is it possible to merge

audience and context? In part, this represents a choice of conceptualization of successful resource discovery: a person who is looking to use a resource (i.e. looking to set up an "is used by" relationship to a copy of the resource) may be conceptualized as either looking for a resource designed for or used by an audience of which they are a member or as looking for a resource for use in educational circumstances similar to their own. Other possible variations which are not shown are those that would arise if one were primarily interested in modelling the context (e.g. for course description) or agents (e.g. for attention profiling). Thus there could be a "has as subject" link for education context and "has interest in subject" links for agent and possibly audience, but these don't seem necessary for learning material description.

A large amount of the educationally relevant information seems to pertain not to the resource itself but to the intended or actual audience or context. This is not surprising if one accepts that "anything used for teaching or learning" may be an educational resource, as suggested in section 2.1, since this implies that the defining quality of an educational resource lies in (or results from) its use, i.e. the users and their context, rather than being inherent to the resource itself. As a consequence, the part of the model that represents the resource itself is not specific to learning materials, rather it is one which is rooted in bibliographic metadata (although at the level of detail used here it seems there is nothing in the model that precludes other types of resource being described). It may be possible to use other resource models in place of the group 1 FRBR entities and there may be advantages in doing so in specific circumstances. It is not clear whether having a single model for learning materials that can be used with any resource type is at all preferable to defining units of a model that can be plugged into other domain models in order to add the entities and properties necessary for learning materials. The second, modular, approach is being followed by the DCMI Education Working Group.

5. Discussion

5.1. The multifaceted nature of metadata for learning resources

During this work we have tried to avoid assumptions about the way in which a learning materials application profile might be used within the UK further and higher education sector. We have not limited the scope of our work to resource discovery, either locally at the level of the service providing the metadata or remotely through a service provider,— we believe that metadata has a valuable role in supporting the management of locally available learning materials within an institution as well as a role in sharing materials and information about their use between and beyond institutions. We have not assumed anything other than the most inclusive definition of what might constitute learning materials, and we have tried to avoid assumptions relating to what metadata schema (e.g. LOM or Dublin Core) would be used. Finally we have not assumed that the learning materials are situated in a single formal repository: they may be in several distinct repositories, or in a VLE, or on the web[4]. The net result has been something akin to scoping an application profile without knowing the application or what is being profiled.

We started with a definition of metadata that asserted that metadata could support resource management, retrieval and use; and this assertion has been echoed in the metadata fields suggested by experts from the various domains we examined. Resource management is seen as one of the core functions of a repository with respect to learning materials (Campbell, 2008) and, for example, knowledge of IPR ownership and licensing conditions, resource use and resource coverage are important in resource management. While full text searching meets most requirements for resource discovery and retrieval for materials that are largely text-based or are closely associated with a textual resource, for non-textual materials the metadata record may be required to furnish this searchable associated text. When dealing with education metadata this study has recognised a requirement for what might be termed "inspirational" metadata, that is, teachers looking for information on how they might use specific learning materials.

Any future work should recognise that metadata may be valuable in supporting resource management, discovery/retrieval and use; and that an application profile to support resource discovery might be quite different from an application profile to support resource management or use.

5.2. Relationship to type-specific metadata

We also based this work on the premise that "learning materials" are not a resource type *per se* in the way that journal articles, images or audio and video recordings can be described as resource types. Instead we adopted a loose and pragmatic definition that a

[4] However any metadata application profile such as LMAP is only relevant where the description of resources is managed, at least to the extent of conforming to a metadata profile, and we recognise that the reality is that this will not apply to most resources in VLEs or on the web.

learning material is anything that may be used in teaching and learning. We believe that a definition along these lines is necessary to meet the needs and expectations of teachers and learners, and to reflect the fact that many learning materials are aggregated multimedia resources. However it is worth remembering that many of the examples of successful repositories (by which we are thinking of arXive, Getty Images, Flickr, YouTube, iTunes, SlideShare and several examples of image or media libraries of smaller scope) all limit their scope to a single simple resource type. Similarly for physical materials, cataloguing in libraries and archives tends to be resource type specific rather than against a single grand schema that can be used for books, archive deposits, journals, audio-visual material etc. A decision needs to be made as to whether the component media assets comprised in an aggregate learning material are treated as significant entities in their own right and merit management as such, or whether they are valuable only as part of an aggregation. Likewise a decision needs to be made as to whether the aggregation has value that is not inherent in component parts individually. Many factors will affect these decisions, and they can only be made on a case by case basis at service level and resource level. The outcome of these decisions will be that some resources are described in "broad-brush" generic terms and others will be described in detail as aggregations of related resources with some parts of significant value in their own right. We should not mandate detailed metadata for all resources, but nor should we ignore the requirement of detailed metadata for some.

Any future work on learning material application profiles should allow resources to be described either as generic learning materials or with type-specific metadata.

