The RAMLET project – A Conceptual Overview

by the IEEE LTSC RAMLET Working Group

Sponsor:

The Learning Technology Standards Committee
of the
IEEE Computer Society

1. Introduction

The LTSC Resource Models for Learning, Education, and Training (RAMLET) Working Group is undertaking a project to produce an IEEE standard that will define a conceptual model for digital aggregations of resources for learning, education, and training applications. The standard will facilitate interoperability by enabling the interpretation of external representations of resource aggregations and their properties. The standard will assume general means and methods for processing resource aggregates and will not address internal compositions, behaviors, or rendering of resources that make up resource aggregates. In addition to the standard, the RAMLET Working Group is developing five recommended practices that will provide mappings of aggregation formats to the conceptual model.

2. Resource aggregations

Resource aggregation is the process of gathering digital resources and describing their structures, so that the resulting resource aggregation can be used for transmission, storage, and delivery to users.

The creators of digital materials for learning, education, and training usually draw together more than one digital resource to create a group of related resources. Creators may apply a structure to this group of digital resources and associate it with descriptions of the educational and technical attributes of the group as a whole and of each individual resource. Information intended to aid in the management over time of the group of digital resources may also be created. Many people have realized the benefits of sharing these materials between communities. To make the materials useful when shared, the accompanying information needs to be persistently associated with the group of digital resources so that the resources and the accompanying information become a single, identifiable, resource aggregation.

To facilitate the exchange and interoperability of resource aggregations, communities of practice have created agreements known as aggregation formats that document the methods of aggregating digital resources into objects that can be exchanged among systems. An aggregation format may be defined by a formal specification or standard, but may also be informal. The defining characteristic is that an aggregation format specifies how to combine digital resources into a structured whole without prescribing the kinds of digital resources, their internal structures, or their intended uses.

A resource aggregation is instantiated using an aggregation definition document, a document that lists and structures the digital resources of an aggregation. Examples include the manifest document defined by IMS Content Packaging (IMS CP)[1], the mets document of Metadata Encoding and Transmission Standard (METS)[2], and the Digital Item Declaration Language (DIDL) document of MPEG–21 Digital Item Declaration (DID).[3]

3. Why and how the conceptual model and mapping recommended practices will be developed

Individual communities have developed different aggregation formats, each focusing on that community's special needs and requirements. There is a need to exchange resource aggregations between communities that may be using different aggregation formats. In these cases, it is necessary to map between aggregation formats. Several use cases were used by the RAMLET Working Group to inform the approach used to develop the RAMLET conceptual model.[4]

Different methods, such as Extensible Stylesheet Language Transformations (XSLT)[5] and Perl scripts[6], have been used to undertake mappings between aggregation formats with varying degrees of information loss. The standard will use a Semantic Web technology approach in the form of a conceptual model to map between aggregation formats. The conceptual model describes the functional similarities of and the relationships among the components of the data models of various aggregation formats. It focuses on the structures of aggregation formats by analyzing and describing the functional characteristics of common structural elements, then abstracting and expressing them in ontological form as a core ontology. Figure 1 shows the relationship between the extent of the conceptual model and aggregation formats.

[1] IMS Specification, IMS Content Packaging, Version 1.1.4, Final Specification.
[2] METS, Metadata Encoding and Transmission Standard: Primer and Reference Manual, Version 1.6, Revised. (See: http://www.loc.gov/standards/mets/METSPrimerRevised.pdf)
[3] ISO/IEC N4813, Information Technology – Multimedia Framework – Part 2 Digital Item Declaration.
[4] The RAMLET Working Group, The RAMLET Project – Use cases.
[5] W3C Recommendation, XSL Transformations (XSLT) Version 2.0. (See: http://www.w3.org/TR/xslt20/)
[6] Perl 5 version 12.2 documentation. (See: http://www.perl.org/)

