Content Packaging Standards for complex objects (Some)

Mahendra Mahey Herbert van de Sompel



- METS Metadata Encoding Transmission Standard – MM – Library of Congress
- MPEG 21 Moving Picture Experts Group - HS – Multimedia Industry
- IMS CP e-learning MM
- All XML based
- CCSDS Panel 2 (not covered today) –
 NASA Space Science
- Feedback from experts group MM



Objectives

- Understand a little about each packaging standard / specification
- Recognise the basic structure
- What they are used for in the 'real' world
- Brief look at strengths and weaknesses of each



Compound and complex digital objects

- Aggregations of related digital objects gathered together to form a logical whole.
- The relationship may be purely structural (e.g. a book and its chapters)
- Complexity is added when we begin to think beyond the structural,
- to a richer set of relationships between digital objects
- and relationships with other kinds of resources (people, organisations, concepts, events etc.)
- Metadata and/or content packaging help us to express structure and relationships





- Metadata Encoding and Transmission Standard (METS)
- Building on Making of America II (MOA) project
 - Texts decaying
 - Structural metadata needed for digital version of book with images
 - Technical metadata needed for digitisation process



- Funded by Digital Library Federation (DLF)
- Principal Author Jerome McDonough
- Library of CongressNetwork
 Development and MARC Standards
 Office (NDMSO) is Maintenance Agency
 (http://www.loc.gov/standards/mets/)
- The METS Editorial Board is responsible for schema content



What is METS?

- METS is an XML Schema designed to create XML document instances expressing:
 - the hierarchical structure of digital library objects
 - the names and locations of the files that comprise those objects
 - the associated metadata.



METS Characteristics

- Extensible
- Modular
- Open standard
- Non-proprietary
- Developed by the library community



METS Document

- 5 Sections
 - Descriptive Metadata
 - Administrative Metadata
 - File Groups
 - Structural Map (Heart of METS Doc)
 - Behaviour

– DAFSB



Descriptive Metadata



Descriptive Metadata Section ">dmdSec>

- Consists of one or more <dmdSec>
- Each <dmdSec> has a unique internal ID
 - <dmdSec ID='dmd001'>
 - </dmdSec>
- Used in structural map to link to particular division of document hierarchy to particular <dmdSec> element
- Allows specific sections of descriptive metadata to be linked to specific parts of the digital object
- Each <dmdSec> can contain pointers to:
 - external metadata <mdRef></mdRef>
 - internal metadata <mdWrap></mdWrap></md/>



Descriptive Metadata Section <mdRef> External Descriptive Metadata

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



Descriptive Metadata Section <mdRef> LOCTYPE

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

- LOCTYPE type of locator
- "URN", "URL ", "PURL","HANDLE","DOI","OTHER"



Descriptive Metadata Section <mdRef> MIMETYPE

- <mdRef LOCTYPE="URN"
 MIMETYPE="application/xml" MDTYPE="EAD"
 LABEL="Berol Collection Finding Aid">urn:x-nyu:fales1735
 </mdRef>
 </dmdSec>
- MIMETYPE MIMETYPE for external descriptive metadata



Descriptive Metadata Section <mdRef> MDTYPE

- <mdRef LOCTYPE="URN"
 MIMETYPE="application/xml" MDTYPE="EAD"
 LABEL="Berol Collection Finding Aid">urn:x-nyu:fales1735
 </mdRef>
 </dmdSec>
- MDTYPE MDTYPE what form of metadata being referenced
- MARC, EAD, VRA (VRA Core), DC (Dublin Core), TEIHDR (TEI Header), DDI (Data Documentation Initiative),
 NISOIMG (NISO Technical Metadata for Digital Still Images),
 LC-AV (Library of Congress Audiovisual Metadata),
 FGDC (Federal Geographic Data Committee Metadata Standard [FGDC-STD-001-1998]), OTHER



Descriptive Metadata Section <mdRef> LABEL

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

 LABEL – describing metadata for viewing METS document, e.g. in TOC of METS document



- Provides wrapper around metadata embedded within METS Document
- Can be
 - XML encoded METADATA
 With the encoding belonging to a namespace other than the METS document namespace
 - 2. Any arbitrary or textual form provided that the metadata is Base64 encoded and wrapped in



Descriptive Metadata Section <mdWrap> Internal Descriptive Metadata

XML encoded METADATA



 Here we see the XML encoding is DC, METS does not come with an own metadata schema, but enables the plug in of different extensions schemas