5.3. A heterogeneous environment

It is unlikely that many sources of metadata will provide complete resource-type specific metadata and complete descriptions of all the educational aspects of a resource, and it is certain that such comprehensive metadata will not be available for all resources. However, it is important to note that for a given resource it may be the case that all this information is available by combining data from several sources. For example, an image bank might describe the content and technical details of an image, while another service elsewhere describes how it has been used in education to illustrate a specific point. The same applies to other resource types, e.g. moving images, audio, text resources etc, however the information available about each of these resource types will be different. Services built on metadata for learning materials will probably have to deal with metadata from multiple sources, some of which will describe resources as generic learning materials and others which will provide metadata specific to various resource types. For example, within an institution there may be images used for teaching and learning that are managed in a dedicated image bank, assessment items that are managed in a dedicated item bank, both of which will have specialised metadata; there may be course resources from the VLE in an associated repository, with metadata from the resource creator and information on what courses they have been used in from the VLE (these resources may include images and assessment items from the other services). Nationally and internationally there will be an even greater variety of services, possibly

with information relating to the same resources that are held in the institution, especially if we consider the resource description potential of social networks, tagging, blogs, the semantic web etc. So a service built to help teachers within an institution discover and appraise what materials are available for use in their teaching would have to cope with heterogeneous metadata from multiple sources in several educational jurisdictions. Building such services is challenging. Application profiles are one approach aimed at trying to reduce this problem by creating a more homogeneous environment within a federation of repositories and related services. The disadvantages of this approach are that the burden of providing the metadata required to join the federation may be too onerous or inappropriate for some sources of information and that the resulting services within the federation may fail to exploit all the information that is available (since not everyone will be in the federation). While application profiles are helpful in this context, other complementary solutions are also necessary for learning materials and probably for other applications as well.

We hope that the JISC will explicitly include learning materials in any work aimed at a better understanding of how to create services based on heterogeneous metadata from multiple sources.

5.4. The requirements for education metadata

One can imagine an approach where descriptions of the educational properties of learning materials are held separately from the type-specific descriptions of the resources, to be combined in a service facilitating resource discovery or management for teaching and learning. Unfortunately, it was noticeable during the course of the study that metadata for some resource types and purposes was better understood than for others, and that metadata for education was one of the domains where the issues were least well articulated and where solutions were least well developed. In many of the areas where resource description was seen as "difficult" there seems to be some suggestion of movement away from structured metadata and towards the approach of providing semi-structured free text descriptions. We see this in the domain of accessibility and in the experience of using learning designs[5] to provide information on how learning materials are used. While this seems a pragmatic solution the following questions remain: what properties should be described and are these properties of the material or of its use; which parts of a resource description need to be provided as structured metadata to be easily intelligible to computers and which can be left as free text to be interpreted by humans; how might both structured metadata and human-oriented resource descriptions be created most efficiently and effectively?

More work is needed to understand the requirements for the description of the educational properties of learning materials and how such a description might be created.

---

[5] Using the term learning design (lower case) to refer to a free text description of what could be encoded in a fully structured form following the IMS Learning Design (upper case) specification.

Work on domain models for resources both for this report and in the DC-Education working group has highlighted that much of what is educationally relevant information about a learning material pertains not to the material itself but to associated entities (the resource creator, intended audience and users) and to their context. Currently this information is frequently not collected, which would assist resource management, or shared, which would assist resource discovery and use. For example, it does not seem to be the case that services hosting and disseminating learning materials receive information from VLEs about which materials are used for which courses. Even where such information is recorded there is no practice of sharing it, which in itself poses many questions: precisely what information should be shared, if copies of an open resource are used in many institutions should the information be collated, and if so where, should the service that manages this information necessarily be the same as the one hosting and managing the resource? Of course such information need only be disseminated to those who have access, or potentially have access, to the learning materials. If the learning materials are only available within a single educational setting (an institution, or maybe just a department within an institution) then gathering the information to aid resource management is less complicated, being reduced to one of integrating the resource management service with the learning and teaching environment. If the resources are shared between institutions, and if successful dissemination of a learning material depends on successful dissemination of information about its intended and actual educational use, then these issues need to be addressed.

Following other JISC application profile studies we have examined the use of FRBR as the basis for a model of the resource being described. Quite apart from the necessary addition of secondary entities (users, audiences, context), the fact that many learning resources are aggregations and available in many formats means that FRBR-based descriptions can become very complex. The model does however seem to have potential in showing relationships between resources in a way that would be useful to teachers and learners. Implementing this sensibly in cataloguing systems, educating those creating resource descriptions in how to apply it to learning materials, and exposing the information in a way useful to users would all pose substantial challenges.

