International Journal on Digital Libraries (2006) 6(2): 139–147
DOI 10.1007/s00799-005-0131-2

REGULAR PAPER

William Reilly · **Robert Wolfe** · **MacKenzie Smith**

MIT's CWSpace project: packaging metadata for archiving educational content in DSpace

Published online: 20 January 2006

Abstract This paper describes work in progress on the research project CWSpace, sponsored by the MIT and Microsoft Research iCampus program, to investigate the metadata standards and protocols required to archive the course materials found in MIT's OpenCourseWare (OCW) into MIT's institutional repository DSpace. The project goal is "to harvest and digitally archive OCW learning objects, and make them available to learning management systems by using Web Services interfaces on top of DSpace." The larger vision is one of complex digital objects (CDOs) successfully interoperating amongst MIT's various learning management systems and learning object repositories, providing archival preservation and persistent identifiers for educational materials, as well as providing the means to richer shared discovery and dissemination mechanisms for those materials. The paper describes work to date on the analysis of the content packaging metadata standards METS (Metadata Encoding and Transmission Standard) and especially IMS-CP (IMS Global Learning Consortium, Content Packaging), and issues faced in the development and use of profiles, extensions, and external schema for these standards. Also addressed are the anticipated issues in the preparation of transformations from one standard to another, noting the importance of well-defined profiles to making that feasible. The paper also briefly touches on the DSpace development work that will be undertaken to provide new import and export functionalities, as the technical specifications for these will largely be determined by the packaging metadata profiles that are developed. Note that the degree of interoperability considered herein might be referred to as "first level," as this paper addresses the packaging metadata only, which in turn is the carrier or envelope for the descriptive (and other kinds of) metadata. It will no doubt be an even more challenging task to ensure interoperability at what might be referred to as the "second level," that of semantic metadata.

W. Reilly (✉) · R. Wolfe · M. Smith
Massachusetts Institute of Technology, Cambridge, MA 02139-4307, USA
E-mail: {wreilly, rwolfe, kenzie}@mit.edu

Keywords Content packaging · IMS-CP · METS · Archiving websites · Learning objects

1 Introduction

The growing corpus of educational materials published on MIT's OpenCourseWare (OCW) [1] web site is regarded as worth careful preservation, to help meet one of the principal requirements of the OCW program, that of ensuring the course materials "remain available to scholars and instructors for inspiration and reuse for the foreseeable future." To best accomplish this task, the MIT Libraries in concert with OCW are at work on a 2-year research project called CWSpace (sponsored by the MIT and Microsoft Research iCampus program), to investigate the metadata standards and protocols required to archive the course materials found in OCW into MIT's institutional repository DSpace [2]. Beyond the initial premise of archiving and preserving OCW course web sites entire, the further CWSpace project goal is "to harvest and digitally archive OCW learning objects, and make them available to learning management systems by using Web Services interfaces on top of DSpace." That is, the content should be not only archived but also refined to a greater degree of granularity than entire courses, and this newly re-packaged, more modular content should then be exposed to wider dissemination among other learning systems via networked interoperability (Web Services, and similar).

The key initial step in this work is to identify the optimal metadata standard for packaging these complex digital objects (the "learning objects," but also the OCW course web sites). This paper describes project work to date on the analysis of the content packaging metadata standards METS (Metadata Encoding and Transmission Standard) [1] and IMS-CP (IMS Global Learning Consortium, Content Packaging) [2], and issues faced in the development and use of profiles, extensions, and external schema for these standards. The goal is to establish a public profile named CWSpace (Open "CourseWare" (CW) [3] in DSpace (Space) [4]), for the use of content packages for learning materials with DSpace.

2 Future vision of interoperability

When considering the value generated in the creation of digital teaching and learning materials, it is an attractive prospect in the world of higher education to maximize the utility of that value by re-use. This sharing calls for interoperability of the materials as complex digital objects (CDOs) designed for seamless movement among the systems for course delivery and the repositories specially designed to support these so-called "learning objects" (LOs). The two domains of teaching and learning technologies and digital libraries and repositories are converging, as increasingly digitally created course materials in higher education are deemed worth preserving and sharing, via the services the digital archives domain has to offer.

For example, managers of learning management systems (LMSs, herein used for the terms "course delivery systems" and "virtual (or collaborative) learning environments" and similar) are experiencing the felt need for more dedicated archival services than they themselves can provide for the course instances they close out at the end of each semester. Course owners (faculty, generally) are interested to learn about the wider range of systems their course might be offered in, the potential for portability of courses, as well as the assurances of permanent storage for those offerings. Librarians and managers of repositories are aware, of course, of these needs in higher education for a more robust level of information management for the fast growing body of work being created as digital learning materials, and have begun to respond with a variety of initiatives. These include web-based consortia of LOs, many of which store some assets but also serve as URL pointers to materials stored elsewhere, with additional services such as peer review and other sorts of community-fostering annotation capabilities. Some examples are MERLOT [5], LOLA [6], GEM [7], CONNEXIONS [8], EDUSOURCE [9], and CAREO [10]. The development of institutional-based "learning object repositories" (LOR) which would contain, preserve, and make discoverable these organized packages of digital assets is largely an area of recent investigation and development. One starting point for a research project into this kind of development activity is the adoption (or adaptation) of an existing repository platform of the type known for use as an "institutional repository."

