Date submitted: 04/08/2010

The Europeana Data Model (EDM)

Martin Doerr
Institute of Computer Science, Foundation for Research and Technology - Hellas
Crete, Greece

Stefan Gradmann
Library and Information Science, Humboldt University
Berlin, Germany

Steffen Hennicke
Humboldt University
Berlin, Germany

Antoine Isaac
Europeana
The Hague, Netherlands

Carlo Meghini
Istituto di Scienza e Tecnologie dell'Informazione
Pisa, Italy

Herbert van de Sompel
Research Library, Los Alamos National Laboratory
Los Alamos, NM, USA

Meeting: **149. Information Technology, Cataloguing, Classification and Indexing with Knowledge Management**

WORLD LIBRARY AND INFORMATION CONGRESS: 76TH IFLA GENERAL CONFERENCE AND ASSEMBLY
10-15 August 2010, Gothenburg, Sweden
http://www.ifla.org/en/ifla76

Abstract:

The Europeana Data Model (EDM) is a new approach towards structuring and representing data delivered to Europeana by the various contributing cultural heritage institutions. The model aims at greater expressivity and flexibility in comparison to the current Europeana Semantic Elements (ESE), which it is destined to replace. The design principles underlying the EDM are based on the core principles and best practices of the Semantic Web and Linked Data efforts to which Europeana wants to contribute. The model itself builds upon established standards like RDF(S), OAI-ORE, SKOS, and Dublin Core. It acts as a common top-level ontology which retains original data models and information perspectives while at

the same time enabling interoperability. The paper elaborates on the aforementioned aspects and the design principles which drove the development of the EDM.

---

1 Introduction to Europeana

Europeana is often presented in public as a portal giving access to millions of objects from all kinds of cultural heritage communities - and even though this way of presenting Europeana conveys some truth it doesn't capture some of the essential characteristics of what Europeana actually is designed to be in the end. Concordia et al. (2010) have tried to make clear that Europeana is not so much a portal characterised by sheer volume, but that the core agenda of our endeavour is to make rich data and functionality available on an API basis. This would allow all kinds of external communities to make use of our rich (and numerous) representations of European cultural treasures for their own needs - and the Europeana portal that is offered at http://www.europeana.eu should in the end be seen as one of the parties making use of this wealth of data and functionality by means of the API.

The idea furthermore is to offer rich semantic contextualisation for the object representations in Europeana in such a way as to enable complex semantic operations on these resources in a way that would not be supported by a traditional digital library interface. In order to enable such functionality the object representations in Europeana need to be systematically connected to Linked Open Data on the WWW or else to semantic contextualisation resources held within the Europeana data space such as thesauri and structured vocabularies migrated to the SKOS standard.

The Europeana technical strategy thus was conceived to both contribute to and benefit from the growing Linked Data paradigm.

2 From ESE to EDM

However, the main obstacle to overcome before moving on this road to Linked Open Data was the Europeana Semantic Elements (ESE) format with its underlying simple and robust data model. Creation and use of the ESE had been unavoidable as a pre-condition for launching a first early prototype of Europeana in November 2008 - but in the meantime its limitations have become visible and - somewhat paradoxically! - prevent from moving into a semantically rich functional model (ESE probably shouldn't have been baptised “semantic” in the first place ...)

In essence, the major problem with ESE is its ‘flat’ modelling approach that doesn’t allow for embedding links to external resources on the web as well as its non-extensibility in terms of more specialised, fine-grained models. Furthermore, the ESE cannot be plugged in Linked Data Namespaces the way this would be required to make Europeana part of such future distributed information architectures – to just sketch the most prominent issues with ESE.

Soon after starting work on the Europeana Semantic Data Layer in work package 1 of the EuropeanaConnect project it thus became evident that ESE would have to be replaced urgently by results of the on-going process for the specification of a Europeana Data Model (EDM) – and it was thus decided to speed up this process for quickly obtaining valid and usable results - and this working phase started in late summer 2009 yielded a version 5 of the EDM which is considered as a candidate for operationalization and which is reported on in the present paper.