5.5. Conclusion

In conclusion, it seems that learning materials cannot be treated as a single resource type, and the underlying requirements for metadata for learning materials are multifaceted, and quite likely different to those for metadata for other types of resource (just as the case for sharing learning materials is quite likely different to the case for sharing resources such as research publications). Success of an application profile for a relatively simple and relatively well understood resource type such as research publications cannot automatically be rolled out to learning materials without first solving many other problems. In our view the first priorities for work aimed at providing descriptions of learning materials within JISC environments should be to clarify how to capture and encode educationally relevant information about resources in a way that this can be used to enrich resource descriptions in pedagogic settings.

6. References

6.1. Articles, web sites and books (by Author)

Allinson, J. (2006). OAIS as a reference model for repositories : an evaluation. UKOLN. URL http://www.ukoln.ac.uk/repositories/publications/oais-evaluation-200607/Drs-OAIS-evaluation-0.5.pdf

Barton, J., Currier, S., & Hey, J. M. N. (2003). Building quality assurance into metadata creation: an analysis based on the learning objects and e-prints communities of practice. In S. Sutton, J. Greenberg, & J. Tennis (Eds.) *Proceedings 2003 Dublin Core Conference: Supporting Communities of Discourse and Practice - Metadata Research and Applications* . Dublin Core Metadata Initiative. URL http://purl.oclc.org/dc2003/03barton.pdf or http://eprints.erpanet.org/83/

Bruce, R. (2008). Metadata for stuff in repositories... URL http://infteam.jiscinvolve.org/2008/04/24/metadata-for-stuff-in-repositories/

Campbell, L. M. (2008). "Better management and sharing of teaching and learning materials by individual teaching practitioners". URL http://blogs.cetis.ac.uk/lmc/2008/08/22/%E2%80%9Cbetter-management-and-sharing-of-teaching-and-learning-materials-by-individual-teaching-practitioners%E2%80%9D/

Currier, S., Campbell, L. M., & Beetham, H. (2005). Pedagogical vocabularies project. report 1: Pedagogical vocabularies review. JISC. URL http://www.jisc.ac.uk/elp_vocabularies.html

Downes, S. (2004). Resource profiles. *Journal of Interactive Media in Education,* 2004 (5) Special Issue on the Educational Semantic Web. URL http://www-jime.open.ac.uk/2004/5/

Duval, E. (ed.). Metadata 2.0 workshop. Katholieke Universiteit Leuven. URL: http://ariadne.cs.kuleuven.be/mediawiki2/index.php?title=Metadata_2.0

Floyd, I. R., & Renear, A. H. (2007). What exactly is an item in the digital world? In *The American Society for Information Science & Technology Annual Meeting 2007, Milwaukee, Wisconsin.* URL https://www.ideals.uiuc.edu/handle/2142/5254

Heery, R., & Patel, M. (2000). Application profiles: mixing and matching metadata schemas. *Ariadne* 25. URL http://www.ariadne.ac.uk/issue25/app-profiles/

Hillmann, D, I., Dushay, N. & Phipps, J. (2004). Improving metadata quality: augmentation and recombination. *DC-2004 International Conference on Dublin Core and Metadata Applications, 11-14 October 2004, Shanghai, China*. URL http://hdl.handle.net/1813/7897

MEG (2003). UK Educational Levels. URL http://www.ukoln.ac.uk/metadata/education/ukel/

Moat Project (no date). Moat project ontology. URL http://moat-project.org/ontology

MIT, (no date). MIT OpenCourseWare home page. URL http://ocw.mit.edu/

Newman, R. (2005). Tag ontology design. URL: http://www.holygoat.co.uk/projects/tags/

Nilsson, M. (2008). Harmonization of metadata standards. Tech. rep., PROLEARN Consortium. URL http://ariadne.cs.kuleuven.be/lomi/images/5/52/D4.7-prolearn.pdf

Nilsson, M., Baker, T., & Johnston, P. (2008). The Singapore framework for Dublin core application profiles. URL http://dublincore.org/documents/singapore-framework/

NISO (2004). Understanding metadata. URL http://www.niso.org/publications/press/UnderstandingMetadata.pdf

Papert, S. (1994). *Mindstorms : Children, computers and powerful ideas.* New York: Basic Books, second ed.

SQA (no date) Catalogue of national qualifications. Available as an Excel spreadsheet via http://www.sqa.org.uk/sqa/3.html

TechDis (no date). Accessibility passport. URL http://www.techdis.ac.uk/index.php?p=9_6_2

Van Assche, F. (2007). Linking learning resources to curricula by using competencies. In D. Massart, & J.-N. Colin (Eds.) *First International Workshop on Learning Object Discovery & Exchange*. European Schoolnet (EUN). URL http://fire.eun.org/lode2007/lode11.pdf

6.2. Specifications, standards and related bodies (by abbreviation)[6]

CLA: Copyright Licensing Agency. http://www.cla.co.uk/

DACS: design arts and copyright society http://www.dacs.org.uk/

DCAM: DCMI Abstract Model. URL http://dublincore.org/documents/abstract-model/

DCAP: Guidelines for Dublin Core Application Profiles. http://dublincore.org/documents/profile-guidelines/

DCEd: DCMI Education Community. Homepage http://dublincore.org/groups/education/

DCES: Dublin Core Metadata Element Set. http://dublincore.org/documents/dces/ (also ISO 15836:2003 Information and documentation—the Dublin Core metadata element set).