This is the essential reasoning behind this MIT research project to archive OCW in DSpace. Below, each of these MIT educational technology initiatives is described, with a particular view to the topic of CDOs, and how each of them will need to be further developed to accommodate the new vision of creating, managing, archiving, and interoperating with LOs.

It is important to note that this paper is about only one fairly fundamental aspect of what it is going to take to arrive at such a vision [11], namely, data interchange standards, in particular the packaging metadata standard. The paper does not address higher level concerns, including semantic metadata standards and uses in practice [12]; intellectual property issues; incentives to contribute to and consult LORs; and, argued by some [13] most important of all, the necessary theory of instructional design that should guide LO creation, description, and reuse, if we are to in fact achieve the real goal of all this activity in both domains, that of facilitating learning.

Finally, it is useful to note that another constituency arguably most directly involved in the creation of content for teaching and learning may in fact have requirements somewhat at odds with the typical services provided by a more heavyweight "institutional repository" solution, as applied to this relatively new task of managing learning materials. That is, the educational content authors and instructional designers, whose more granular level work product (e.g. stand-alone learning objects, and similar) is today usually contributed directly to existing LMSs and perhaps their supporting content management systems (CMSs). The needs of these contributors are more likely to place emphasis on functionality that supports versioning and updating, over the services for long-term preservation and for wider dissemination and access that an archival repository can provide. This need for more responsive and lighter weight management of teaching content still in development may prove to be a real challenge to the ready adoption of learning object and courseware archives, at least for those kinds of materials that are considered never quite finished or ready for "final publication."

3 CWSpace project, vis-à-vis complex digital objects

3.1 OCW content

The courseware material under consideration for this project (OpenCourseWare's course web sites) differs in a few interesting ways from the materials anticipated for (re-)use in the above-described future vision of interoperating LOs (Fig. 1).

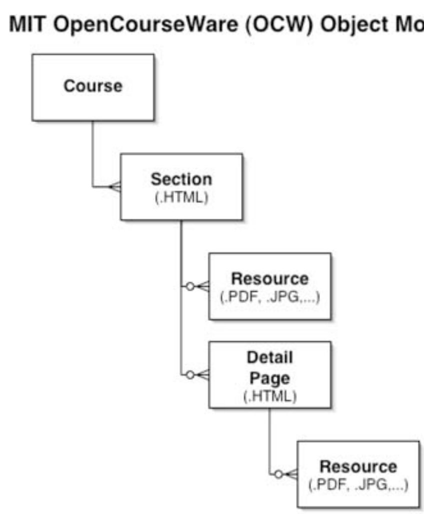

Fig. 1 MIT OpenCourseWare object model

To begin, the primary unit of processable record for initial OCW purposes has been the course, not a more granular LO. The lowest level of granularity in the OCW object model is a "Resource," which, being by definition a single file (e.g. a .PDF, or a .JPG, etc.) cannot be used to represent multiple file LOs. This, as we shall see, is the most challenging – and germane to the topic of CDOs – issue to be addressed.

Secondly, OCW is a courseware publication, not a learning management system, so the CWSpace research project's discoveries, proposed standards and profiles must take into account not only OCW requirements and use cases, but also, to the extent feasible, those of LMSs, both at MIT (Stellar and SloanSpace) and more widely (Sakai, Blackboard et al.). An OCW course is a statically served, unchanging web site, serving as the published record of a particular instance (semester offering) of a given course. No interactivity is found in the course publication, and no student authored content (discussion boards, etc.) is part of the course materials.

Finally, a third noteworthy distinction (disclaimer, really) concerning the publicly available OCW material, as compared to content in active LMS course offerings, is that all copyright has been cleared for OCW use, whereas a number of intellectual property rights and even privacy issues will come into play in a wider envisioned universe of exchangeable LOs that while technically interoperable may well have considerable hurdles to overcome regarding open access.

3.2 DSpace platform

As noted above, an institutional repository like DSpace does not provide a high degree of versioning capabilities and other "lighter weight" content management kinds of functionality. So for the mid-term future, discussion of the aptness of DSpace for service as a LOR primarily concerns the internal repository metadata record, and the import/export packaging formats to be supported.

