Demystifying publishing to Europeana:
A practical workflow for content providers

George Pavlidis and Vasileios Sevetlidis

Athena Research Centre, University Campus at Kimmeria, GR-67100 Xanthi, Greece

Corresponding author: George Pavlidis (gpavlid@gmail.com)

ABSTRACT

Every organisation needs to introduce itself in a target audience, promote its products and establish its presence amongst other competitive parties. This also applies to cultural institutions such as museums, libraries, archives, private or public stakeholders. Usually their approach involves the use of IT and especially the Web. Through these technologies they document, digitise, study, disseminate and exhibit their cultural wealth. Of common practice is the usage of ontologies and metadata schemas to provide access to those assets. The European Commission adopted the vision of a single access point to the European Cultural Heritage, which would be open to the public. In order to achieve this vision the European Digital Library was founded and the Europeana project was established. Europeana became an organisation that provides technological expertise and legal support to those stakeholders who want to share their assets. In this paper we describe the network structure of Europeana, the metadata technology and finally the workflow that someone needs to comprehend and apply in order to become partner of Europeana as a content provider.

1. INTRODUCTION

Europeana is a cultural heritage organisation whose goal is to provide access to Europe's entire heritage, to bring cultures together, to offer different approaches and point of views of any individual event, to enchase the European identity and to create new ways for people to engage with their cultural history through one single access point, the Europeana portal. More precisely, Europeana aims to develop a European Digital Library containing digitised material about the European scientific and cultural heritage. In particular, the metadata description schema, known as EDM (Europeana Data Model) (Europeana, 2013a, 2013b), has adopted CIDOC-CRM core (Doerr et at., 2007), while the consortium emphasises the need for linking existed descriptions of the digitised material in EDM descriptions, according to the Linked Data approach (Haslhofer and Isaac, 2011; Isaac et al, 2012).

Europeana published a fact-list (Europeana, 2013c) stating that the network contains over 30 million records from more than 2300 European galleries, museums, archives and libraries. Those records consist of textual material such as books, journals, newspapers, letters diaries and archival papers, multimedia content such as paintings, photographs, music, videos, TV and radio broadcasts, films, and the latest addition of digital material the 3D objects. All those digital items are accessible by everyone through the Europeana portal, which supports 31 languages (Petras 2013).

The content that Europeana gives access to is delivered primarily through a network of cultural heritage institutions (see Figure 1), whose goals are to promote their content, provide better user experience, and cooperate with like-minded professionals across Europe. The content that they provide to Europeana is the metadata description of their collection of digital objects, which is displayed in the portal and links back to their site. Additionally, Europeana through its expertise is able to provide to its partners tools and technology on data modelling, guidance over digital artefact documentation, consultation on copyright issues, and usage metrics. The portal that contains and displays the digital cultural items constitutes a single access point for multiple purposes. Users can search and discover a great variety of information, without leaving the main portal. In other terms, the technology that integrates the content material meets the user expectations that in the fast, easy and accurate information discovery. The fact that a significant number of European cultural institutions joined the Europeana network, means that the idea of information sharing and re-use is spreading among the stakeholders.

2. EUROPEANA STRUCTURE

As every competitive organisation sets up a strategy for future expansion, and course correction, Europeana does just the same. Europeana, as published in its strategy plan, focuses primarily on four subjects: aggregation, facilitation, distribution, and engagement (see Figure 2) (Europeana, 2011a).

In order to gather more digitised content, standardise, ingest it, and finally to apply the technology of Linked Data, Europeana needs to expand its contributor network. This network is divided in groups with certain responsibilities. A content provider for Europeana is any organisation that provides digital content accessible via Europeana. The digitised content is the metadata description of cultural heritage objects, enriched with a permanent URL link. Europeana uses the link to point to the original source of the digital object, through the portal. Europeana only ingests and indexes the institutions' metadata, while the digital object remains at the original institution. A content provider usually doesn't communicate directly with Europeana in terms of content delivery. This task is up to an intermediary level of organisations, namely the aggregators. Aggregators are defined as organisations that collect metadata from a group of content providers and transmit them to Europeana. Aggregators gather material from individual organisations, standardise the file formats and metadata, and channel the latter into Europeana ac-cording to the Europeana guidelines and procedures. The Network is presented in Figure 3; Europeana is at the top of the hierarchy, in the middle layer are the aggregators who collect the provided content form the bottom nodes, which represent the content providers. Aggregators also support the content providers with administration, operations and training (Europeana, 2011b). Europeana developed a metadata schema, the Europeana Data Model (EDM) (Europeana, 2013d, 2013b) in order to comply with the Linked Open Data specifications (Heath and Bizer, 2011). Linked Data is about connecting related data that were previously unrelated throughout the Web. This simply means:

- The discovery of data may be more precise
- The reusability of data may be more efficient
- The information extraction may lead to machine generated knowledge

After the delivered content is transformed to the EDM metadata schema, then the aggregator transmits those data to Europeana via OAI-PMH (Lagoze et al., 2005), which in return exposes the collected metadata to search engines, making the content accessible on the web.

The Europeana Network is an open forum for experts, content providers and aggregators, providers

of technical, legal and strategic knowledge, researchers and the creative industries. The analysis of domain representation of Europeana Network members indicates that libraries and museums are well represented. Archives, audio-visual archives and publishers are certainly an area for development to get more members in the Network. Europeana Network tends to be more inclusive than its former state, the Council of Content Providers and Aggregators (CCPA), was. This forum represents the diversity of Europeana. It is part of the governance of Europeana itself with six elected Network Officers sitting on the Board of the Europeana Foundation (Bergman-Tahon et al., 2013).
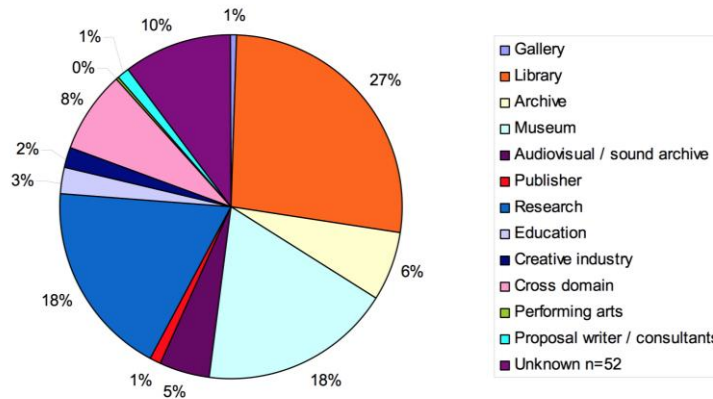

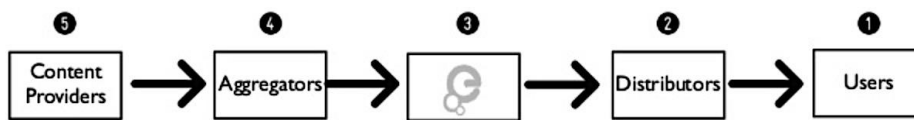**Figure 1. Europeana content providers affiliation**

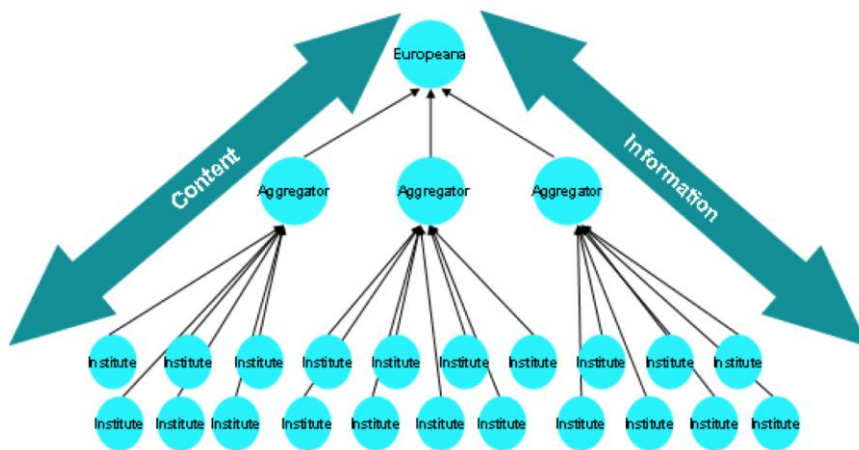**Figure 2. Subjects of Europeana strategic plan**