3 Design Principles of EDM

3.1 W3C Standards and Semantic Web

In its first incarnation, the WWW essentially was a very large hypertext application with lots of interlinked ‘documents’ (= web pages) each of which is assigned a Unified Resource Identifier (URI). This picture changes with the semantic web, which – among other things – now includes representations of real world entities (so called non-information resources, cf. http://www.w3.org/TR/cooluris), and these as well are assigned a URI. In a recent attempt to further clarify the fundamentals of the Semantic Web the notions of ‘Linked Data’ (http://linkeddata.org/) or ‘Data Web’ have developed which introduce the idea of a web document (a so called ‘information resource’) representing the non-information resource and the idea of systematic redirection from the latter to the former.

A crucial change that adopting EDM will bring to Europeana is compatibility with this Semantic Web paradigm. Over the past years, the World Wide Web Consortium (W3C) has introduced a number of standards to enable the representation and sharing of machine-accessible, structured data over the web. The first one is Resource Description Framework (RDF) which allows representing structured information about any resource in the form of simple triple statements (subject, predicate, object). The vision guiding RDF is that resources can be described by means of semantically meaningful connections between them. For example, the triple (ec:ulysses, ex:author, ex:james_joyce)[1] describes the book Ulysses by connecting its identifier to another that stands for James Joyce, using an author typed link which denotes the relation between a book and its author. EDM entirely follows this triple description approach.

Link types are crucial to the ability of RDF to convey meaningful knowledge. These properties, together with types that can be assigned to the subject and object resources (classes) are defined in ontologies. The word “ontology” is understood here as a synonym of “conceptualization” (Gruber 1993) as opposed to the meaning that the word has in philosophy: “the science of what is, of the kinds and structures of objects, properties, events, processes and relations in every area of reality.” (Smith 2003). Ontologies, which are defined by means of the RDF Schema (RDFS) and Web Ontology Language (OWL) standards, contain both informal definitions, in the form of human-readable documentation, and formal definitions, in the form of constraints and rules that allow detecting inconsistencies or deriving new facts from asserted ones. An ontology can for example define classes for books, paintings and persons, one author property, and formally state that all resources connected to books by the author property will be of the person type. It can also formally define another class object as a superclass of both book and painting. Running an inference engine on top of data for a collection and books and paintings, and querying for all objects created by one person would allow retrieving all these objects without prior knowledge of their specific type, a crucial feature when information integration is required.

The Semantic Web approach indeed enables the combination of various ontologies in the same descriptions. One can deploy different views on top of the same assertions, or build assertions that combine different vocabularies tailored to specific needs so as to match the requirements of a more general application. RDF data for a book catalogue may thus re-use an ontology for book description, to represent the core data of book records, and an ontology for persons to finely describe the authors of the books, as given in an authority file.

[1] Where ex: denotes an arbitrary example namespace.

EDM will re-use some of the reference ontologies already available, such as the W3C-sponsored Simple Knowledge Organization System (SKOS). SKOS defines a model to represent the elements of Knowledge Organization System such as thesauri, classification schemes, and their likes. SKOS features a main class to describe concepts. Adapting standards like ISO2788 to a concept-based modelling approach, it coins properties for the labels of these concepts (e.g., skos:prefLabel for the preferred label of a concept, skos:altLabel for the alternative ones), for semantic relationships between these concepts (skos:narrower, skos:broader, skos:related) and for general concept documentation (skos:scopeNote, skos:definition, etc.). Importantly, SKOS is tuned towards matching across concept schemes, e.g. by linking concepts from different thesauri which are semantically equivalent using the skos:exactMatch property. This technically enables applications to navigate through a semantic layer of concepts from different sources, leveraging such conceptual network to access objects that are originally described using different – but semantically related – concepts.