DSpace is well established as an institutional repository, in which research reports, data sets, whitepapers, scholarly publications, images, and videos are archived and preserved. The Dublin Core set of metadata elements is used to catalog and describe this material, some of which had richer source metadata records (e.g. MARC library cataloging or VRA Core, etc.) that had to be crosswalked down to Dublin Core. LOs – teaching and learning materials that have outcomes, levels of difficulty, and other specialized pedagogical metadata – are usually described with the more complex IMS/IEEE Learning Object Metadata (LOM) [14]. In the current version of DSpace, this too would have to be crosswalked down to Dublin Core, while future versions of DSpace may permit a more integrated use of this kind of specialized metadata. (MIT's SIMILE project [15] is investigating Semantic Web technologies for this purpose.)

Import (or "ingest") and export functionality in DSpace is an area ripe for further development, especially as driven by the specifications of new compound digital object packaging standards and in particular the profiles developed for them. Therefore the CWSpace activity with METS and IMS-CP is of interest both for the requirements for ingest processing as well as determining the output export options that need to be supported. Processing will be architected upstream or downstream of the DSpace core functionality – as pre- or post-processing plug-in modules – and may include transformation services (e.g. XSLT or custom Java development) on submitted packages prior to ingest, or exported packages prior to delivery.

3.3 MIT LMSs: SloanSpace, Stellar (Sakai)

The local MIT LMSs have expressed support for the IMS-CP standard for packaging of both LOs and entire courses.

At the administrative level of handling entire courses, one data interchange use case currently being looked at most closely is one in which these systems would archive completed courses at the end of each semester to DSpace. Instead of devising *N* number of import and export ad hoc arrangements with other individual systems, these LMSs hope to take advantage of standards for packaging and protocols, so as to be able to go to DSpace as the LOR for putting and also getting entire courses.

Course authors manipulating course components (LOs) will use new Web Services-based functionality to permit searching, selecting, and retrieving content and metadata packages from LO repositories. Course authors would then typically use a specialized desktop tool (such as RELOAD [16], Giunti Packager [17], or the HarvestRoad Explorer [18]) for editing and authoring environment for LO content itself. The resulting packages would be uploaded from the desktop tool to the LMS. Optionally, an LO might be submitted independently to the DSpace LOR.

4 Learning objects

Defining just what is meant by the term "learning object" can be challenging. One of the more straightforward definitions for learning objects reads, "any digital resource that can be reused to support learning." [19] Even in this concise definition we see a variety of factors expressed (each of which can serve as the topic of its own lengthy discussion): being digital (vs. not); issues surrounding reusability; what exactly it means to "support" learning; and many debates about learning itself (pedagogical approaches, etc.).

For the CWSpace project focussed on archiving OCW materials in DSpace, the most relevant concept from that definition is the fundamental one of identification of what a "resource" is, precisely. The above definition, while helpful, still begs the question (as they all do) of how large or small that "digital resource" might usefully, or most appropriately, be. This is particularly the case with OCW material, which does not have a logical representation in its object model to identify an aggregation of files as a LO.

4.1 LOs were not expressly planned for

In planning for the task of publishing (eventually) all of MIT's course offerings under the umbrella of a single web-based Institute publication, OpenCourseWare faced a considerable challenge. In order to provide a cohesive presentation of what was a truly wide variety of course material types and highly divergent approaches to organizing those materials, the OCW object model devised to support course publishing requirements was purposely designed to be very straightforward and normalizing. Courses, Sections within Courses, and Resources linked to from Sections is the essence of the model.

Whether entire courses are usefully regarded as LOs may depend on one's point of view; whether sections of course material arranged by kind (all Problem Sets, for example) are good candidate LOs will depend on your use case; whether the resources of individual file .PDFs (or .JPGs, or other types) are usefully treated as LOs again will vary depending on content and on intended purpose for (re-)use.

The vision of making available MIT's courseware material via this publication has been largely course-centric, as this has best served most of the purposes to which OCW has been put: presenting MIT's courses to the World Wide Web, making courses available to translation partners, making courses available to educational partners, providing faculty with personal copies of the courses they have taught, and exchanging entire courses with the two learning management systems at MIT, Stellar and SloanSpace. The need to work with smaller parts of course offerings has not been a prominent use case to date. Plans are now in place for assessing the CMS system capabilities to permit the identification of candidate learning objects, their primary entrypoint files, the necessary supporting files and structure, and a metadata record descriptive of the learning object entire.

5 Content packaging

5.1 Candidate standards

The CWSpace project stands in the overlapping portion of a Venn diagram of two circles, one representing the world of Libraries and Repositories, the other the universe of Learning and Teaching Technology. Not surprisingly, each of these domains has created its own approach to a metadata standard for packaging CDOs.

In our project we have considered in turn the standard from the world of digital libraries, METS ("Metadata Encoding and Transmission Standard"), followed by investigation of the standard from the world of learning technology, at both the IMS Global Learning Consortium and the IEEE, IMS-CP ("Content Packaging"). It's important to note that a third circle (to date unexplored by this project) might be added to the Venn diagram, to represent the ISO standard from the Moving Pictures Experts Group, MPEG21-DIDL ("Multimedia Framework," and "Digital Item Declaration Language," http://xml.coverpages.org/mpeg21-didl.html). This has been generally designed for CDOs consisting of videos, audio tracks, images, etc., of interest especially to "the content, financial, communication, computer and consumer electronics sectors," [20] although at least one interesting case study comes out of the Los Alamos National Laboratories (LANL) [21] where they have selected MPEG21 for archival and repository purposes for their CDOs.