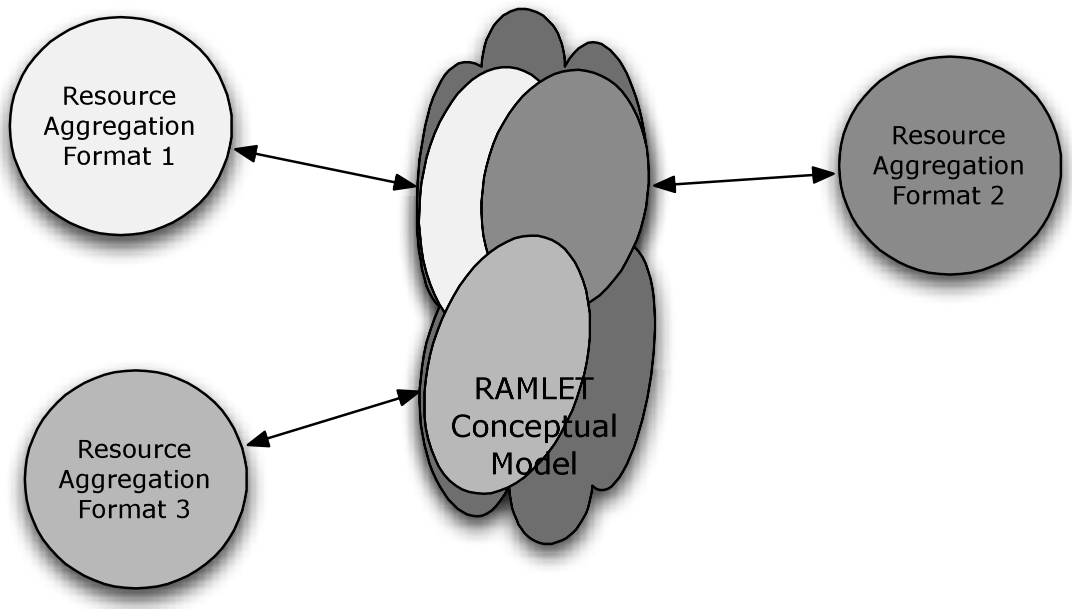

Figure 1 – The extent of the conceptual model and aggregations formats

Significant concepts within the conceptual model map to significant concepts of different aggregation formats, thus identifying the overlaps and differences among them. Not all concepts within the conceptual model may be present in any particular aggregation format. However, for those aggregation formats that have been authoritatively mapped to the conceptual model in the course of its development, all structural details of the aggregation formats are represented in the model.

The conceptual model does not address generic content models, such as that described in the Abstract Learning Object Content Model (ALOCOM)[7], or informational properties of the content of resource aggregations. For example, the conceptual model is not concerned with the semantic relationship between sections and chapters of a book or with the interactive properties of different media types. Because overlaps exist between the structural descriptions provided by the conceptual model and generic content models, software developers, content managers, and other information specialists will need to consider how these different informational aspects relate to their specific circumstances.

A clear set of principles was used in the development of the conceptual model.

- Describe only the semantics of the functionality of each component as described in the data models of the aggregation formats.

[7] More information on the ALOCOM project is available at http://www.cs.kuleuven.be/groups/hmdb/alocom/index.php

- Do not describe the epistemological status or pedagogical import of a component.
- The upper and lower boundaries of the model are, respectively, the classes resourceAggregation and digitalResource and its children. Although two of the boundary classes have real-world examples, it is not a requirement that boundary classes have examples.
- Between the boundary classes of the conceptual model, include only classes that have real-world examples.
- Do not describe the internal structures of the digital resources that are aggregated. Digital resources are treated as black boxes.
- Do not describe the epistemological, instructional, or educational purposes of the digital resources that are aggregated.
- Do not specify processing methods for resource aggregations.
- Do not describe the internal structures of descriptive metadata schemas.

NOTE – The description of the epistemological, instructional, or educational purposes of the digital resources could be the role of another standardization activity, based on the previously mentioned work of the ALOCOM project and the Organization for the Advancement of Structured Information Standards (OASIS) Darwin Information Typing Architecture (DITA) Learning and Training Content Specialization Subcommittee.[8]

4. Why use an ontological approach and OWL?

The conceptual model was developed using an ontological approach and the set of principles described above. Five aggregation formats, IMS CP, METS, MPEG–21 DID, Atom Syndication Format (Atom)[9], and Open Archives Initiative Object Reuse and Exchange (OAI–ORE)[10], were analysed, and a core ontology was constructed that represents an abstract generalization of the aggregation formats. For example, the ontology's topNode class represents the root data element of a resource aggregation and corresponds to the IMS CP manifest, Atom feed, METS mets, OAI–ORE resourceMap, and MPEG-21 DID DIDL data elements. By mapping each component of multiple aggregation formats to the core ontology, mappings can then be generated between aggregation formats.