Other ontologies that are of specific interest for EDM include Dublin Core (DC) and Friend-of-a-Friend (FOAF). Dublin Core gives a compact vocabulary to describe the core features of culture objects (creators, relations to other resources, subject indexing, etc.) in a Semantic Web-enabled fashion that fits a very wide range of needs. DC is used as the basis for ESE: keeping it as a part of the model thus gives direct compatibility with legacy Europeana data. It also enables these providers that do not want to provide richer descriptions to keep to a simple vehicle for the data they submit. It finally makes EDM data better fit to sharing and re-use: DC is used by many applications, which could be very easily adopted to consume EDM data.

FOAF is an ontology used to describe persons in RDF, originating from web profile description requirements. It could thus be fit – though with some adaptation or extension – to describe the many persons that take crucial part in the context of Europeana objects.

3.2 "Object Reuse & Exchange" framework

Typical object representations in Europeana mostly will be compound entities consisting of several parts, such as for instance metadata attributes, a thumbnail picture and a static html landing page (to give just a very simple example). For this and other reasons the OAI Object Reuse & Exchange (OAI-ORE) specifications (http://www.openarchives.org/ore/1.0/to) were chosen as the structural modelling framework for the EDM ontology. OAI-ORE defines an approach for the identification and description of sets of Web resources. In order to be able to unambiguously refer to a set of Web resources, a new Resource is introduced that stands for collection of other Resources. This new Resource, named an Aggregation, has a URI just like any other Resource on the Web. Since an Aggregation is a conceptual construct, it is modelled as a non-information resource that does not itself have a Representation, but rather is described by another Resource. This latter Resource is named a Resource Map; it has a URI and a machine-readable Representation that provides details about the Aggregation. In essence, a Resource Map conveys which Aggregation it describes (the ore:describes relationship in Figure 1), and it lists the Aggregated Resources that are part of the Aggregation (the ore:aggregates relationship in Figure 2, a subproperty of dcterms:hasPart). In addition, a Resource Map can express relationships and properties pertaining to all Aggregated Resources, as well as metadata pertaining to the Resource Map itself. For example, Figure 2 shows that authorship and modification time of the Resource Map are conveyed (the dcterms:creator and dcterms:modified relationships, respectively). A Resource

Map can also express relationships of the Aggregation, Aggregated Resources, and the Resource Map itself with any arbitrary other Resource.