The use of METS for the capture of web site structure is still in relatively early days, with work at New York University [22] and elsewhere, largely under the auspices of the Center for Research Libraries [23], as well as the Library of Congress's NDIIPP (National Digital Information Infrastructure and Preservation Program) [24]. Our analysis of the comparatively consistent and straightforward information architecture of the OCW course web sites mapped well onto the METS approach for recording web sites, with its use of the <par> and <area> children elements to the <div>s that represent HTML pages, and a <fptr> child element to those <div>s that represent linked "Resources" (OCW term for the .PDF files and similar that contain most of the actual course content). These are effectively the leaf nodes.

IMS-CP is not known for use in the capture of web sites, but its content model is not fundamentally very different from that of METS. The IMS-CP manifest is essentially composed of two parts: "resources" (accounting for all files), and "organizations" (depicting the presentation, arrangement, "organization," of those resources) (Fig. 2).

In both METS and IMS-CP we discovered a feature we originally expected would be useful to our content: the flexibility to record more than one arrangement of content (METS' multiple <structMap>s; IMS-CP's multiple <organization>s). This would provide the means to represent not only the OCW course web site entire, but also the set of discrete learning objects we had hoped we would encounter in processing the course on ingest. In practice, as

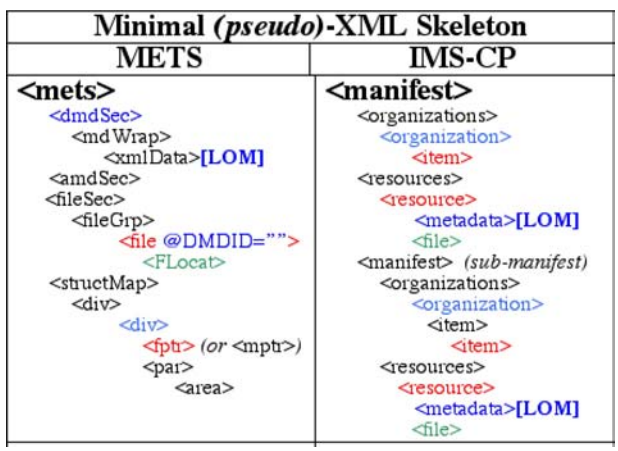

Fig. 2 Pseudo-XML indicating essential skeleton of elements used in METS and IMS-CP

we did not process learning objects, our IMS-CP manifest currently uses one organization only.

5.2 Selection rationale (IMS-CP for higher education)

Our initial bias, coming at this problem initially from the library and repository side of things, was to regard METS as the useful standard for packaging the OCW web sites, particularly as we learned more about the METS for web sites activity, and as we considered future development of the target repository platform, DSpace, which is considering the use of METS with its asset store.

But it was the wider consideration of other learning management systems at MIT that caused us to consider the alternative from the educational technology world as the more generally useful package description for our target audience: IMS-CP. Below are some detailed topics of our work with profiling our use of the IMS content package.

While it is an important goal to ultimately be able to transform the received package to METS, it should be noted that in fact the options remain open to modify DSpace to accept and ingest directly IMS-CP and/or METS. In any event, LMSs and other learning technology systems need only concern themselves with their familiar IMS-CP package when dealing with DSpace as an LOR.

5.3 Transform of IMS-CP to METS for LOR

The process of transforming from METS to IMS-CP (note this is the *reverse* of the CWSpace ingest problem) has been looked at closely at the University of California, Berkeley [25], with the conclusion that a transformation of basic examples appears feasible. However, in order for this to work in production, the key finding was that profiles would need to be established for both the METS input and the IMS-CP output. As regards the trip in the other direction, from IMS-CP to METS (what the CWSpace project needs, for ingest), the same principles no doubt apply.

In sum, as these packaging metadata standards are by design fairly open, some standardization of their use must be determined, via profiles, for the particular use of them with DSpace serving as an LOR.

6 Customizing IMS-CP for CWSpace

Customized use of a given schema (e.g. IMS-CP) for a particular purpose (e.g. use with DSpace) may entail either constraining or extending the schema (or a combination of both of these). This is accomplished by means of profiles (documentation), extensions (your own schemas), and external schema.

Profiles constrain the instance documents of a schema. A valid XML instance document that matches the profile will always be a valid document against the original schema. Profiles may be expressed in an XML Schema document (.xsd) or in a narrative document (e.g. MS-Word) typically used to depict the object model and to describe the intended constraints and "best practices" for usage.

