A Cohesive Approach to Format Registries

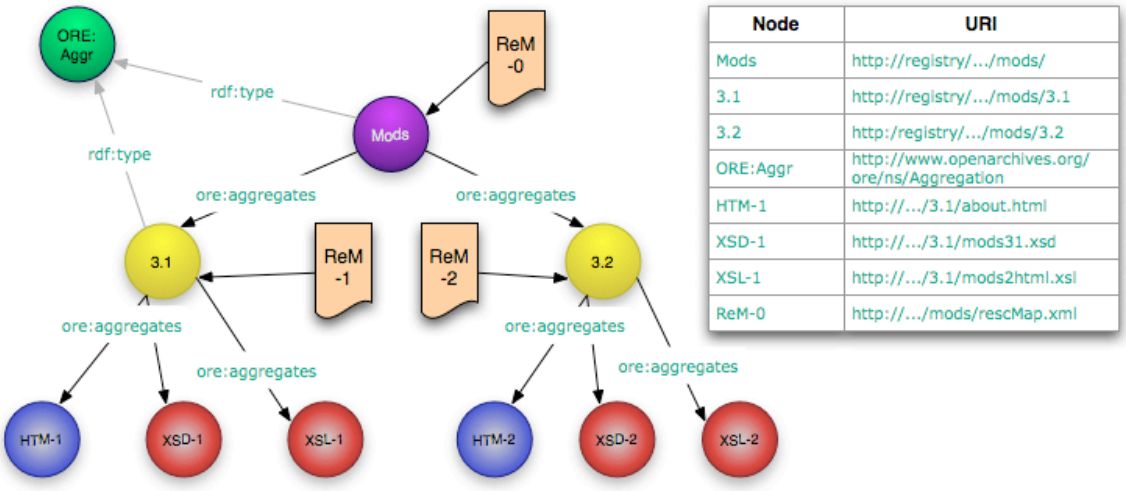
Authors: Robert Sanderson, Herbert Van de Sompel
Research Library, Los Alamos National Laboratory
Version: 2009-10-15

Digital Library protocols frequently require identifiers for the formats in which they make records available, such as XML schemas like MODS, Dublin Core profiles, PRISM and so forth. The current situation is that each protocol maintains its own list of different identifiers for the same common schemas. The maintainers of these protocols recognize the increasing need for a shared registry of schema identifiers, in exactly the same way as digital preservation systems require shared registries for digital object formats.

Equally, similar information is required for this use case as for the digital preservation case -- it is important for client and server developers to have access to both technical and descriptive metadata, such as human readable documentation, schema files, XSLT stylesheets for transforming between formats and other related information. A client might even retrieve the XSLT stylesheet directly from the registry when it encounters a format that it cannot natively render.

We have the following suggestions as to the integration of the format registry into the web infrastructure:

> An HTTP Request Header could be defined that acts in the same way as current content negotiation (the Accept header) but instead of mime types, it would use the URIs for the formats. This would allow the retrieval of specific versions of a format, rather than any version. It would also enable data in a particular schema, rather than format to be retrieved without registering unique mime types with IANA.
>
> It is also hoped that the registry would adopt Linked Data principles. When resolving the URI identifiers for the formats, the system would redirect the client to appropriate descriptions. A format like MODS would be an ORE Aggregation of its various versions (3.1, 3.2, 3.3, ...) Each version would then be an ORE Aggregation of the descriptive and technical metadata files, and hence the description retrieved would be an ORE Resource Map.



Node	URI
Mods	http://registry/.../mods/
3.1	http://registry/.../mods/3.1
3.2	http:/registry/.../mods/3.2
ORE:Aggr	http://www.openarchives.org/ore/ns/Aggregation
HTM-1	http://.../3.1/about.html
XSD-1	http://.../3.1/mods31.xsd
XSL-1	http://.../3.1/mods2html.xsl
ReM-0	http://.../mods/rescMap.xml

Figure 1: ORE Modeling of Format Registry