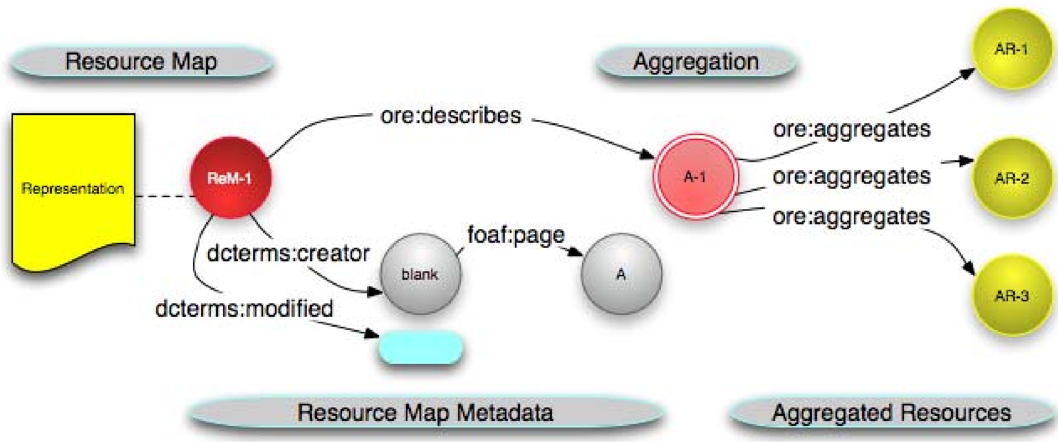

Figure 1: The core components of the OAI-ORE Data Model

In addition, the data model allows expressing that an Aggregated Resource is itself an Aggregation (nesting Aggregations). To that purpose, an ore:isDescribedBy relationship (the inverse of ore:describes), and a subproperty of rdfs:seeAlso) is expressed between the Aggregated Resource and a Resource Map that states that the Aggregated Resource is itself an Aggregation. Also, for discovery purposes, the data model allows a Resource Map to express that an Aggregated Resource of a specific Aggregation is also part of another Aggregation. This is achieved by means of the ore:isAggregatedBy relationship (the inverse of ore:aggregates) between the Aggregated Resource and that other Aggregation. Furthermore, the use of non-protocol-based identifiers that can be expressed as URIs (such as URNs) is quite common for referencing cultural heritage assets. In order to support this practice, the ore:similarTo relationship between an Aggregation and a somehow equivalent resource identified by a non-protocol-based URI is expressed. The specificity of ore:similarTo is situated between rdfs:seeAlso and owl:sameAs. It should be pointed out that the Linked Data community is still debating the need for equivalence expressions other than owl:sameAs (http://events.linkeddata.org/ldow2010/papers/ldow2010_paper09.pdf).

Note that the URI that denotes an Aggregated Resource in a particular Aggregation is no different than the URI that identifies that Resource independent of the Aggregation. However, many uses cases – including the Europeana one – require a distinction to be made between referencing a Resource as such, and referencing the same Resource when it acts as an Aggregated Resource in an Aggregation. Citation in context and provenance tracking are examples. To accomplish this differentiation, OAI-ORE introduces the notion of a Proxy. A Proxy is a Resource that stands for an Aggregated Resource in the context of a specific Aggregation. The URI of a Proxy provides a mechanism for denoting a Resource in context. Figure 2 shows the ore:ProxyFor and ore:ProxyIn relationships between a Proxy and an Aggregated Resource and an Aggregation, respectively. It also illustrates how citing the Aggregated Resource is different from citing its Proxy: the former cites a Resource "as is", the latter cites that Resource as it exists in the context of a specific Aggregation. In order to work seamlessly in the Web and to provide context information to OAI-ORE aware clients, resolution of HTTP URIs assigned to Proxies must lead to the Aggregated Resource, and the response must include a HTTP Link Header (https://datatracker.ietf.org/doc/draft-nottingham-http-link-header/) that points to the Aggregation.