Extensions represent the introduction of new elements and attributes, thereby "extending" the base schema of the standard. Extension schema are the mechanism for achieving the extension; they are written in a separate file of schema information (.xsd) the developer can maintain and publish at a permanent URL for others to access and use.

6.1 Profiles

Establishing a profile is useful for two reasons: (1) it publishes to others the profile and namespace for use of IMS-CP with DSpace, and (2) it creates the extensible mechanism for future additional constraint requirements. It makes the declaration that this is how to use IMS-CP if you are exchanging LOs or course web sites with DSpace. This is effectively a proposed profile for "courseware" ("CW") in DSpace ("Space"): hence the (lower-cased) namespace name of "cwspace."

6.2 Extensions

Systems exchanging LOs with DSpace will be required to support IMS-CP Level 1 compliance, as that guarantees preservation of extensions.

The details of how IMS-CP is extended for CWSpace are seen below. In essence, some additional information was deemed necessary to label the content package itself, and some information already available down in the content (and its metadata) was regarded as usefully repeated up in the package header, for ease of identification of its contents when working with the package.

6.3 External schema

The CWSpace profile for IMS-CP does not indicate any particular external schema, though the use of IMS/IEEE LOM is anticipated, and DSpace will have the appropriate crosswalk for transforming LOM to Dublin Core. However, other educational systems may have reason to employ other descriptive metadata (e.g. VRA Core for image collections), and subsequently other crosswalks would be developed for DSpace to handle those.

6.4 Issues encountered in profile development

6.4.1 Label information for content, for package

The content packaging specification provides very little in the way of recording identifying or labelling information for the package itself. This may be because in the majority of anticipated use cases for content interchange, the package itself might be regarded as a mere temporary wrapper, an envelope used one time to convey contents to a system, after which it becomes essentially disposable. Additionally, very little provision is made in the package header to reveal descriptive or other useful information found down within the contained content files and their metadata.

In practice, we have determined that for our CWSpace project each of these concerns might benefit from a few additional pieces of information in the package. For the labelling of the package itself, it is useful to capture a few new fields of information about its preparation, largely for purposes of quality control and similar.

Concerning the identifying information for the content within the package, it is anticipated that for both human readers and for software that consumes these packages a few key pieces of information brought up (duplicated) from out of the content and its metadata will make for easier, surer processing.

New information to present in content package header (see also the sample XML following):

- Package Labelling (cwspace: namespace)
  - Package Type
    - Initial enumeration of values [“Course”or “LO”].
      - (Other ideas: [“Section”, “LOs”, “Courses”])
  - Package Version
    - Likely a date-time stamp when the package was created (e.g. “2004-12-06T11:54:17”)
    - Note: Not strictly “new.” This can use the inherent manifest/@version attribute; no need for additional cwspace: element or attribute
  - Source System
    - System that wrote the IMS-CP manifest, created the package .ZIP file (e.g. “OCW”, “DSpace”, “Stellar”, “SloanSpace”)
  - Content Provider
    - System that published the content (e.g. “OCW”, “Stellar”, “SloanSpace”)
  - Checksum
    - Reserved; as yet undefined (e.g., for now, “string”)
      - One possible use: indication of number of submanifests (if any)
  - Profile
    - Name of profile (e.g. “CWSpace”)
  - Profile Version
    - (e.g. “1.0”)
- Content Labelling (lom: namespace) (extracted from LOM record for root object of package)
  - Title
    - Title of root object in package content.
      - For “Course” packages, course title, from lom:general/lom:title (e.g. “Technologies of Word 1450-2000”)
      - For “LO” packages, title of LO, also from lom:general/lom:title
  - Identifier
    - Used to record identifier number or code
      - For “Course” packages, Course Number, from lom:general/lom:catalog/lom:entry (e.g. “21H.418”)
      - For “LO” packages, an Identifier may need to be developed. Not available at present. To begin, may well be course-related.
  - Version
    - Used to indicate version of the content (not package)
      - For “Course” packages, from lom:lifecycle/lom:version (e.g. “Fall 2002”)
      - For “LO” packages, a Version may need to be developed. “Package Version” may be one facet of this LO Version.
  - Primary Contributor/Responsibility.
    - Used to record the name of one entity (person, usually preferred) who created or was responsible for the object
      - For “Course” packages, from lom:lifecycle/lom:contribute[lom:role=”Author”]/lom:entity (e.g. “Ravel, Jeffrey S.”)
      - For “LO” packages, recording a Contributor may need to be instituted in lom:lifecycle/lom:contribute [lom:role=”Author”]/lom:entity (e.g. “Ravel, Jeffrey S.”)
  - Date Contributed
    - Date this “Primary Contributor” submitted the object.
      - For “Course” packages, from lom:lifecycle/lom:contribute/lom:date (e.g. “2004-09-28”)
      - For “LO” packages, recording this Date may need to be insituted, in lom:lifecycle/lom:contribute/lom:date (e.g. “2004-12-16”)