Arbitrary binary or textual form (base 64)



 Here we see the base 64 data is contained in the <binData></binData> tags

Administrative Metadata



- Contains admin metadata for
 - Files comprising digital object
 - Information pertaining to original source object used to create digital object



4 parts

- 1. Technical Metadata
 - Information regarding files' creation, format and use characteristics
- 2. Intellectual Property Rights Metadata
 - Copyright and licensing information
- 3. Source Metadata
 - Descriptive and administrative metadata regarding analogue source from which the digital object derives

4. Digital Provenance Metadata

 Source/destination relationships between files, including master / derivative, migrations/transformations (original digitisation and current incarnation)



- Elements follow same model as <dmdSec>
- <mdRef> to external metadata
- <mdWrap> embedding metadata within document
- Or both



- All <amdSec> elements must carry a unique ID attribute
- Elements in a structural map or file elements may be linked to the appropriate <amdSec> elements describing them



- A <file> element within a <fileGrp> might then identify this administrative metadata as pertaining to the file it identifies by using an ADMID attribute to point to this <amdSec> element:
- <file ID="FILE001" ADMID="AMD001">
 <FLocat</p>
 LOCTYPE="URL">http://dlib.nyu.edu/press/testimg.tif</FLocat</p>
 </file>



File Groups



File Groups <fileGrp>

- <fileGrp> lists all files which comprise single version of digital object
- Consists of one or more <fileGrp> each with its own unique internal ID
- There may be separate <fileGrp> elements for thumbanails, master archival images, pdf etc



File Groups <fileGrp> Example

<fileGrp>

```
<fileGrp ID="VERS1">
    <file ID="FILE001" MIMETYPE="application/xml" SIZE="257537"</pre>
    CREATED="2001-06-10">
      <FLocat LOCTYPE="URL">
        http://dlib.nyu.edu/tamwag/beame.xml
      </FLocat>
    </file>
                                                      XML encoded
  </fileGrp>
  <fileGrp ID="VERS2">
    <file ID="FILE002" MIMETYPE="audio/wav" SIZE="64232836"</pre>
    CREATED="2001-05-17" GROUPID="AUDIO1">
      <FLocat LOCTYPE="URL">
        http://dlib.nyu.edu/tamwag/beame.wav
      </FLocat>
    </file>
                                                      WAV format
  </fileGrp>
  <fileGrp ID="VERS3" VERSDATE="2001-05-18">
    <file ID="FILE003" MIMETYPE="audio/mpeg" SIZE="8238866"
CREATED="2001-05-18" GROUPID="AUDIO1">
      <FLocat LOCTYPE="URL">
         http://dlib.nyu.edu/tamwag/beame.mp3
      </FLocat>
    </file>
                                                      MP3 format
  </fileGrp>
</fileGrp>
```



File Groups <fileGrp> <FLocat> and <FContent>

- <file> elements may possess <FContent> rather than <FLocat>
- <FContent> is used to embed the actual contents of the file (base 64)
- <FContent> has CHECKSUM to indicate an MD5 checksum value for the file



Structural Maps



Structural Map <structMap > (Mandatory)

- Hierarchical structure which can be presented to users for navigation. It can be logical (i.e. to digital object) or relate to physical object itself
- Uses nested <div> division
- Each <div> carries info about what kind of <div> it is
- Structural Link between two div elements from different <structMap>

<mets:structLink>

<mets:smLink xlink:from="div1" xlink:to="div2">

May contain multiple METS pointers <mptr> and file pointer <fptr> elements



Structural Map <structMap > Pointers

- <mptr> multiple METS pointer
- Specify separate METS documents as containing relevant file information for <div> containing them
- Keeps METS file small
- fptr> file pointer elements specify files
 (or locations within files) within the
 current METS document's <filegrp>
 section that corresponds to portion in
 hierarchy represented by current <div>