Figure 3. The Europeana Network

3. EUROPEANA SCHEMA

Taking advantage of digital era in cultural heritage domain usually means a series of actions that convert artefacts in digital duplicates. These actions are interpreted in digitisation of the artefacts, objects documentation, virtual exhibition and information sharing. Extensive research has been done for the documentation of an object. In the beginning, stakeholders were using rational databases. After the evolution of the Web, the need of information distribution emerged. So, the description of an object in Extensive Markup Languages (XML) structures came as a solution. A XML structure, provides several advantages, for example a XML structure offers the ability of creating custom tags, and defining object properties. When the structure of an XML schema is defined to fit a specific description e.g. the documentation of a cultural heritage object, then the structure

becomes a metadata schema. Bearing in mind that a metadata schema could be applicable on a specific situation, an extensive list of metadata standard schemas have been created. Europeana in order to comply with its needs, created —and later evolved— a metadata schema, that fits for the documentation of authorship, intellectual rights, as well as information sharing and re-use through Linked Open Data technology.

3.1. *Europeana Semantic Elements*

The Europeana Semantic Elements (ESE) provides a basic set of elements for describing objects in the cultural heritage domain in a way that is usable for Europeana. It is a Dublin Core-based application profile providing a generic set of terms that can be applied to heterogeneous materials thereby providing a baseline to allow contributors to take advantage of their existing rich descriptions. ESE produces a flat record, where it is not always possible to tell if a value applies to the original object or to its digital representation. The XML schema for ESE checks for basic conformance to this specification and gives instruction about the ordering of the XML elements (Europeana, 2013e). Some generic rules are considered in order to map resources into ESE:

- Map as many as possible of the original source elements to the available ESE elements.
- Always use the more specific *dcterms* refinements if the semantic of the source term clearly corresponds to the narrower term e.g. *dcterms:spatial* or *dcterms:temporal* instead of *dc:coverage*.
- Providers are encouraged to include *xml:lang* attributes in all appropriate metadata elements.
- The persistent link to the provided object should be given as a URL. These may need to be constructed from metadata values and information external to the metadata.
- If it is difficult to decide which ESE element to map a source term to, consider how best to meet expectations of the user and the functionality of the system.
- Where there are multiple values for the same element repeat the element for each instance of the value.
- Consider how the data would perform in response to "who, what, where and when" queries. This therefore encompasses names, types, places and dates.
- To ensure that the data will be meaningful when displayed in the new context, consider adding a prefix or suffix.

3.2. *Europeana Data Model*

The Europeana Data Model (EDM) (Europeana, 2013d) is the proposed data architecture for structuring the data that Europeana will be ingesting, enriching and publishing in the future. It improves the Europeana Semantic Elements. While ESE was offering a "flat" approach to the data, EDM, on the contrary, has the potential for accommodating the richness of all data standards, as well as the diversity of sectors represented in Europeana (museums, archives, audio-visual collections and libraries). EDM allows a clear distinction between the provided objects (painting, book, movie, archaeology site, archival file, etc.) and its digital representation, as well as between the object and the metadata record describing this object. It gives to the digitised object a new dimension, allowing multiple linking between its digital representations accessible over the web.

4.   TECHNICAL WORKFLOWS

In this section we try to provide cultural stakeholders with a comprehensive workflow, a reference toolkit, for two major Europeana possibilities. The first workflow summarises how an institution can become a *Europeana partner*, which is a first step to be able to publish data to Europeana. The second workflow is a comprehensive view of the steps to publish cultural content to Europeana.

4.1. *Becoming Europeana partner and Data Submission*

An institution has to follow some steps, to be able to submit data into Europeana. Partners contributing *indirectly* to Europeana will be provided with the necessary requirements for the next steps by the *aggregator* or *project* with which they are involved. The aggregator team offers assistance and guidance throughout the workflow. The steps involved when contributing content to Europeana are as described in Figure 4.