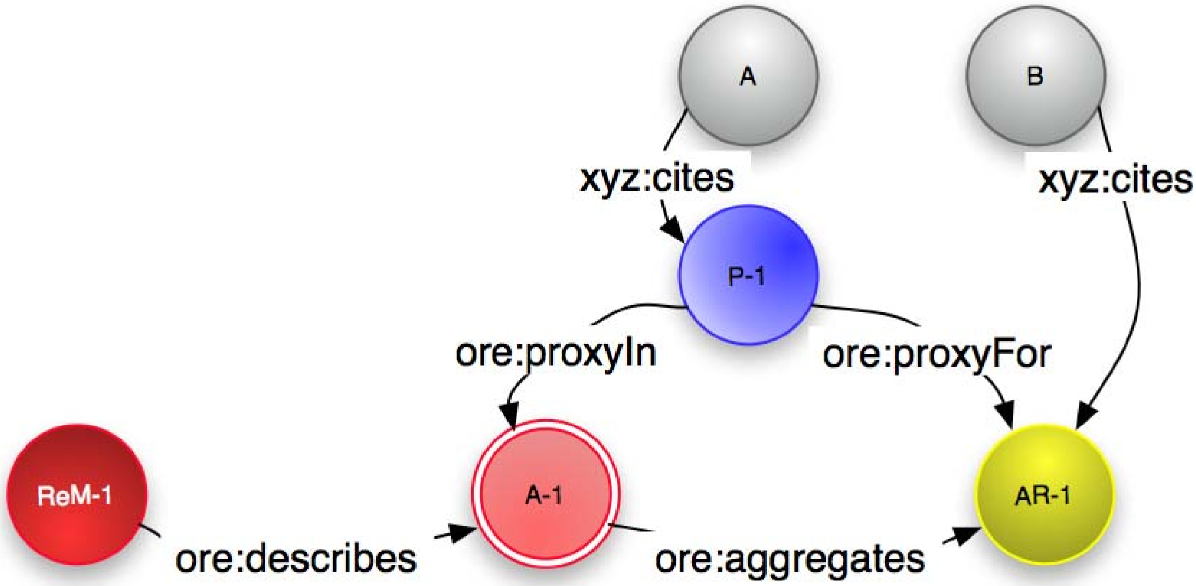

Figure 2: OAI-ORE Proxy: referencing an Aggregated Resource in Context

3.3 EDM and Linked Open Data

The goal of Linked Data is to enable the sharing of structured data on the Web. To this end, Linked Data relies on the Web architecture as the enabling backbone, and on RDF as a representation language. It makes recommendations on how to make accessible RDF data on the web, with the overall vision of the web as a data commons typically called the Semantic Web.

The vision of the Semantic Web has informed to a large extent the design of the EDM. In particular, it has led to the adoption of RDF as the EDM meta-model, and to the decision of making any object of interest in the Europeana information space (whether a Cultural Heritage Object, or a contextualization entity such as a person, a place, a concept and the like) a resource, identified by a HTTP URI. This choice enables the normalization of the values in the Europeana descriptions to the HTTP URI format, a de facto standard supported by the web architecture.

Linked Data adds a fundamental dimension to this vision, because through Linked Data Europeana can use the HTTP URIs in its information space also as links enabling access to structured descriptions of the corresponding objects. These links act therefore as connectors of the Europeana information space with the information space of other authorities, allowing Europeana to collect additional knowledge about people, places, concepts, and so on. Needless to say, the so collected knowledge is expected to play a major role for improving the usability of Europeana in important aspects such as the performance of the discovery functionality.

3.4 Abstraction of relevant semantic relationships

The EDM does not bind the representation of ingested metadata to one common schema but acts as a common top-level ontology according to which metadata compliant with other, original data models can be expressed. It seeks coverage rather than the most common fields. This allows for integrating the distinct information perspectives and needs of the various communities providing data to Europeana. The original richness of community standards like LIDO, CIDOC CRM, MARC or EAD can thus be preserved whereas access by implicit top-level relationships will guarantee unprecedented recall and precision across data provided in various formats.

The particular challenge was to find a set of semantic relationships that are highly relevant to structure and access information about social-cultural artefacts, that are abstract enough to cover and subsume potentially thousands of more specialized relationships and that are yet expressive enough to close the current precision gap between keyword search and access by discipline specific fields. Core schemata like DC, VRA and the CIDOC CRM ontology are highly generic but no one covers the other sufficiently.

Excluding management of identity, we finally distinguish five fundamental semantic relationships which are further specialized, for instance by Dublin Core:

- Classification into categories that can be expressed with SKOS.
- Part decomposition of anything and incorporation of information resources into another one.
- Similarity, i.e. the relation between things or information resources that share some common features by chance, by influence or by a related derivation history as described by FRBR (Doerr and LeBoeuf, 2007).
- Aboutness, i.e., the entities or ideas a thing or information resource represents, presents, refers to or is about.
- History of an item, i.e., the things, people, places, times, events something had contact with, has existed at, “has met”. More analytically, all historical relationships can be explained and expanded as presence in events and the related event parameters.

All referred values of the above relationships may be represented as URIs into adequate Linked Open Data, for instance VIAF (www.viaf.org) for people, and gazetteers for places. People may also be represented by the FOAF ontology and by their relations to events in the EDM.

4 EDM and community specific representation schemas

The “Web language” RDF allows for declaring subsumption between properties, in other words, that a relationship used to associate some item with a particular value implies (is “subproperty of”) more general relationships to that value, exactly as “broader terms” in a thesaurus would do. Consequently, one can search for an item with this association under a more general relationship that was not declared in the data. This feature of RDF has not yet widely been exploited by user communities designing metadata formats, but actually is the most important means to integrate community specific representation schemas with more generic or “core” schemas. For instance, “dcterms:created” implies or refines “dc:date”. The EDM provides the most radical generalizations of metadata properties proposed so far.