Structural Map <structMap> (Very simple)

```
<structMap TYPE="logical">
 <div ID="div1" LABEL="Oral History: Mayor Abraham Beame" TYPE="oral history">
   <div ID="div1.1" LABEL="Interviewer Introduction" ORDER="1">
      <fptr FILEID="FILE001">
        <area FILEID="FILE001" BEGIN="INTVWBG" END="INTVWND" BETYPE="IDREF" />
      <fptr FILEID="FILE002">
        <area FILEID="FILE002" BEGIN="00:00:00" END="00:01:47" BETYPE="TIME" />
      <fptr FILEID="FILE003">
        <area FILEID="FILE003" BEGIN="00:00:00" END="00:01:47" BETYPE="TIME" />
   </div>
   <div ID="div1.2" LABEL="Family History" ORDER="2">
      <fptr FILEID="FILE001">
        <area FILEID="FILE001" BEGIN="FHBG" END="FHND" BETYPE="IDREF" />
      </fptr>
      <fptr FILEID="FILE002">
        <area FILEID="FILE002" BEGIN="00:01:48" END="00:06:17" BETYPE="TIME" />
      </fptr>
      <fptr FILEID="FILE003">
        <area FIL FID="FIL F003" BEGIN="00:01:48" FND="00:06:17" BETYPE="TIME" />
     </fptr>
   </div>
   <div ID="div1.3" LABEL="Introduction to Teachers' Union" ORDER="3">
      <fptr FILEID="FILE001">
        <area FILEID="FILE001" BEGIN="TUBG" END="TUND" BETYPE="IDREF" />
      </fptr>
      <fptr FILEID="FILE002">
        <area FILEID="FILE002" BEGIN="00:06:18" END="00:10:03" BETYPE="TIME" />
      </fptr>
      <fptr FILEID="FILE003">
        <area FILEID="FILE003" BEGIN="00:06:18" END="00:10:03" BETYPE="TIME" />
      </fptr>
   </div>
 </div>
</structMap>
```



Behavior



Behavior Section < METS: behavior Sec >

- Associate executable behaviours with content in the METS object
- Has interface definition element abstract definition of behaviour
- Behavior Mechanism module of executable code that implements and runs behaviors defined by the interface definition
- Can be implemented as linkages to distributed web services



Behavior Section < METS:behaviorSec > Example of Watermarking behavior

- <METS:behaviorSec ID="DISS1.0" STRUCTID="S1" BTYPE="uva-bdef-image-w:101" CREATED="2002-05-25T08:32:00" LABEL="Watermark Behaviors" GROUPID="DISS1" ADMID="AUDREC1" STATUS="A">
- <METS:interfaceDef LABEL="Photo Watermark Behavior Definition" LOCTYPE="URN" xlink:href="uva-bdef-image-w:101"/>
- <METS:mechanism LABEL="Watermarking Behavior Mechanism for Images" LOCTYPE="URN" xlink:href="uva-bmech-image-w:112"/>
- </METS:behaviorSec>



METS Header

 METS header contains information about the METS object (METS file), NOT about the content



METS Profiles

METS Profile describes the usage of METS for a special scenario:

- what extension schemas are used?
- what authority files?
- usage of attributes and elements

METS-profile schema available; profile is an XML file, which is not machine readable.

"registry" on METS website available - many



METS Resources!

- Family of XML data standards: METS, MODS, MIX, PREMIS, TEI, and EAD
- METS Implementations: LC, OCLC, RLG, California Digital Library, Harvard, Princeton, National Library of Portugal, National Library of Wales, University of Indiana, Stanford, New York University, University of Göttingen, Oxford University, etc etc...
- METS Software Tools: Harvard METS Toolkit, Harvard DRS METS Archive Tool (Dmart) for Audio Deposit, CDL 7train METS Generation Tool, MEX Authoring Tools (Das Bundesarchiv), ContentE (Biblioteca Nacional Digital, Portugal), METS Navigator (Indiana University Digital Library Program) ResCarta Metadata Creation Tool (ResCarta Foundation) etc
- METS listserv: over 500 subscribers



METS Resources

http://www.loc.gov/mets



IMS Content Packaging



What is IMS Content Packaging?