DID: ISO/IEC 21000-2:2003, Information technology—Multimedia framework (MPEG-21) -- Part 2: Digital Item Declaration

FOAF: Friend of a Friend project. http://www.foaf-project.org/

FRBR: Functional Requirements for Bibliographic Records. http://www.ifla.org/VII/s13/frbr/index.htm

IMSAcMD: IMS AccessForAll Meta-data http://www.imsglobal.org/accessibility/index.html

IMSCP: IMS Content Packaging. http://www.imsglobal.org/content/packaging/

IMSQTI: IMS Question and Test Interoperability specification. http://www.imsglobal.org/question/

ISOAcMD: ISO/IEC 24751-3:2008, Individualized Adaptability and Accessibility in E-learning, Education and Training Standard Part 3: "Access for All" Digital Resource Description.

ISO MLR: ISO/IEC 19788, Information technology—Learning, education and training—Metadata for learning resources.

---

[6] Availability of standards and specification documents:

- ANSI/NISO Standards are available at no cost from http://www.niso.org/standards/
- Dublin Core specifications are available at no cost at the URLs given.
- EEE Standards may be purchased online via http://standards.ieee.org/
- IETF Requests for Comments (RFCs) are available at no cost from the URLs given.
- IMS specifications are available at no cost from the at the URLs given.
- ISO and ISO/IEC publications may be purchased via http://www.iso.ch/
- Other specifications are available from the URLs provided.

LOM: IEEE 1484.12.1—2002, Standard for Learning Object Metadata.

MARC: MARC Standards. http://www.loc.gov/marc/

MDRSIG: CETIS Metadata and Digital Repository Special Interest Group. Homepage URL http://wiki.cetis.ac.uk/Metadata

METS: Metadata Encoding & Transmission Standard. http://www.loc.gov/standards/mets/

MIX: "NISO Metadata for Images in XML (NISO MIX)" http://www.loc.gov/standards/mix/ (see also Z39.87)

OAI-ORE: Open Archives Initiative Object Reuse and Exchange. http://www.openarchives.org/ore/

OAIS: ISO 14721:2003, Space data and information transfer systems – Open archival information system – Reference model. Also available at http://public.ccsds.org/publications/archive/650x0b1.pdf

ODRL: Open Digital Rights Language. http://odrl.net/

OpenURL: ANSI/NISO Z39.88-2004 The open URL framework for context-sensitive services

PINGBACK. Pingback specification. http://www.hixie.ch/specs/pingback/pingback

PLUS: Picture Licensing Universal Scheme http://www.useplus.com/

PREMIS: PREMIS Data Dictionary for Preservation Metadata. http://www.loc.gov/standards/premis/

PRISM: Publishing requirements for industry standard metadata. http://www.prismstandard.org/

SWAP: Scholarly Works Application Profile. http://www.ukoln.ac.uk/repositories/digirep/index/SWAP

TRACKBACK: TrackBack technical specification. http://www.sixapart.com/pronet/docs/trackback_spec

UK LOM Core: The UK Learning Object Metadata Core. URL http://zope.cetis.ac.uk/profiles/uklomcore

vCard: IETF RFC 2435: 1998 MIME Content-type for Directory Information, http://tools.ietf.org/html/rfc2425, and IETF RFC 2426: 1998 vCard MIME Directory Profile http://tools.ietf.org/html/rfc2426

XCRI: eXchanging course-related information. http://www.xcri.org/

XrML: eXtensible rights Markup Language. http://www.xrml.org/

Z39.87: ANSI/NISO Z39.87-2006, Data Dictionary – Technical Metadata for Digital Still Images.

A1. Properties of entities in the model of resources

The tables in this appendix show a suggestion of what properties might be worth describing for each of the entities in the model for resources presented in section 4.3, and the range of relevancy of these properties in terms of the metadata domains discussed in section 3.

It must be stressed that this is incomplete and debatable: a complete version is not possible with the information gathered during this scoping study. Indeed it would probably require an actual application profile to be agreed and then stakeholders from each of the domains could indicate which properties were relevant to their work. Other issues with the tables below are the following. The columns represent such different concepts that it is difficult to compare them. So, for example, the "Scholarly Communication" column represents the information required for disseminating journal papers, which cannot easily be compared with the information in the "Technical" column which indicates those properties that are of a technical nature. Also, some columns such as that for scholarly communication stand alone; others make sense only if combined with another, for example the properties checked in the "Curation and preservation" column are an incomplete set unless merged with properties for a resource type.