Example XML snippet for instance of these elements and attributes:

```
<manifest cwspace:packageType=”Course” version=”2004-12-06T11:54:17”
  identifier=”OcwWeb_21H-418Technologies-of-Word-1450-2000Fall2002” >
<metadata>
  ...
   <cwspace:packageMetadata>
      <cwspace:sourceSystem>OCW</cwspace:sourceSystem>
      <cwspace:contentProvider>OCW</cwspace:contentProvider>
      <cwspace:checksum>string</cwspace:checksum>
      <cwspace:profile>CWSpace</cwspace:profile>
      <cwspace:profileVersion>0.1</cwspace:profileVersion>
   </cwspace:packageMetadata>
   <lom:general>
      <lom:title>Technologies of Word 1450-2000</lom:title>
      <lom:catalog>
         <lom:entry>21H.418</lom:entry>
      </lom:catalog>
   </lom:general>
   <lom:lifecycle>
      <lom:version>Fall 2002</lom:version>
      <lom:contribute>
      <lom:role>Author</lom:role>
      <lom:entity>Ravel, Jeffrey S.</lom:entity>
      <lom:date>
         <lom:dateTime>2004-09-28</lom:dateTime>
      </lom:date>
   </lom:contribute>
  </lom:lifecycle>
</metadata>
</manifest>
```

6.4.2 IMS-CP manifest organization: Two approaches

In addition to the “resources” and “organizations” noted above, by means of its “submanifest,” the IMS-CP also provides a way to package up multiple CDOs for easy aggregation and disaggregation.

We at first intuited that IMS-CP’s sub-manifest mechanism would be useful to our processing of an entire OCW course web site, because as we process the site our plan was to at the same time look for candidate disaggregatable objects. These would include LOs – which can potentially be nested (e.g., a Java applet inside an HTML textbook inside a course) – as well as possibly disaggregating OCW course "Sections."

There were, we felt, a number of advantages to this approach. These included making it easier for any process consuming the course to identify the candidate disaggregatable objects. Likewise the subsequent authoring of the new manifest required by the disaggregated object would be simplified, as the manifest would already be written and the parent manifest would not need to be reparsed to assemble the necessary information for a disaggregatable collection of files. Additionally, the availability of a full <manifest> header for additional, even custom, metadata about the disaggregatable object was seen as useful. In particular this included some OCW-specific needs like capturing the date-time stamp of the latest version of a section of a course.

We discovered, however, that a couple restrictions inherent in using sub-manifests prevented our adopting that approach, and we chose instead a single organization with the set of lighter weight nested <item> elements to record everything found in the web site. We also elected to process separate IMS-CP packages for individual LOs.

6.4.3 XML ID attributes

The first restriction concerned XML ID attributes, with their requirement of being unique within an entire XML document. A submanifest, representing a disaggregatable (and therefore stand-alone) object, must contain a local reference to all files required for its operation. When certain resources need to appear in multiple submanifests, each time they must be assigned unique IDs, as if they were different objects.

For example, in Table 1 we see a simple, minimal set of files and links within them to demonstrate the issues involved. In essence, a single .PDF (LectureNote01) is pointed to from two pages: a page listing all Lecture Notes, and the Calendar page, providing a temporal-oriented access to the same material.

In Table 2 we see two different approaches to authoring a content package for that minimal set of files. So, if multiple Items in the course (e.g., Table 2, Column B. (No Submanifests), Lines 6 & 12: @IDREF = "LectureNote01.pdf") reference a given Resource (e.g., Table 2, Column B., Line 28: @ID = "LectureNote01.pdf"), there is no problem.

But with the use of submanifests, if those Items happen to be located within separate, sibling submanifests (e.g. Table 2, Column A. (Uses Submanifests), Lines 16 & 42: @IDREF = "cal_LectureNote01.pdf" and @IDREF = "lns_LectureNote01.pdf" respectively) – which would be common enough in OCW's content – then the Resource would have to appear in both submanifests, each appearance being assigned its own *unique* @ID. These would need to be distinguished via some scheme of affixing prefixes or similar (e.g. Table 2, Column A., Lines 32 & 58: @ID = "cal_LectureNote01.pdf" and @ID = "lns_LectureNote01.pdf" respectively) .

The difficulty this was likely to create (different @IDs for the exact same resource (e.g. Table 2, Column A., Lines 33 & 59: <file href = "LectureNotes/LectureNote01.pdf"/> for *both*) was viewed as likely to create more trouble than the benefits using submanifests might bring.

Table 1 Minimal example set of files and links

Minimal Set of Files, Links
/Calendar/index.htm
<a href to LectureNote01.pdf>
<img src to spacer.gif>
<link href to core.css>
/LectureNotes/index.htm
<a href to LectureNote01.pdf>
<img src to spacer.gif>
<link href to core.css>
/LectureNotes/LectureNote01.pdf
/images/spacer.gif
/styles/core.css

6.4.4 'Dependency' refactoring

Another consideration against submanifests, fairly important to our web site-centric main use case, is that they cause you to lose the benefits of the IMS-CP <dependency> mechanism. This permits simple refactoring of repeating material, so it can be represented by a named file grouping.