Step 1 - Partner Request to Europeana: The first step is to get in touch with Europeana. Contacting with Europeana is simple as visiting their website[1] and especially their contact form[2]. The stakeholder has to fulfil and apply a partnership request form[3]. Afterwards, an Aggregation team will be notified by Europeana and shortly after they will communicate with the stakeholder. The communication is about the possibilities of the provided content and guidance through the overall process.

---

[1] *http://pro.europeana.eu*
[2] *http://pro.europeana.eu/web/guest/contact*
[3] *http://pro.europeana.eu/provide-data*

Step 2 - Data Exchange Information: A partner institution interested in contributing data to Europeana will need to fill in the Data Exchange Information Form provided by the aggregation team. The form identifies types and numbers of objects, data owners etc. With this information, Europeana will discuss whether it is possible to provide the data directly or if the submission should go through other aggregators such as projects or national initiatives.
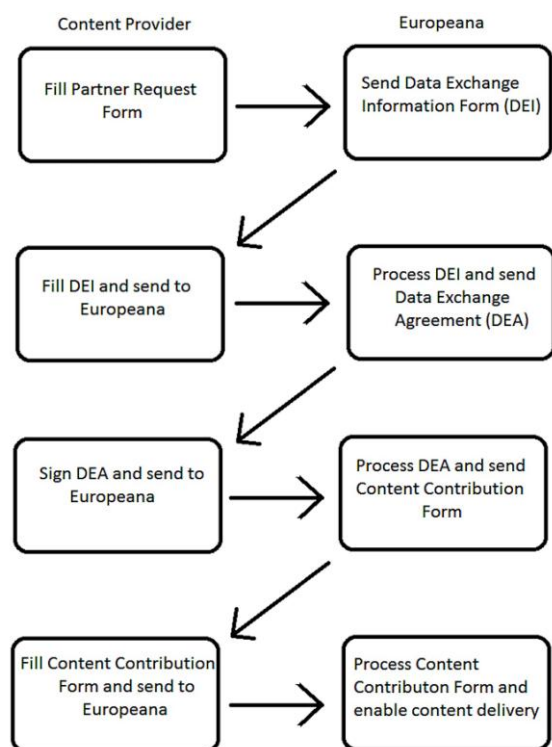

Figure 4. Partner's registration process

Step 3 - Data Exchange Agreement: The cultural heritage organisation and the Europeana Office formalise their direct collaboration by signing the Europeana Data Exchange Agreement[4].

Step 4 - Content Information Exchange: The cultural heritage organisation submits the Content Contribution Form (provided by the Europeana Aggregation team) containing: type of submission (new/update), licensing and metadata information, and the transfer mechanism that will be used (preferably OAI-PMH). The Europeana Aggregation team reviews the submission request and feeds back to the organisation.

Step 5 - Technical Guidance and Feedback: Alongside Step 4, the Europeana Aggregation team validates the data transfer infrastructure of the cultural heritage organisation. Subsequently, the team provides guidance and feedback on data collection

(harvesting) issues and metadata immigration, and formalisation (mapping). The submitted datasets must conform to the Europeana Semantic Elements (ESE) v3.4 or Europeana Data Model (EDM) specifications' and the Mapping and Normalisation Guidelines[5].

Step 6 - Publication: The Europeana Office runs the operations through the Europeana production environment and the publication process. When all processes are completed, the cultural heritage organisation receives a notification from the Europeana Aggregation team that the data has been published on the Europeana portal.

Step 7 - Update: When the cultural heritage organisation updates or creates new datasets, the Europeana Operations team will start again from step 6.

4.2. Data submission workflow

Each aggregator can receive the provider's resources in a metadata schema that meets its expectations, for example project "OpenUP!" makes use of the ABCD(EFG) schema (OpenUp! Project, 2013), the CARARE project (Fernie et al., 2013) has an enhanced subset of the CIDOC-CRM schema (Doerr et al., 2007), and the ATHENA project has created the LIDO (Cobum et al., 2010) (also a CIDOC-CRM subset). The workflow that defines the process of data submission may vary from case to case. Every aggregator has the liberty to expand the general outline of the submission workflow, in order to fit its expectations. The general outline consists of 5 steps as presented in **Figure 5** and described below.