The tables are presented here in order to flesh out the model for resources and to show how it handles the metadata requirements elicited during this project. We also believe that completing such a table might be of use in building an understanding of the interplay between metadata requirement from different communities, which would be necessary for an application profile that cuts across many distinct resource types and purposes.

Table A1.1: Possible properties of Works and their range of use

Suggested properties of works as described in the model for resources presented in section 4.3, and the relevancy of these properties to the range of metadata discussed in section 3. It must be stressed that this is incomplete and debatable.

Possible properties of Works and their range of use	Scholarly Communication	Image	Moving Image	Geospatial	Complex Object	Technical	Personal & Organizational	Resource Discovery	Collection Management	Curation & preservation	Rights	Accessibility	Usage	Assessment	Education
Title	X	X	X					X							
Abstract / description	X	X	X					X							?
Subject Classification or HasAsSubject relation	X	X	X	X				X	X						X
Keywords (free text)	?	X	X					X						?	X
Rights info(?)		?	?					?	?		X				?
Has adaptation	X	?	?					?				?			?
Whole/part relation			?		X			X	X					?	X
IsCreatedBy (Agent) relation	X	X	X					X	X		X				X

Table A1.2: Possible properties of Expressions and their range of use

Suggested properties of expressions as described in the model for resources presented in section 4.3, and the relevancy of these properties to the range of metadata discussed in section 3. It must be stressed that this is incomplete and debatable.

Possible properties of Expressions and their range of use	Scholarly Communication	Image	Moving Image	Geo spatial	Complex Object	Technical	Personal & Organizational	Resource Discovery	Collection Management	Curation & preservation	Rights	Accessibility	Usage	Assessment	Education
Title	X	?	X					X							
Description	X	?	?					X						?	X
Date available	X	X	X											X	
Status	X	?	X			X				?				X	
Version	X	?	X			X		X	X	?				X	
Language	X		X					X						?	
Genre/type	X	?	?					X						?	
Modalities?												X			
Copyright holder	X	X	X					X	X	X	X			X	
Bibliographic citation	X							?							
References	X							?							

Possible properties of Expressions and their range of use	Scholarly Communication	Image	Moving Image	Geo spatial	Complex Object	Technical	Personal & Organizational	Resource Discovery	Collection Management	Curation & preservation	Rights	Accessibility	Usage	Assessment	Education
Time to use														X	X
Pedagogic approach															X
Has Version	X	?	?					X	X			X		?	
Has Translation	X		X					X						?	
Is Part Of (collection)		X						X	X					?	
Whole-part relation			?		X			X	X	?				X	
Is Edited By (agent) relation		?	?						X	?					
Is Intended For (audience / context) relation														X	X

Table A1.3: Properties of manifestations and their range of use

Suggested properties of manifestations as described in the model for resources presented in section 4.3, and the relevancy of these properties to the range of metadata discussed in section 3. It must be stressed that this is incomplete and debatable.

Properties of manifestations and their range of use	Scholarly Communication	Image	Moving Image	Geo spatial	Complex Object	Technical	Personal & Organizational	Resource Discovery	Collection Management	Curation & preservation	Rights	Accessibility	Usage	Assessment	Education
Format	X	X	X			X		X		X		?			
Date modified	X	X	X					X	X						
Scale factor / magnification		X													
Other technical		X	X		?	X				X		?			
Date of inclusion in collection									X						
Accessibility information								X				X			
Whole-part relation					X			X	X						?
Is Published By (agent) relation								X		X	X				X

Table A1.4: Properties of items and their range of use

Suggested properties of copies as described in the model for resources presented in section 4.3, and the relevancy of these properties to the range of metadata discussed in section 3. It must be stressed that this is incomplete and debatable.

Properties of items and their range of use	Scholarly Communication	Image	Moving Image	Geo spatial	Complex Object	Technical	Personal & Organizational	Resource Discovery	Collection Management	Curation & preservation	Rights	Accessibility	Usage	Assessment	Education
Date available	X														
Rights, terms, conditions	X	X	X					X	X	?	X				X
Location (=URL =identifier)	X	X	X			X		X	X	?					
Whole Part relation															
Is Used By (agent) relation								X	X				X		
Is Used In (context) relation															

Table A1.5: Properties of Agent and their range of use

Suggested properties of agents as described in the model for resources presented in section 4.3, and the relevancy of these properties to the range of metadata discussed in section 3. It must be stressed that this is incomplete and debatable.

Properties of Agent and their range of use	Scholarly Communication	Image	Moving Image	Geo spatial	Complex Object	Technical	Personal & Organizational	Resource Discovery	Collection Management	Curation & preservation	Rights	Accessibility	Usage	Assessment	Education
Type of agent	X						X								X
Name	X						X	X							
Family name	X						X								
Given name	X						X								
Mailbox	X						X								
Homepage	X						X								
Affiliation							X	X							?
Location							X								?
Workplace homepage	X						X								
Age?															?