For example, the set of some dozen files that are required in the rendering of almost every HTML page can be grouped together into a set called "webrenderrequirements" (e.g. Table 2, Column B (No Submanifests)., Line 19–23). That named set can now be referenced with a single line under the HTML page's Resource entry (e.g. Table 2, Column B., Lines 26 & 33), vs. repeating the dozen or so lines needed to explicitly list each of the 12 files needed for web rendering. But when this same approach to refactoring is assessed in the context of using sub-manifests, it is largely lost (e.g. Table 2, Column A. (Uses Submanifests), Lines 27–31 & 53–57 repeat identical listings of files (gifs, css, etc.)).

In sum, because each sub-manifest must by definition be prepared to be disaggregated and therefore ready to stand on its own, it must repeat within its local scope any supporting information that might have been able to be held only globally, and referenced via a 'dependency' (as in fact we saw in reviewing the "Column B" (no sub-manifests) approach just above).

Generally, we feel that the use of the IMS Content Package for entire course web sites is pushing the limits of what it was designed for, especially in terms of the use of its otherwise convenient and sensible sub-manifest mechanism. A more conservative use of its simple nested <item> elements should prove versatile enough for us to process course web sites. When smaller learning objects are packaged for separate submission we anticipate the same content package

Table 2 Minimal set of files, as recorded in manifests (two approaches)

Column A. Using Sub-Manifests

```
<manifest>
  <organizations>
    <organization>
      <item identifier="item_CalendarSection"
        identifierref="CalendarSection"></item>
      <item identifier="item_LectureNotesSection"
        identifierref="LectureNotesSection"></item>
    </organization>
  </organizations>
  <resources/>
  <manifest identifier="CalendarSection">
    <organizations>
    <organization>
    <item identifier="item_Calendar.index.htm"
        identifierref="Calendar.index.htm">
      <item identifier="item_cal_LectureNote01.pdf"
          identifierref="cal_LectureNote01.pdf">
    </item>
    </item>
  </organization>
  </organizations>
  <resources>
    <resource identifier="Calendar.index.htm">
     <file href="Calendar/index.htm"/>
    <dependency
identifierref="cal_webrenderrequirements"> </dependency>
    </resource>
    <resource
identifier="cal_webrenderrequirements">
    <file href="styles/core.css"/>
    <file href="images/spacer.gif"/>
    <!–{plus 10 more similar files}–>
    </resource>
  <resource identifier="cal_LectureNote01.pdf">
     <file href="LectureNotes/
LectureNote01.pdf"/>
  </resource>
  </resources>
  </manifest>
  <manifest identifier="LectureNotesSection">
  <organizations>
  <organization>
    <item identifier="item_LectureNotes.index.htm"
        identifierref="LectureNotes.index.htm">
      <item identifier="item_lns_LectureNote01.pdf"
        identifierref="lns_LectureNote01.pdf">
      </item>
    </item>
  </organization>
  </organizations>
  <resources>
  <resource identifier="LectureNotes.index.htm">
    <file href="LectureNotes/index.htm"/>
    <dependency
identifierref="lns_webrenderrequirements"> </dependency>
  </resource>
  <resource identifier="lns_webrenderrequirements">
  <file href="styles/core.css"/>
  <file href="images/spacer.gif"/>
  <!–{plus 10 more similar files}–>
  </resource>
  <resource identifier="lns_LectureNote01.pdf">
  <file href="LectureNotes/ LectureNote01.pdf"/>
  </resource>
  </resources>
  </manifest>
</manifest>
```

Column B. Using Nested Items (no sub-manifests)

```
<manifest>
  <organizations>
    <organization>
      <item identifier="item_Calendar.index.htm"
          identifierref="Calendar.index.htm">
       <item identifier="item_LectureNote01.pdf"
         identifierref="LectureNote01.pdf">
    </item>
  </item>
    <item identifier="item_LectureNotes.index.htm"
        identifierref="LectureNotes.index.htm">
      <item identifier="item_LectureNote01.pdf"
      identifierref="LectureNote01.pdf">
       </item>
     </item>
     </organization>
  </organizations>
  <resources>
     <resource
identifier="webrenderrequirements">
    <file href="styles/core.css"/>
    <file href="images/spacer.gif"/>
      <!–{plus 10 more similar files}–>
  </resource>
  <resource identifier="Calendar.index.htm">
     <file href="Calendar/index.htm"/>
     <dependency
identifierref="webrenderrequirements"> </dependency>
</resource>
   <resource identifier="LectureNote01.pdf">
  <file href="LectureNotes/
LectureNote01.pdf"/>
    </resource>
  <resource identifier="LectureNotes.index.htm">
      <file href="LectureNotes/index.htm"/>
     <dependency
identifierref="webrenderrequirements"> </dependency>
    </resource>
   </resources>
</manifest>
```

design principles outlined here will provide the packaging metadata solution needed.