To develop the conceptual model, each of the five aggregation formats were analyzed and applied to the developing ontology. The ontology was refined to avoid inconsistencies until the conceptual model was expressed as a core ontology. Mappings were then defined between the conceptual model and the aggregation formats and expressed as ontologies. These mappings will be published as IEEE recommended practices.

---

[8] More information on the DITA Learning and Training Content Specialization Subcommittee is available at http://wiki.oasis-open.org/dita/LearningSubcommittee
[9] Networking Group RFC 4287, The Atom Syndication Format. (See: http://tools.ietf.org/html/rfc4287)
[10] ORE Specifications and User Guides, Open Archives Initiative, October 2008. (See: http://www.openarchives.org/)

The conceptual model, which will be expressed in human-readable tables in the standard, was first implemented in the World Wide Web Consortium (W3C) Web Ontology Language (OWL).[11] OWL was selected for use as a formalism for expressing the structural concepts of the conceptual model and their interrelationships for the following reasons:

- OWL is an open standard.
- OWL allows expression of the relationships between components explicitly.
- OWL allows a well-understood method of specialization.
- OWL is part of a a Semantic Web technology stack that was designed to deal with heterogenous data.
- OWL is supported by readily available and well-developed tools, including open source tools that were available to the RAMLET Working Group.

The OWL expression of the conceptual model was developed first because the explicit semantics require the model to be described rigorously, and this formalism can be enforced with OWL development tools.

Ontology mappings defined by RAMLET are bidirectional, i.e., an IMS CP resource aggregation can be expressed using RAMLET terms and vice versa. To map resource aggregation instances from one aggregation format, such as IMS CP, to another, such as METS, two steps are required. First, a mapping is made from the IMS CP instance to the core ontology and then from the core ontology to the METS instance.

This approach can scale as the number of aggregation formats increases. To include a new aggregation format, such as Really Simple Syndication (RSS)[12], only a bidirectional mapping between the core ontology and the new aggregation format would need to be expressed. Without the RAMLET conceptual model, instead of a single mapping from RSS to the core ontology, individual mappings from RSS to all other aggregation formats of interest would need to be developed (e.g., RSS to Atom, RSS to METS, RSS to MPEG–21 DID, etc.).

RAMLET provides an integrative conceptual model that covers IMS CP, MPEG–21 DID, METS, Atom and OAI–ORE. Coverage of these aggregation formats implies that all data elements of these formats are represented in the conceptual model, which enables lossless transformation from the aggregation formats to the conceptual model. However, a potential loss of information exists when translating from an aggregation instance in one format to the core ontology and then to an aggregation instance in another format (e.g., METS to RAMLET to IMS CP), depending on the semantic overlap between the source and target aggregation formats. Although the transformation between resource aggregations is not always lossless, important structural relationships should be preserved, thus achieving a maximal preservation of meaning.

---

[11] W3C Recommendation, OWL 2 Ontology Language Document Overview. (See: http://www.w3.org/TR/owl2-overview/)

[12] RSS Specification, RSS Advisory Board. (See: http://www.rssboard.org/rss-specification)

5. Extension and use of the core ontology

In addition to a maximal preservation of meaning, the RAMLET approach has the advantage of scalability as the number of aggregation formats increases. Adding an aggregation format requires only one additional mapping ontology to establish formal relationships between the new format and all existing mapped formats. While the addition of a new format would leave the mapping from the core ontology to the existing mappings unchanged, the maximal preservation of meaning from the additional mapping to the core would probably require an extension to the core. Such an extension would capture those components of the new aggregation format that are unique to that format.

The core ontology can also be used to develop new aggregation formats, both because it records many of the components that have already been used in content aggregation, and – via the mappings – because it indicates how widely those components are used in existing formats. In addition, generalizations about how the various components relate to each other are captured by the core and mapping ontologies. Consequently, it is easier for communities to follow good practice and design new formats that stay within the bounds of the semantics of existing formats as much as possible, adding new functionality only where absolutely necessary.

Because the RAMLET core and mapping ontologies have been formulated using standard Semantic Web technology, extensions to the ontologies can be easily accommodated by defining equivalence, refinement, inclusion, or any other relationship between the extensions and the existing ontology using OWL and other Resource Description Framework (RDF)[13] vocabularies, such as RDF Schema (RDFS)[14] and Simple Knowledge Organization Systems (SKOS)[15].