Table A1.6: Properties of Audience and their range of use

Suggested properties of works as described in the model for resources presented in section 4.3, and the relevancy of these properties to the range of metadata discussed in section 3. It must be stressed that this is incomplete and debatable.

Properties of Audience and their range of use	Scholarly Communication	Image	Moving Image	Geo spatial	Complex Object	Technical	Personal & Organizational	Resource Discovery	Collection Management	Curation & preservation	Rights	Accessibility	Usage	Assessment	Education
Role	X						X								X
Language															X
Level? and/or Age range?															X
Access preferences												X			
Teaching/learning preferences															X
Cultural sensibilities?		X	X					X							?
Relation to expression (is intended for)								X							?
Relation to education context (is studying in / is teaching in)															X
relation to copy (is used for)												?	X		?

Table A1.7: Properties of Context and their range of use

Suggested properties of works as described in the model for resources presented in section 4.3, and the relevancy of these properties to the range of metadata discussed in section 3. It must be stressed that this is incomplete and debatable.

Properties of Context and their range of use	Scholarly Communication	Image	Moving Image	Geo spatial	Complex Object	Technical	Personal & Organizational	Resource Discovery	Collection Management	Curation & preservation	Rights	Accessibility	Usage	Assessment	Education
Learning outcome / competency														X	X
Level?														X	X
Language?														X	X
Pedagogic approach ...														X	X
Course details ...														X	X
Relation to expression (Is intended for)								X						X	X
Relation to copy (is used in)									X		X			X	X
Relation to audience														X	X

A2. An example of applying FRBR to a learning material[1]

The example shows the application of the FRBR model to a course unit (a lecture, plus supporting material such as power point slides and problem sheet). The example is based on a module of an Open Yale course[2] with the modification that whereas in the original the overhead notes were written and displayed on the blackboard during the lecture delivery in the example they are envisaged as being PowerPoint slides and made available separately.

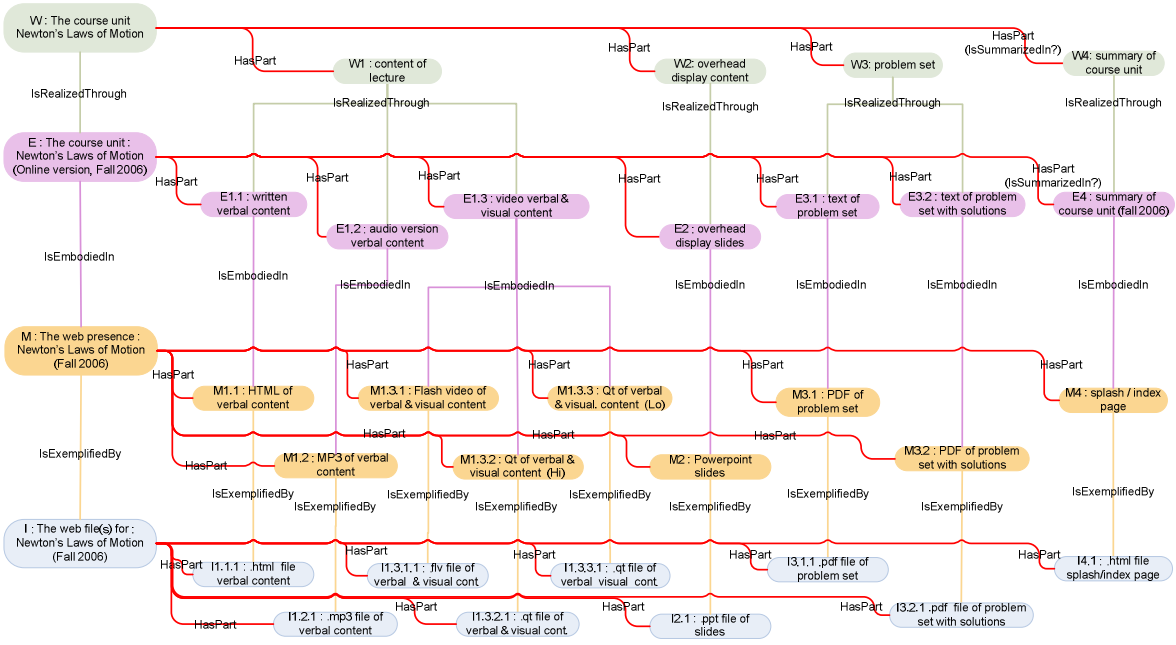

Figure 1: a graphical representation of the breakdown of an online course module into FRBR components.

Works

The course unit is treated as an aggregate work, W, comprising: W1, the content of the lecture; W2, the overhead display content; W3, the problem set; and W4, a summary of the course unit.

W: The course unit "Newton's Laws of Motion"
- has part W1: content of lecture
- has part W2: overhead display content
- has part W3: problem set
- has part W4: summary of course unit

[1] Many thanks to John Robertson (JISC CETIS and CAPLE, Strathclyde University) for his help in preparing this section.