- XML based packaging format developed by IMS (International Metadata Standard) for Global Learning Consortium)
- Origins around learning materials (transport and reuse) but extended into other areas such learner information.
- Aggregation of resources, structure with (rich) description
- Application profiles: SCORM/SCORM2004, IMS Common Cartridge



Background and overview

- First released in 2001
- V.1.1.4 current public release
- V.1.2 in development



Current status of v1.2

- Public draft 1 released early 06
- IMS internal draft 2 documentation complete by early March 07 and testing
- Full public release early 2008

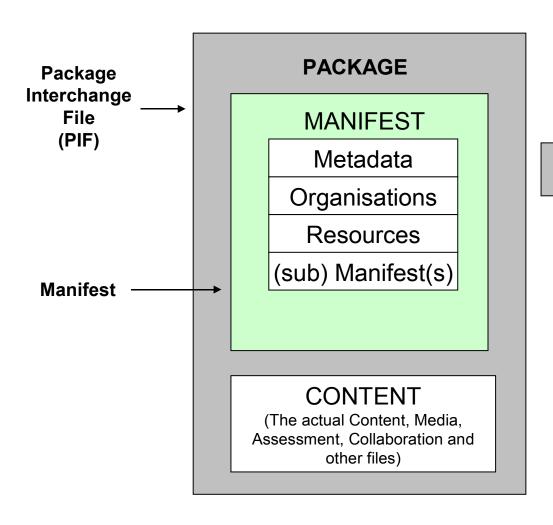


What constitutes an IMS CP?

- Manifest file (xml) with 3 main sections:
- Metadata (description)
- Organizations (structure)
- Resources
- (submanifests)
- All contained within PIF (packaging interchange file)
 - commonly a zip file



What constitutes an IMS CP?



External Packages...



CP 1.2 Conceptual Model

LOGICAL PACKAGE

Bound Interchange Package e.g. PIF, files on CD...

INTERCHANGE PACKAGE

MANIFEST

Metadata

Organisations

Resources

Child Manifests

FILES

e.g. media, assessment, collaboration, control files

External Interchange Packages

External Manifests

External Metadata

External Files



Simple stand alone package

- Simple stand alone package all resources in interchange (zip) file, usually one level of manifest no child manifests.
- Package supported by Virtual Learning
 Environments and Learning Object Repositories
- SCORM packages example of this stand alone zip, but they need to package level metadata in a separate file which is linked to and from the manifest. SCORM also extends support to tracking and sequencing of content.



Bare Manifest

- Bare Manifest doesn't contain all resources, instead references links in a known repository.
- Can be interchanged either using just the manifest file or a zip containing only the manifest.
- Advantages:
 - Same resources can be re-used many times
 - Resources can be updated at any time
 - Resources can be tracked (publisher/repository owner)
 - Access can be controlled
- Disadvantage:
 - Consuming system needs reliable internet connection



Composite or meta-package

- Composite or meta-package one step on from bare manifest, links to whole packages not just content assets.
- Most likely linking to a well defined structured set of self contained packages e.g a course.
- Advantages:
 - moves notions of reusability of content forward
- Disadvantages:
 - quite a new concept, so not all systems may be able to provide adequate functionality.



Archive Package

- Archive package gathers a number of resources and captures their basic structure (not intended to use with learners).
- Typically used to exchange 'raw' content which can be made into a complete package at a later stage.



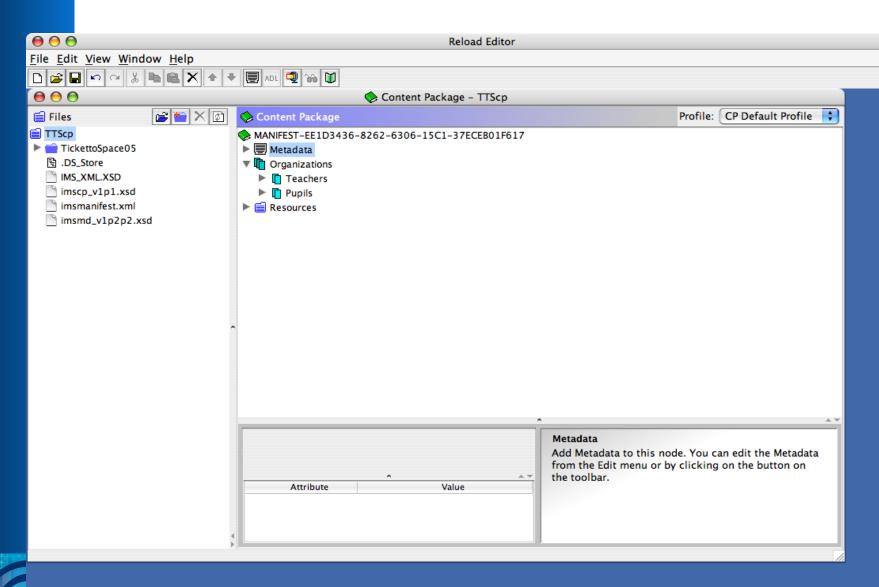
Specialised package

- Specialised package doesn't conform to common type but are used to convey specific types of content,
- Most of these packages are combinations of packages with other IMS specification content e.g. ePortfolio, QTI, Learning Design.
- CP really just providing a convenient aggregation for different XML data descriptions.