Step 1 - Data mapping: the content provider may or may not provide resources directly in aggregator's preferable metadata description schema. As a consequence, it is aggregators' responsibility to convert the provided data to the native schema, and this process is called Metadata Mapping.

Step 2 - Data quality check: the content providers should check the quality of their resources before submitting them. For the purpose of quality control, providers should use specific software tools that aggregators have provided with. If the submitted data don't comply with aggregator's quality standard, then they have to be corrected and resubmitted.

Step 3 - Repository: a toolkit that serves as a data collector (harvester) stores the mapped metadata into specialised aggregators' central databases. Those databases of metadata descriptions are being called repositories.

Step 4 - Mapping to EDM (or ESE): the metadata that are mapped in aggregator's native schema are

---

transforming to either ESEv3.4 or EDM metadata schema.

Step 5 - Transmission of data to Europeana: Europeana collects the metadata from the aggregators that are prepared via transfer protocols such as OAI-PMH, or via FTP requests.

5.   CASE STUDY: MUSICAL INSTRUMENT MUSEUMS ONLINE

The Musical Instrument Museums Online or MIMO delivered to Europeana more than 43,000 record items regarding musical instruments across six countries and it was the first time that an aggregator used the Europeana Data Model (EDM). MIMO is a project that reflects the process that a data provider needs to do in order to deliver data to Europeana (Europeana, 2011c). This section describes the process from the data providers' point of view, concerning the original data mapping, the data quality check, and the mapping from one metadata scheme to another.
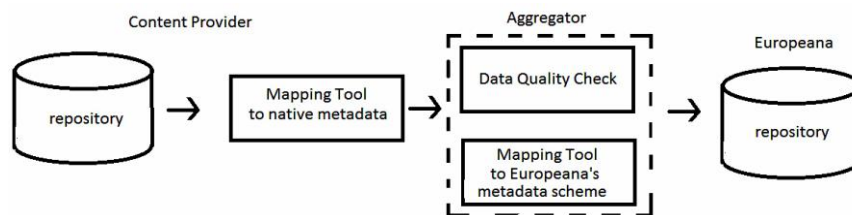

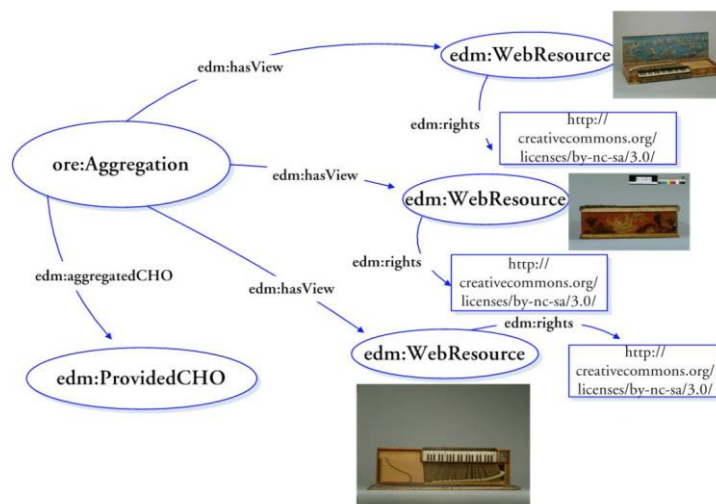**Figure 5. The process of content submission to Europeana**

Figure 6. Clavicorde lié dit de Lépante from Cité de la Musique

Initially, a data provider has to check whether the format that his data are being represented. In the case of MIMO, the original data were not according to the EDM, but they were being represented in LIDO's metadata scheme. Here is where Step 1 or the data submission workflow should be taken.