[2] available at http://oyc.yale.edu/physics/fundamentals-of-physics/content/sessions/lecture03.html

Discussion:

The reason for treating this collection of works as being aggregated into a single work is that the works collectively seem to have a single objective (to explain Newton's laws of motion) and it is likely that the component works were conceived as being complementary to each other in achieving this aim. So, while the lecture, problem set and overhead display content each stand in their own right as independent endeavours, the aggregation of them aims to achieve something that cannot be achieved by those works independently. The course summary is modelled as a part of the aggregate work, though it could also be a separate work summarizing the course unit (indeed it might even be a version of the metadata for the course).

Some relationships are not modelled here. The course unit is a part of a larger work (the course) and has a sequential relationship with other parts of that course. The course as a whole, and its constituent course units, might be an updated version of a previous course, or may have been subsequently updated to give another version. Also, in certain circumstances the problem set might be an aggregation of individual problems which are themselves independent works.

The reading assignment is modelled here as a related work (at expression level since it might change year on year, see below). In other courses the reading list may have a similar role to the problem set here, the creation of which represents enough intellectual effort to justify it being considered a work.

The overhead display content is envisaged as being an independent work, in some cases it may be no more than a summary of the lecture content, or even just an expression of that content.

Expressions

One of the expressions through which the main aggregate work realized is the online version of the course unit from Fall 2006, E, which comprises expressions of the component works. The content of the lecture is realized through a written transcript (which is assumed to be edited), an audio recording, and a video recording (which includes sound). The overhead display content is conceived as being realized through a set of slides. The problem set has two expressions as text, with and without solutions. The summary of the course unit is realized through text, graphical layout and the expression of relationships (as hyperlinks).

W: The course unit "Newton Laws of Motion"

E: The course unit "Newton Laws of Motion" (Online version, Fall 2006)

W1: content of lecture

E1.1: written verbal content

E1.2: audio version of verbal content

E1.3: video of verbal and visual content

W2: overhead display content
    E2: overhead display slides

W3: problem set
    E3.1: text of problem set without solutions
    E3.2: text of problem set with solutions

W4: summary of course unit
    E4: summary of course unit (Fall 2006)

E: The course unit "Newton Laws of Motion" (Online version, Fall 2006)
    has part E1.1: written verbal content
    has part E1.2: audio version of verbal content
    has part E1.3: video of verbal and visual content
    has part E2: overhead display slides
    has part E3.1: text of problem set without solutions
    has part E3.2: text of problem set with solutions
    has part E4: summary of course unit (Fall 2006)

Discussion

The realization of the main aggregate work modelled here is just one of its possible expressions. Others would include the "real-life" expressions, i.e. the lecturer in the classroom with handouts etc., which apparently occur twice a year. Treating these as expressions parallels the treatment of performances of a musical work by the composer as an expression in the FRBR report (assuming that the content of the lecture does not change significantly between course offerings). This is assuming that there is no significant change in the lecture between instantiations which would be sufficient to model different instantiations as realizations of distinct but related works.

It seems natural that the realization of an aggregate work should result in an aggregate expression comprising relevant expressions of the component works. There would certainly be value in thus showing the sibling relationships between the various expressions in order to facilitate discovery of a suitable record of, say, the lecture content.

The reading assignment is not shown, but would be modelled as a related manifestation of an independent work at the level of manifestation since the assignment may change without significantly affecting the nature of the course unit. This parallels the inclusion of "references" as an attribute of an expression in the Scholarly Works Application Profile[3].

The video recording of the lecture performance is not modelled as a distinct work. The content of the video recording is not significantly different to that of the other recordings: it contains some hand waving and some notes on the blackboard but the

[3] http://www.ukoln.ac.uk/repositories/digirep/index/Model

omission of these from the other recordings was more the result of limitations in the media used for these recordings than any intellectual effort (alternatively the inclusion of, say, overhead display content may signify that the video recording comprises manifestations of two works). Secondly it is assumed that the creation of this recording does not involved any significant cinematography. Thus the video recording is treated as on a par with the recordings in other media, and follows examples of audio books as being expressions of the same work as printed books[4]. Showing the relationship of these different formats to the same work may have accessibility benefits, for example when needing to provide a version suitable for students with hearing difficulties.

Manifestations

The online version of the course unit from Fall 2006 is embodied in a section of the Open Yale course website. The entry point for this section is the "splash" page, and the parts of it are manifestations embodying the expressions described above; again these are modelled as component parts of an aggregate object. The only expression embodied in more than one manifestation is the video of the lecture which has manifestations in different formats and quality.