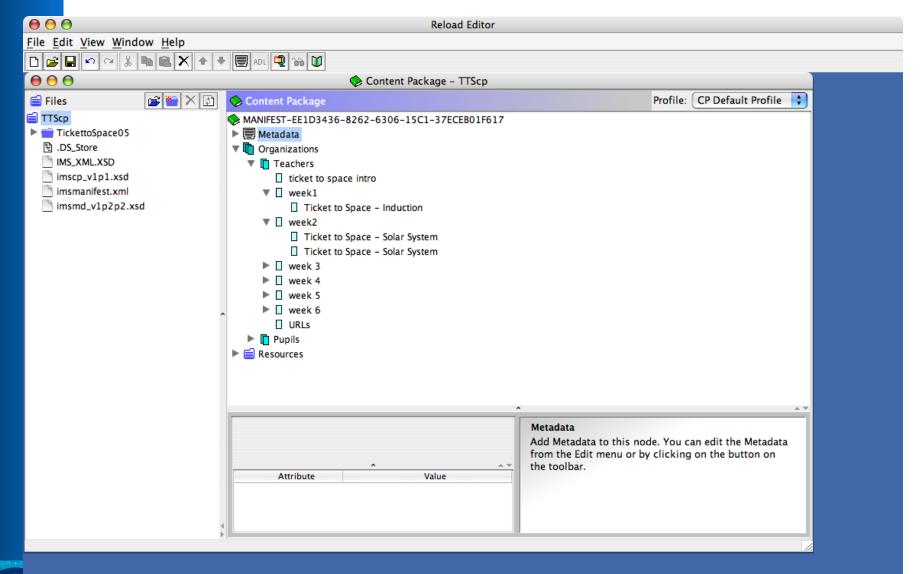


EXAMPLE standalone package

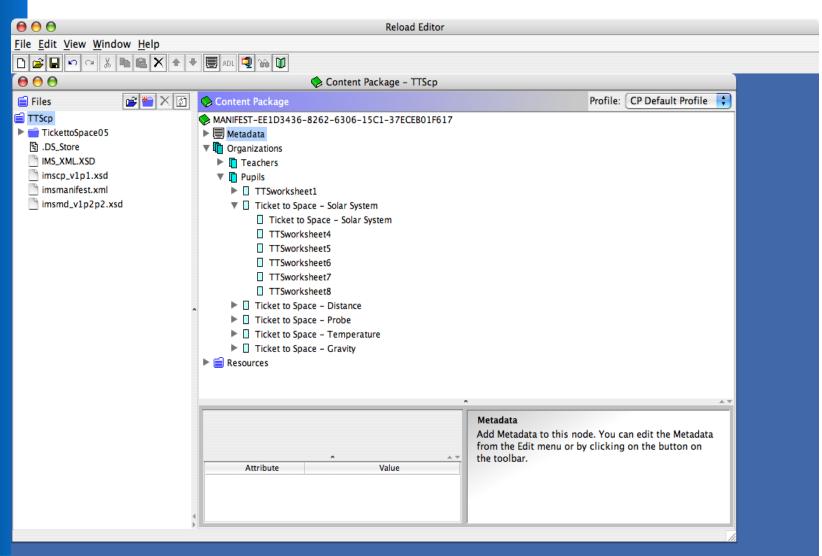




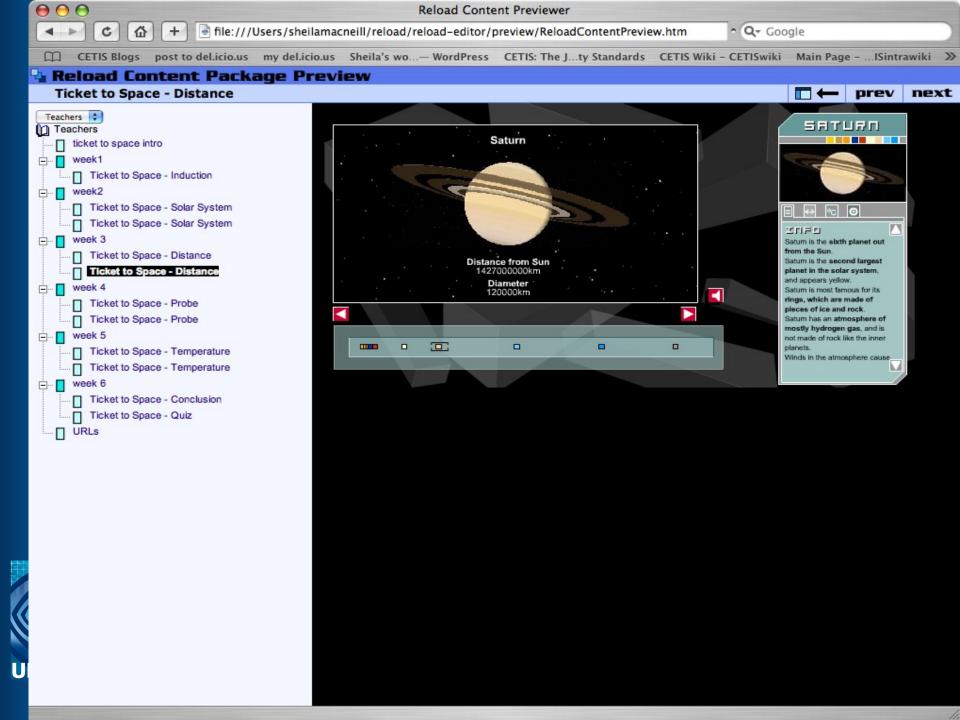


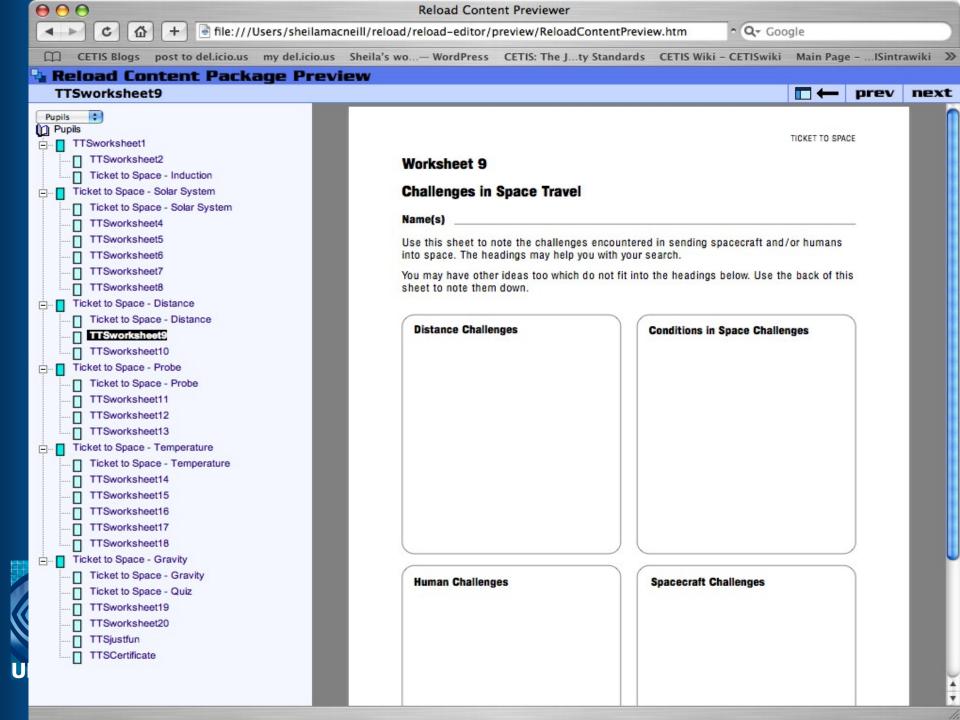












CP V.1.2 - updates

- General clarification of information model and binding
- Providing alternative resources with different accessibility characteristics
- Extending resource vocabulary
 - used to be just 'webcontent' and scattered IMS specific types (QTI items etc.)
 - now is an extensible vocabulary in a VDEX file (can be extended by anyone)
- Extending organization vocabulary
 - used to be just 'hierarchical' now extensible (e.g. TopicMaps, some METS integration)
- New mechanism to refer to remote manifest structures (ipointer)



IMS Content Packaging

- IMS Content Packaging : http:// www.imsglobal.org/content/packaging/index.html
- IMS Common Cartridge : http:// www.imsglobal.org/commoncartridge.html
- ADL/SCORM : http:// www.adlnet.gov/scorm/index.cfm
- JISC CETIS briefing papers: http:// zope.cetis.ac.uk/static/briefings.html
- RELOAD: http://www.reload.ac.uk



Meeting in February 2007, Bristol

- Invite experts to talk about standards
- Discuss strengths and weaknesses of each standard



What CP to use?