Likewise, the ontologies can be combined with existing vocabularies or ontologies that are designed to describe aspects of a resource aggregation other than aggregation functions. One such combination that is immediately obvious is the augmentation of the RAMLET conceptual model with descriptive metadata schemes. Many resource aggregations contain descriptive metadata of some kind that would be well worth preserving in a RAMLET–enabled environment. Another case would be the combination of the RAMLET conceptual model with generic content models, such as ALOCOM or DITA.

6. The binding of the core ontology

An OWL ontology can be represented or bound in several different ways so that it can be used by different implementation systems. An RDF/XML[16] representation is provided as downloadable files (see **Error! Reference source not found.**). The advantage of

---

[13] W3C Recommendation, RDF Primer. (See: http://www.w3.org/TR/rdf-primer/)
[14] W3C Recommendation, RDF Vocabulary Description Language 1.0: RDF Schema. (See: http://www.w3.org/TR/rdf-schema/)
[15] W3C Working Group Note, SKOS Simple Knowledge Organization System Primer. (See: http://www.w3.org/TR/skos-primer/)
[16] W3C Recommendation, RDF/XML Syntax Specification. (See: http://www.w3.org/TR/REC-rdf-syntax/)

RDF/XML is that it is among the more widely implemented RDF binding formats and can easily be translated into other binding formats using a variety of tools. A Terse RDF Triple Language (Turtle)[17] representation will be provided in the standard, because it is one of the more human readable RDF bindings.

In terms of which RDF statements get expressed where, it is possible to express both the core ontology and all the mapping ontologies in a single document. For some implementations, that may well be the most expedient way of making using the standard. For the standard itself, however, the core ontology and each of the mappings will have their own namespaces and are represented in their own files, mainly to aid maintainability. A change to one ontology can leave the other ontologies unaffected.

In order to express how the mapping ontologies relate to the core ontology, the file set for the standard and associated recommended practices will be structured so that the core ontology file imports the other files. By extension, it is advisable to explore the mapping ontologies through the core ontology in this manner, because it will place the mapping-to-core relationships in their complete context. Looking at the mappings in isolation of the core ontology means that similarities with other formats will be missed as will crucial information about the structure of the relevant core ontology classes and properties. The core ontology, as distributed, will import the latest version of each mapping ontology. Persistent identifiers are allocated to each version of the mapping ontologies, and the core ontology can be edited to import specific versions of the mapping ontologies, if required, by using the relevant persistent identifiers.

In the standard, class and property names will be represented in camel case for human readability and in lowercase in the binding files for processability. The underlying reason is that, though aggregation formats may prescribe case, actual aggregation definition documents are rather variable in practice. Tools also vary in their case sensitivity. Therefore, in order to aid transformation and querying, it is easiest to reduce all incoming data to lowercase, and the ontology binding files reflect this.

Finally, although the cardinalities for data types will be included in the downloadable files for the core ontology, they not will not be included in the class tables in the standard because the cardinality requirements for the core ontology are minimally restrictive. The cardinality requirements for the aggregation formats that are mapped to the core ontology in the mapping recommended practices vary and are more restrictive. Therefore, cardinality requirements for each aggregation format will be found in the mapping recommended practices.

7. Implementation Experiences

Various options and architectures for the implementation of a RAMLET–based infrastructure have been proposed in a RAMLET implementation study report[18]. It concluded

[17] W3C Team Submission, Turtle – Terse RDF Triple Language. (See: http://www.w3.org/TeamSubmission/turtle/)
[18] Kraan, W., *RAMLET* implementation *study report*, JISC–CETIS, August 2008. (See: http://digitalcommons.bolton.ac.uk/iec_reports/28/)

that, although a centralized repository using RAMLET–structured RDF was one way of providing transformations from one aggregation format to another, a more distributed architecture was likely to be more resilient and easier to implement. In such a distributed architecture, existing repositories of content aggregations could implement an agent or SPARQL Protocol and RDF Query Language (SPARQL)[19] endpoint to expose their contents, which could then be queried in RAMLET terms and the return transformed into the desired aggregation definition document. Work on the required XSLTs and SPARQL queries continues.

[19] W3C Working Group Note, SKOS Simple Knowledge Organization System Primer. (See: http://www.w3.org/TR/skos-primer/)