## METS 2 RDF

## Some tentative conclusions (7th of July 2015)

Checking Tom’s proposal against our SPAR examples leads us to some remarks:

* **Where is the METS 2.0 model? The xsd file is rather old now and does not correspond to Tom’s proposal.**
* Save the fileSet / structMap distinction! We need to say technical stuff about file groups that don’t correspond to a logical/physical division… (same distinction in XFDU: informationPackageMap =structMap, dataObjectSection =fileSec) and sometimes put in common file-specific technical information whose repetition would be very verbose.
* In general, the METS concepts have to be thoroughly examined to determine how useful they will be in RDF. Some of them have already been taken over by the PREMIS ontology (file, bitstream, message digest and message digest algorithm, size, format, etc.). Some others are meaningful in XML but not really (tmho) in RDF (header, METS document, file section, etc.). Some others are specific to METS and should become properties and classes: structural map, file group, use, parallel and sequential files, order, metadata, etc.
* If we choose to reuse existing ontologies, we will come up with a “Frankenstein” ontology. But reusing may be the wisest choice.
* In RDF, there is no “METS document”, a fortiori no “METS file”. Just triples in a big graph. So the frame we use until now should be forgotten. The assertions we will make should be on the digital document the triples concern, not on a “document” that no longer exists. If we want to say something about the triples created on a single document at a certain time, couldn’t we use the named graph mechanism? Or if we want to keep a class that describes the whole metadata created on a digital document (as METS was designed to be a “digital object binding”), maybe use another word as “mets:description”?
* To document structural relationships between parts of the METS, we used OAI-ORE in SPAR. Tom proposes a specific METS property: mets:hasPart. We consider that it can be synonym to ore:aggregates, and that we need the inverse property: we might want to refer to higher levels of hierarchy (especially when describing an archive fonds). Our SPAR “set” level is an example: in every METS manifest describing an “information package” we mention a root division which is at a higher level than the digital object we describe. I am not sure if this is compatible with the data model Tom proposed.

|  |  |
| --- | --- |
| Tom’s proposal | SPAR mechanism |
| <METS document> mets:hasStructuralMap <root div>.  <root div> a "periodical”;  mets:hasPart <div1>.  <div1> a “issue”. | <ark:/12148/bpt6k948k23j> a “issue”;  ore:isAggregatedBy <ark:/12148/bc6pk495kj>;  ore:aggregates <ark:/12148/bpt6k948k23j/f1>.  <ark:/12148/bc6pk495kj> a “periodical”.  <ark:/12148/bpt6k948k23j/f1> a “page”. |

* Metadata: do we keep the distinction descriptive / technical / provenance / source? A property “hasMetadata” and a class “metadata” could be useful to mention the creation date and context or authors of the metadata. What class do we use to name these metadata? Tom’s proposal suggests to replace technical metadata by premis:objectCharacteristics. Isn’t it too generic?
* Structural map: Tom’s model keeps it just as a property (hasStructuralMap) but its range is a mets:division. I think we should keep it, because one METS document may have several structural maps and we need to characterize it as “logical”, “physical”, etc.
* File group: this is very useful for us to put in common object characteristics (file format, resolution, color depth, etc.). So we should keep it. Reducing verbosity is a major issue: one needless triple, when archiving millions of documents, is quite a thing.
* Working on identifiers will probably be necessary, but I don’t know if it is in the scope of the WG…
* Why Tom proposed a mets:file class? Can’t we manage with the premis:file alone?
* Two years ago, we led a conceptual study on the SPAR data model, and it turned out that the distinction between the intellectual entity and its digital representation was not clear enough. Maybe we should take this into account when thinking about the METS data model.
* Alignment between PREMIS representation concept and METS notions (division, file group?) does not seem straightforward to us.
* Question on the ordered and labeled divisions in Tom’s model: without ORDER or LABEL there is a direct link between division and file (slide 35), but with ORDER / LABEL there is a blank node between them (slide 45)?? Instead of the blank node, we should have a “page 1” division (that could have several manifestations, ex.: “P00001.tif” and “P00001.xml”).

|  |  |  |  |
| --- | --- | --- | --- |
| METS concept | Definition of the XML element | Definition of the RDF element | Class / property |
| mets:mets | METS document represented in the XML file by the METS root element | Assertions about a digital document describing its structure and metadata | mets:description?  mets:manifest?  mets:document?  mets:packagingInformation? |
| mets: |  |  |  |
|  |  |  |  |