The data provider moves on to check if the data about to be delivered are in compliance with the aggregator's rules and data policies. EDM allows different resources to be connected to each other, in order to enrich the aggregated object. MIMO provided digital resources that span a wide range of multimedia content from audio-visual to images of musical instruments. Those objects could have more than one digital representation, as shown in an ex-

ample in Figure 6[6]. In order to comply with the EDM policy regarding usage and access rights, a decision was made concerning the identity of each object: the aggregator provided one *edm:rights* element to each *edm:WebResource*. This was during *Step 2 of the submission workflow*. In order to implement the data mapping MIMO had to link the LIDO properties to EDM properties. For this purpose MIMO used the EDM mapping template[7]. The process raised some issues connected with the use of literals and references in EDM's properties. Europeana does not use any controlled list to define its contributors, so

---

[6] *http://europeana.eu/portal/record/09102/33711D0A88B8AD 6D719112CDD9D37EE77526BC94.html*

[7] *http://labs.europeana.eu*

MIMO decided to use literals for the properties *edm:dataProvider* and *edm:provider*. In contrast, the properties *edm:isShownAt*, *edm:isShownBy*, and *edm:rights* support references, i.e. URIs taken from controlled lists. By the time that MIMO was about to deliver its data, the implementation of EDM retained the rights management approach of ESE (one rights statement per record). Generally, EDM intents to provide a separate rights statements per resource at the *ore:Aggregation* level and at the *edm:WebResource* level. With this option enabled the rights statements could be concurrent. Another objective that MIMO had to deal with was the further enrichment of the objects. MIMO had a lot of information additionally to the description that the objects provided themselves. Thus, two different approaches have been chosen:

- The first approach was to link the existing data to existing resources available in the semantic web (for example, in order to provide spatial information, MIMO used the contextual entity *edm:Place*, using GeoNames[8] to match a URI regarding the coordinates of a location related to the object).
- The other approach was to publish a new resource.

MIMO as an aggregator itself, created a domain-specific thesaurus using the standard SKOS model. This kind of information was linked to EDM via the entity *skos:Concept*. The metadata representation was in flat XML structure using the RDF syntax, due to the fact that EDM XML schema was not stable during MIMO's implementation.

6.   CONCLUSION

Organisations that promote access in information have an interest in reaching out more users and increasing their website trafficking. Cultural heritage organisations have the same interests. Nowadays, users are attracted by the functionality that a service offers and the interface it provides. Europeana is an organisation that aims to develop a European Digital Library that contains digitised material about the European scientific and cultural heritage. Europeana through its network of content providers, and generally its partners emphasises in innovative ways to explore and exploit the current knowledge, and transfer it to the public domain. By integrating the Europeana API, organisations can give users access to an unparalleled source of rich content. Professionals in the heritage sector might be interested in Europeana because it is a platform for knowledge exchange between librarians, curators, archivists and

the creative industries. Also, Europeana is a prestigious initiative endorsed by the European Commission, and is a means to stimulate the creative economy and promote cultural tourism.

There are many reasons why providing data to Europeana is beneficial for an institution. Here are just some of them:

- Europeana enriches users' experience; users can find not only an institution collections but also related information held in other countries, or in other formats, so the data, the content and the institution gain in visibility from association with linked material.
- Users today expect data to be integrated; to be able to see videos, look at images, read texts and listen to sounds in the same place
- Europeana will expose metadata to search engines, making deep web content accessible.
- Europeana drives traffic to the institution's site by linking users back to your website.
- Europeana provides a set of APIs through which the content of Europeana may be reused or returned in its enriched form by Europeana partners.
- Knowledge transfer is a key reason for being part of the Europeana Network. Europeana gives an opportunity to keep up-to-date with leading thinkers and practitioners in the fields of metadata standards, multi-linguality, semantic web, usability, geolocation and others.

The process of being Europeana partner is constantly being simplified and every cultural institution is worth considering being a member of Europeana Network. Europeana makes effort to automate the process of content submission, in order to be easeful and open to both partners and individuals, who want to share their information.