One important distinction to note in comparing this work done to date on packaging entire course web sites, as contrasted with the anticipated future work of packaging and archiving more complex and dynamic classes of learning objects, is that it may prove unwieldy to expect that the archival platform itself be able to also render those more complex materials. Static web sites require little more than a web server to provide a reasonably similar rendition of themselves; DSpace and other archival repositories are able to provide this. But the potential complexity of the software requirements to fully present some teaching and learning materials may call for functionality beyond what the archive can or perhaps ought to provide. In these cases the use of the archival platform will need to remain closer to its core purpose: storage, preservation, access. It will need to yield the "front-end" ground to those applications better suited for rendering these complex objects. Development permitting the two to integrate services will permit archival repositories to continue to deliver considerable value to the long-term management as well as the regular use of digital teaching and learning materials.

Acknowledgements The CWSpace project gratefully acknowledges the support of the MIT and Microsoft Research alliance program, MIT iCampus [26]. The project also appreciates the cooperation of staff from MIT OpenCourseWare, MIT Stellar [27], and MIT SloanSpace [28].

References

1. METS http://www.loc.gov/standards/mets/ See also "METS Structure Diagrammed" http://sunsite3.berkeley.edu/mets/diagram/ (2005)
2. IMS-CP IMS Global Learning Consortium, Content Packaging Best Practice Guide http://www.imsglobal.org/content/packaging/cpv1p1p4/imscp_bestv1p1p4.html
3. MIT OpenCourseWare (OCW). http://ocw.mit.edu (2005)
4. DSpace http://www.dspace.org (2005)
5. MERLOT http://www.merlot.org/Home.po (2005)
6. LOLA http://www.lolaexchange.org (2005)
7. GEM http://www.thegateway.org/ (2005)
8. CONNEXIONS http://cnx.rice.edu/ (2005)
9. EDUSOURCE http://www.edusource.ca/english/home_eng.html (2005)
10. CAREO http://www.careo.org/ (2005)
11. Yee, R.: Semantic Interoperability Problem. (2003) Accessed August 22, 2005, from the World Wide Web: http://raymondyee.net/ wiki/SemanticInteroperabilityProblem
12. Godby, C.J.: What Do Application Profiles Reveal about the Learning Object Metadata Standard? (2004) Accessed August 22, 2005 from the World Wide Web: http://www.ariadne.ac.uk/issue41/godby/
13. Wiley, D.A.: Connecting learning objects to instructional design theory: A definition, a metaphor, and a taxonomy. In: Wiley, D.A. (ed.), The Instructional Use of Learning Objects, p. 29 (2000) (online version). Retrieved December 6, 2004, from the World Wide Web: http://reusability.org/read/chapters/wiley.doc
14. IMS Global Learning Consortium, Learning Object Metadata (LOM).http://www.imsglobal.org/metadata/mdv1p3pd/imsmd_bestv1p3pd.html; IEEE WG12: Learning Object Metadata http://ltsc.ieee.org/wg12/ (2004)
15. MIT SIMILE project (Semantic Interoperability of Metadata and Information in unLike Environments). http://simile.mit.edu (2005)
16. RELOAD (Reusable Learning Object Authoring and Delivery) http://www.reload.ac.uk/ (2005)
17. Giunti eXact Packager http://www.learnexact.com/l1.php?click = 4 (2004)
18. HarvestRoad Explorer http://www.harvestroad.com/products/hive-modules.cfm (2005)
19. Wiley, D.A.: (2000) (op. cit.) p. 7
20. MPEG21-DIDL "Multimedia Framework" and "Digital Item Declaration Language," http://xml.coverpages.org/mpeg21-didl.html (2004)
21. Bekaert, J.: Using MPEG-21 DIDL to Represent Complex Digital Objects in the Los Alamos National Laboratory Digital Library. (2003) Accessed December 11, 2004 from the World Wide Web: http://dlib.org/dlib/november03/bekaert/11bekaert.html
22. New York University, Digital Library Team http://library.nyu.edu/diglib/index.html (2005)
23. Center for Research Libraries http://www.crl.uchicago.edu/ (2005)
24. National Digital Information Infrastructure and Preservation Program (NDIIPP) http://www.digitalpreservation.gov/ (2005)
25. Yee, R., Beaubien, R.: A preliminary crosswalk from METS to IMS content packaging. (2003) Accessed August 22, 2005 from the World Wide Web: http://iu.berkeley.edu/rdhyee/Filer/filetree/2004/yee_beaubien_mets_to_imscp_paper.pdf
26. MIT iCampus http://icampus.mit.edu/ (2005)
27. MIT Stellar http://stellar.mit.edu (2005)
28. MIT SloanSpace http://sloanspace.mit.edu (2005)