Europeana foresees that any community may declare an Application Profile (Heery and Patel 2000) to enrich the precision of EDM for their particular subset of data. All generally relevant associations made in such Profiles should explicitly refine one or more properties of the EDM, and thereby ensure recall even for request by users not aware of the community-specific semantics. Vice-versa, Europeana has spent much effort to collect realistic metadata requirements from communities, i.e. metadata formats underpinned by large sets of actual data that can be provided to Europeana, and to verify their compatibility with EDM. Ultimately, the EDM is the result of generalizing over this input.

The relevant communities for Europeana can be subdivided into the libraries, digital libraries, audio-visual, archival and museums sectors. The museum sector currently has the highest internal diversity both of sub-disciplines and metadata formats, which is equalled by the diversity of physical things they administer. To a certain degree museums expose minimal metadata in Dublin Core. In the US, more popular is VRA and CDWA, as it captures the physical location of things. All three formats are subsumed by the EDM. However, satisfactory museum metadata are very complex, and a general agreement has been made in the form of the CIDOC CRM (ISO21127:2006), which declares rich common semantics of metadata elements in an RDF compatible form, but does not, by itself, prescribe the use of any fields. Only recently, most important international stakeholders in the museum sector have agreed on an explicit format for harvesting museum data, called LIDO. It is CRM compatible and is already underpinned by the production of large data sets in this format by the European Project ATHENA. LIDO originates in an event-centric reformulation of CDWA-Lite. Data in LIDO can be transformed into EDM, but also into the far richer CRM. Hundreds of other museum formats can also be transformed into the CRM. The CRM in turn is subsumed by EDM. Therefore representatives of the museum community have voted for the CRM as community application profile for museums under the EDM. Dublin Core metadata can be produced by rules from CRM metadata (Kakali et al. 2007).

For the archival community, collection level descriptions such as EAD play a major role. They fit neatly under the EDM, in particular the notion of ore:aggregration allows for describing archival “fonds”. The International Council of Archives just started the discussion about a common conceptual model similar to FRBR or the CRM. In the meanwhile, with the CRM historical facts associated with archival contents can be described in more detail than just on the EDM level (Stasinopoulou et al. 2007). Further, collection-level descriptions in Dublin Core are quite convenient and becoming popular for archival descriptions. For the library sector, MARC is still the dominant format, but many metadata can also be provided in the simpler MODS format, which could be used as application profile under the EDM, but Dublin Core is already a close match to MODS records. With FRBR, the library sector turns its interest to richer metadata since the late 1990ties. Even though there are many implementations of FRBR, the precise semantics of “Works” and “Expressions” are still controversial and the practical uptake in libraries is delaying. However, the need for aggregating content along derivation hierarchies is unquestioned and can be implemented without the controversial classification of the aggregation levels. FRBRoo, the ontological interpretation of FRBR (Doerr and LeBoeuf, 2007), has identified core relationships subsumed by the EDM “similarity” properties. Therefore, EDM can indeed represent uncontroversial core notions of FRBR.

5 Object-centric and Event-centric approach

Quite naturally, the design of metadata schemata has concentrated mostly on directly assigning attributes to the object in the collection as “finding aids”. Behind this approach stands a centuries-old tradition of good and highly successful librarianship. It is based on the paradigm that the user knows a topic the object is about, a material property or some associations of the object with a related item, such as the author or editor. It further assumes, that once the user gets access to the object, his information needs will be satisfied by the object itself. This paradigm can and has been completely transferred to digital collections, even more, since in addition to metadata access retrieval by content is available.

With the time our information systems are getting more powerful, people start to recognize that metadata have an enormous documentary information value in their own right. Also, under the extraordinary large scale of modern collections, increasing synonymy limits significantly the power of simple attributes to identify things. Finally, people start describing huge collections of “non-verbose” objects, such as images or objects in a museum, the understanding of which may completely depend on the metadata.