E: The course unit "Newton Laws of Motion" (Online version, Fall 2006)
    M: The web presence for "Newton Laws of Motion" (Online, Fall 2006)

E1.1: written verbal content
    M1.1: HTML version of verbal content

E1.2: audio version of verbal content
    M1.2: MP3 version or verbal content

E1.3: video of verbal and visual content
    M1.3.1: Flash video version of verbal and visual content
    M1.3.2: QuickTime version for high bandwidth of verbal and visual content
    M1.3.2: QuickTime version for low bandwidth of verbal and visual content

E2: overhead display slides
    M2: PowerPoint version of display slides

E3.1: text of problem set without solutions
    M3.1 PDF version of problem set without solutions

E3.2: text of problem set with solutions
    M3.2.1 PDF version of problem set with solutions

E4: summary of course unit (Fall 2006)
    M4 HTML "splash" page

M: The web presence for "Newton Laws of Motion" (Online, Fall 2006)
    has part M1.1: HTML version of verbal content

[4] See the treatment of "Harry Potter and the Goblet of Fire" by William Denton at <http://www.frbr.org/eg/hp-goblet-1.html>.

has part M1.2: MP3 version of verbal content
has part M1.3.1: Flash video version of verbal and visual content
has part M1.3.2: high bandwidth QuickTime version of verbal and visual content
has part M1.3.2: low bandwidth QuickTime version of verbal and visual content
has part M2: PowerPoint version of display slides
has part M3.1 PDF version of problem set without solutions
has part M3.2.1 PDF version of problem set without solutions
has part M4 HTML "splash" page

Discussion

The possibility of one of the component expressions relating to two of the component works is raised above (the video recording might capture the lecture and the overhead display); a similar possibility might occur if one or more of the recordings were to be packaged with the overhead content in a single manifestation. This might well occur if the slides were made available as a "slidecast", a set of slides streamed with automatic transitions synchronized with an audio recording[5].

Showing the relationships between alternative manifestations may have accessibility benefits, for example allowing a user to obtain the material in a format compatible with their requirements.

Items

Individual files are considered to be the relevant items for each manifestation[6]. There will be multiple items for each manifestation (the files on the server, the files down-loaded for viewing, copies of files kept elsewhere), although this is not shown in the diagram, and there is no particular interest in enumerating the items here.

Discussion

It's not really clear what the item of the aggregation actually is, except as a notional entity (i.e. the collection of files on the server); however it seems clear that not all copies of the component items will be part of a complete copy of a complete aggregation item. One case where there could clearly be seen to be a complete aggregation item would be where the whole thing were available in a single archive file, e.g. a .zip file, or in a similarly packaged format (e.g. an IMS Content Package—although this case may be considered a different manifestation from the website).

Knowledge of the whereabouts of items exemplifying particular manifestations and how these items relate to expressions may be useful in activities related to digital curation.

---

[5] http://www.slideshare.net/faqs/slidecast

[6] For a discussion of this see I. R. Floyd and A. H. Renear, "What exactly is an item in the digital world?" in *The American Society for Information Science & Technology Annual Meeting 2007, Milwaukee, Wisconsin.*, October 2007. [Online]. Available: http://hdl.handle.net/2142/5254

A3. Contributors

Informants

For each metadata domain we identified one or more representative informant, for example from a JISC Advisory service or other expert group. The informants were:

Gayle Calverley	Manchester University / time-based AP project	Moving images (time based media)
Lorna M. Campbell	JISC CETIS	Education
Sarah Currier	Intrallect Ltd.	Education, context, rights
Ellen Daly	IRISS	Images
Mick Eadie	VADS/Image AP project	Images
David Giaretta	DCC	Curation / preservation
Andy Heath	Independent consultant	Accessibility
Sarah Higgins	DCC	Curation / preservation
Linda Kerr	Intute	Discovery, collection management
Sheila MacNeill	JISC CETIS	Education
Tony Mathys	Edina	Geospatial
Lesa Ng	Intute	Discovery, collection management
Sharon Perry	JISC CETIS	Accessibility
Peter Rainger	Key2Access Ltd.	Accessibility
Mary Jane Steer	Intute	Discovery, collection management
Ian Watson	IRISS	Images
Grant Young	TASI	Images
Rowin Young	JISC CETIS	Assessment

Expert working group

This work was supported by an expert working group who commented on a draft outline of this report presented at a face to face meeting in Glasgow. Members of this group also commented on the work at other stages. The group comprised:

Julie Allinson	University of York (formerly of SWAP)
Gayle Calverley	Manchester University / Time-based media AP
Lorna M Campbell	JISC CETIS (convenor)
John Casey	JORUM
Phil Cross	Intute
Sarah Currier	Intrallect Ltd. / DC Education working group
Neil Fegen	JISC CETIS
Pete Johnston	Eduserv Foundation
Sheila MacNeill	JISC CETIS
Peter O'Hare	JORUM
Sharon Perry	JISC CETIS
John Robertson	JISC CETIS / Repositories Research Team
Rowin Young	JISC CETIS