---

[8] *http://www.geonames.org*

ACKNOWLEDGEMENTS

REFERENCES

Bergman-Tahon, A., Friberg, A., Wittgren, B., Urtegaard, G., Verwayen, H., Scholz, H., Oomen, J., Edwards, L., and Poole, N. (2013) Europeana network - past, present, future, *Europeana Tech. Rep.,* Available: http://pro.europeana.eu/documents/297450/1505773/Europeana+Network+-+Past-Present-Future.pdf

Cobum, E., Light, R., Mvkenna, G., Stein, R., and Vitzhum, A. (2010) Lido - lightweight information describing objects version 1.0, *ICOM-CIDOC Working Group Data Harvesting and Interchange Tech. Rep.,* Available: http://www.lido-schema.org/schema/v1.0/lido-v1.0-specification.pdf

Doerr, M., Ore, C.E. and Stead, S. (2007), The CIDOC conceptual reference model: A new standard for knowledge sharing, *in Tutorials, Posters, Panels and Industrial Contributions at the 26th International Conference on Conceptual Modeling - Volume 83, ser. ER '07.* Darlinghurst, Australia: Australian Computer Society, Inc., 2007, pp. 51–56, Available: http://dl.acm.org/citation.cfm?id=1386957.1386963

Europeana (2011a) Strategic plan 2011 - 2015, *Europeana Tech. Rep.,* Available: http://tinyurl.com/7ythxty

Europeana (2011b) The Europeana Aggregator Handbook, *Europeana Tech. Rep.,* Available: http://pro.europeana.eu/documents/858566/858665/Aggregators+Handbook

Europeana (2011c) EDM Case Study: MIMO and EDM, *Europeana Tech. Rep.,* Available: http://pro.europeana.eu/mimo-edm.

Europeana (2013a) The Europeana data model for cultural heritage, *Europeana Tech. Rep.* Available: http://pro.europeana.eu/c/document_library/get_file?uuid=ef2baffc-f078-41d9-be5f-76a3427f198f&groupId=51031

Europeana (2013b) Europeana Data Model Primer, *Europeana Tech. Rep.,* Available: http://pro.europeana.eu/documents/900548/770bdb58-c60e-4beb-a687-874639312ba5

Europeana (2013c) Europeana – facts and figures, *Europeana, Tech. Rep.,* Available: http://pro.europeana.eu/documents/900548/c861c263-9acf-4fd6-a174-736346419a4e

Europeana (2013d) Definition of the Europeana Data Model v5.2.5., *Europeana Tech. Rep.* Available: http://pro.europeana.eu/documents/900548/0d0f6ec3-1905-4c4f-96c8-1d817c03123c

Europeana (2013e) Europeana Semantic Elements Specification and Guidelines, *Europeana Tech. Rep.,* Available: http://pro.europeana.eu/documents/900548/2eee7beb-b9d8-4532-a089-8e8d6df38ce7

Fernie, K., Gavrilis, D., and Angelis, S. (2013) The CARARE metadata schema, v.2.0, *CARARE Tech. Rep.,* Available: http://carare.eu/eng/Media/Files/CARARE-metadata-schema-Version-2.0

HASLHOFER, B. AND ISAAC, A. (2011) data.europeana.eu: The Europeana linked open data pilot, *International Conference on Dublin Core and Metadata Applications*, vol. 0, Available: http://dcpapers.dublincore.org/pubs/article/view/3625

Heath, T. And Bizer, C. (2011) *Linked Data Evolving the Web into a Global Data Space*, Morgan and Claypool Publishers, Available: http://www.uni-koblenz-landau.de/campus-koblenz/fb4/west/teaching/ws1213/seminar-web-science/linked-data.pdf

Isaac, A., Clayphan, R., and Haslhofer, B. (2012) Europeana: Moving to linked open data, *Information Standards Quarterly*, vol. 24, no. 2/3, Available: http://eprints.cs.univie.ac.at/3489/

Lagoze, C., Sompel, H., Nelson, M., and Warner, S. (2005) Implementation guidelines for the open archives initiative protocol for metadata harvesting, *Open Archives Initiative,* Available: http://www.openarchives.org/OAI/2.0/guidelines.htm

Openup! Project (2013), Openup! to ESE/EDM documentation, *OpenUp! Tech. Rep.* Available: http://open-up.eu/node/1238

Petras V. (2013) Multilingual and semantic interoperability in cultural heritage information systems, *in W3C Multilingual Web Workshop*, Available: http://www.w3.org/International/multilingualweb/rome/slides/09-petras.pdf