This caused a raised interest in providing more expressive and coherent records of the provenance and histories of objects. Since the mid 1990’s, several communities world-wide began to understand (IndeCs, CRM, ABC, OPM) that the complexity of describing historical relationships can be normalized and dramatically be simplified, if, instead of putting the objects in the centre of the description, historical events, which “mediate” in a sense all dynamic relationships between people, things, time and space, are put in the centre of documentation. Besides providing more details, this allows for detecting with high precision objects that are related through a common history, which is synonymous to shared participation in events.

Describing explicitly events may in the worst case double the size of some metadata with frequently multiple attributes (such as “creation_date, creation_place, creator”) pertaining to a shared event, but it increases considerably expressive power. For instance, precisely seeking ancient Egyptian objects imported to Crete in Bronze Age cannot be solved in a generic way without event description. Event documentation is a requirement of the cultural heritage community. Object-centric and event-centric descriptions of the same item can be transformed into each other, with a controlled loss of precision in the object-centric direction. The Europeana model represents these transformations internally. The obvious trade-off between complexity and expressive power of both aspects can therefore be flexibly adjusted to the user needs and quality of ingested sources.

In this sense, the relationship “has met” is a powerful innovation of the EDM. It bridges elegantly “object-centric” and “event-centric” modelling, and extends the genericity of DC:date to the other fundamental categories of non-information resources (events, agents, places, material and immaterial objects) and thereby provides a semantic coverage achieved by no other data standard so far. This coverage guarantees extensibility without losing access by the fundamental semantic relationships implicit in most of specialized data fields.

6 Validation and Potential of EDM

The EDM had reached a stable state with version 5. The expert group felt that it would be necessary to start evaluating and validating the model at this stage in order to eventually include it in the specifications of the 'Danube' release of Europeana; validation was to be based on real life examples from a variety of different backgrounds.

Therefore, Europeana organized four "community meetings" with representatives from archives, audiovisual archives, libraries, and museums. Each community provided typical sample data from their collections. The aim was to find out how well the various community standards could be mapped to the EDM. The results were very encouraging. Of course, each community came up with different detail issues, but these mainly pertained to questions of display and retrieval. As a common top-ontology the EDM itself proofed very flexible and stable and is able to accommodate community specific classes and properties as specializations. This series of meetings will be concluded by an expert meeting in Pisa in June where feedback from the communities will be integrated into the EDM.

6.1 Validation

The archives delivered example files of finding aids for archival material encoded in EAD. The distinct feature of such archival descriptions is the deep hierarchical structure and strong focus on fine grained and contextual description. The EDM properties for part decomposition and incorporation demonstrated its ability to handle descriptions of collections which contain several levels of finer grained sub-descriptions where each intermediate level contains contextual information.

The museums mainly provided examples encoded in museumdat and LIDO. The strong event-based approach of museumdat/LIDO fit very well into EDM. The provided classes and event-centric properties offered modelling possibilities which were flexible enough to integrate the rich event-centric descriptions of LIDO by means of typing events and creating sub-classes and sub-properties as specializations of EDM ones. The museum community, however, suggested to replace EDM classes and properties with CRM ones wherever possible and to use CRM entities without a counterpart in the EDM part of a museum application profile.

The audiovisual archives constitute a very heterogeneous community which provides very diverse objects and applies a variety of different encoding standards. In the audiovisual realm, the entity which is described is often difficult to identify and so is its nature. There is no such clear focusing perspective as with Museums (LIDO) or Archives (EAD) in the AV community and their material contains much more digital born objects. Still the EDM proofed to be able to integrate the diversity and richness of the AV provided examples.

The library community provided a number of sometimes very complex examples, all of which were successfully mapped to the EDM. It was evident, however, that modelling librarian data in the EDM would benefit a lot from an extension of the model including the FRBR group 1 categorization into work, expression, manifestation and item. The librarian experts agreed that the introduction of RDA, once operational, would substantially intensify the need to include the FRBR categories, eventually as part of a community application profile.

The conclusion from all community meetings was that the EDM, once validated and implemented, will be a good integrating tool for modelling different cultural artefacts from very different kinds of backgrounds.