- The purpose the standard was originally used for gives us clues as to who uses the standard and what for
- Usually a good indicator of how the CP is used in the real world



METS (Strengths)

- Closely tuned to community concerns
- Managed standard at global level. Profiles reviewed by editorial board.
- Can have many file pointers, either parallel or sequential.
- Workflow management tools
- Pointers to files or into files
- Extensible simple to create profile., Good for preservation
- Flexible but concrete
- High take up with institutions and originating from library world
- Ability to incorporate other metadata+ digital objects themselves
- Useful as a preservation mapper
- Very precise
- Relatively simple. Deals with simple ordered tree structural relations.
- Includes and defines different types of M/D
- METS very suitable format for constructing submission information packages that will result in rich and well designed data objects within Fedora



METS (Weaknesses)

- Too tied to XML and tree structures may not be able to handle other data structures, lack of abstract model
- Not based on Abstract Model?
- Does not use GURIs
- Libraries background and digital book focused than resource oriented
- Verbose
- Not tuned for interoperability
- XML schema not machine readable
- Profile dictates vocabulary, not controlled standard vocabulary
- To exchange documents some profiling required
- No example of transmission of objects
- Lack of GUIS for more general use than librarian community
- No tool to build METS manifests
- Take up by a large diversity of projects not feasible for small scale projects
- Interoperable. METS exported from DSpace won't ingest into FEDORA Internal links not externally referenced



MPEG 21 DIDL (Strengths)

- Very extensible and flexible
- Plugins available for major platforms
- Well established
- Standard freely available
- Industry backing
- Rights expression and management
- Flexible and extensible, use of existing ID schemes, use of existing DRM schemes
- Cost XML and Java Tools no fees related to use of Standard
- Not coming from library background



MPEG 21 DIDL (Weaknesses)

- Not coming from a library background and hence low uptake
- Industry standard seen negatively in education (i.e. business model different?)
- Needs heavy profiling (most educational institutions would not have the resource that LANL has?)
- Relationship between items at profile level, must be a hiearchy.
- Sequencing and versioning
- Flexibility, hence interoperability issues
- Some mystery about standard itself
- Expertise needed to deploy its use with whatever repository software the institution uses
- Expertise needed to construct DIDL XML files
- Commitment necessary from institution for its deployment
- Possible issues with OAIPMH
- Flexibility
- Questions over 'bit equivalence' concept
- Standard format for encoding profiles



IMS CP (Strengths)

- Simple and Flexible
- Tuned for Display
- Ability to package other content packages
- Most common denominator user interface rather than standard
- Ubiquity within e-learning and tools available and easy to use
- Multiple implementations



IMS CP (Weaknesses)

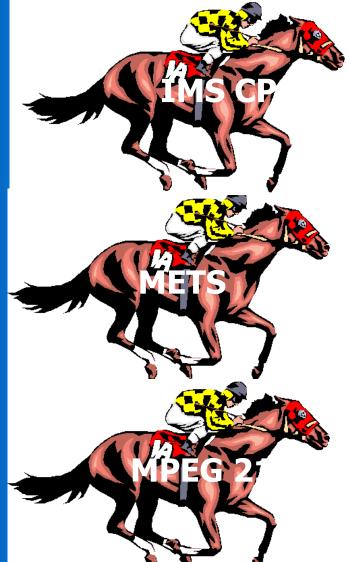
- Too simple and lack of sophistication and encodes a restricted set of relationships
- Limited to simple ordered tree structures
- Just learning community based not transferable to other areas
- Issues of Interoperability with the standards
- Doesn't enforce assignment of Global unique Identifiers
- Not enough people in libraries know about it



Which one wins?



Backing the right horse, riding them all, or changing jockeys?





Digital Library Heaven