6.2 Potential

The new "Europeana Data Model" (EDM) will replace the "Europeana Semantic Elements" (ESE) which currently underlies the data space of Europeana. The EDM offers greater expressivity and flexibility and allows for richer and truly semantic representations of the millions of objects from all kinds of cultural heritage communities in Europeana.

In comparison with prior data models EDM realizes a very high level of abstraction. It is the most radical generalization of metadata properties in the cultural heritage area so far and it does not bind the representation of ingested metadata to one common schema. The EDM carefully integrates well-established ontologies like SKOS, Dublin Core, and FOAF in order to allow for rich and interoperable descriptions of Europeana objects. As a common top-level ontology it allows for integrating the distinct information perspectives and needs of the various communities providing data to Europeana and to preserve the original richness of community standards like LIDO, CIDOC CRM, MARC or EAD.

The EDM uses RDF(S) as its meta-model and URIs to identify structured information about cultural heritage objects. The structural modelling framework for the EDM ontology is provided by the OAI Object Reuse & Exchange (OAI-ORE) specifications. This open architecture of the EDM makes Europeana compatible with the Semantic Web paradigm and enables it to become part of the emerging Linked Open Data community. In fact, the EDM provides a migration path for cultural heritage institutions from their currently mostly closed information architectures to open, linked environments – for the benefit of both these institutions and the WWW community.

The validation attempt in the community meetings demonstrated that the EDM has the potential to successfully function as a common top-level ontology for many different kinds of more specialized data models from a various knowledge domains. The possibility for communities to agree on an application profile to increase the precision for their particular data collections in EDM should allow for relatively "loss-free" integration of any data model.

Finally, the EDM offers new use scenarios for the diverse data ingested in Europeana. For instance the EDM based architecture enables contextualization of object representations creating new heuristic scenarios for scholars – more specifically in the 'digital humanities' – but this would be the subject of a paper of its own …

7 Literature

- Cesare Concordia, Stefan Gradmann, Sjoerd Siebinga (2010): Not just another portal, not just another digital library: A portrait of Europeana as an application program interface. In: International Federation of Library Associations and Institutions 36(1), pp. 61–69. (http://dx.doi.org/10.1177/0340035209360764)

- Martin Doerr, Patrick LeBoeuf (2007): Modelling Intellectual Processes: the FRBR - CRM Harmonization. In: C. Thanos, F. Borri, and L. Candela (eds.): Digital Libraries: R&D, LNCS 4877, pp. 114–123, 2007. (First DELOS Conference on Digital Libraries, February 2007 Tirrenia, Pisa, Italy.)

- Thomas R. Gruber (1993): Toward Principles for the Design of Ontologies Used for Knowledge Sharing. In: Nicola Guarino and Roberto Poli (eds.): Formal Ontology in Conceptual Analysis and Knowledge Representation.

- Constantia Kakali, Irene Lourdi, Thomais Stasinopoulou, Lina Bountouri, Christos Papatheodorou, Martin Doerr, Manolis Gergatsoulis (2007): Integrating Dublin Core Metadata for Cultural Heritage Collections Using Ontologies. In: Proceedings of the International Conference on Dublin Core and metadata Applications (DC-2007), pp. 128-139. 27-31 August, 2007. Singapore.

- Rachel Heery, Manjula Patel (2000): Application profiles: mixing and matching metadata schemas. Ariadne 25. (http://www.ariadne.ac.uk/issue25/app-profiles/intro.html)

- Barry Smith (2003): Ontology. In: L. Floridi (ed.): Blackwell Guide to the Philosophy of Computing and Information, Oxford: Blackwell, pp. 155–166.

- Thomais Stasinopoulou, Lina Bountouri, Constantia Kakali, Irene Lourdi, Christos Papatheodorou, Martin Doerr, Manolis Gergatsoulis (2007): Ontology-based Metadata Integration in the Cultural Heritage Domain. In: Proceedings of the 10th International Conference on Asian Digital Libraries, pp.165-175. Hanoi, Vietnam, December 10-13, 2007.