

## Metadata aggregation via linked data: results of the Europeana Common Culture project

Nuno Freire<sup>1</sup>[0000-0002-3632-8046], Enno Meijers<sup>2</sup>[0000-0002-2884-3523], Sjors de Valk<sup>2</sup>[0000-0001-5997-1600], Julien A. Raemy<sup>3,4</sup>[0000-0002-4711-5759], Antoine Isaac<sup>5,6</sup>[0000-0001-9767-6979]

<sup>1</sup> INESC-ID, Lisbon, Portugal

<sup>2</sup> Dutch Digital Heritage Network, The Netherlands

<sup>3</sup> HES-SO University of Applied Sciences and Arts Western Switzerland, Haute école de gestion de Genève, Switzerland

<sup>4</sup> University of Zurich, Zurich, Switzerland

<sup>5</sup> Europeana Foundation, The Hague, The Netherlands

<sup>6</sup> Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

nuno.freire@tecnico.ulisboa.pt, enno.meijers@kb.nl,  
sjors@sjorsdevalk.nl, julien.raemy@hesge.ch,  
antoine.isaac@europeana.eu

**Abstract.** Digital cultural heritage resources are widely available on the web through the digital libraries of heritage institutions. To address the difficulties of discoverability in cultural heritage, the common practice is metadata aggregation, where centralized efforts like Europeana facilitate discoverability by collecting the resources' metadata. We present the results of the linked data aggregation task conducted within the Europeana Common Culture project, which attempted an innovative approach to aggregation based on linked data made available by cultural heritage institutions. This task ran for one year with participation of eleven organizations, involving the three member roles of the Europeana network: data providers, intermediary aggregators, and the central aggregation hub, Europeana. We report on the challenges that were faced by data providers, the standards and specifications applied, and the resulting aggregated metadata.

**Keywords:** data aggregation; linked data; datasets; cultural heritage, Semantic Web.

## 1 Introduction

Nowadays, digital cultural heritage (CH) collections from libraries, museums and archives are widely available on the web. Many of these collections do not contain natural language texts (e.g., pictures, videos, music), and others that are of textual nature often lack machine-readable representation that can be used for indexing by search engines. In order to make these resources findable, CH institutions have traditionally turned to creating and exploiting metadata (that is, data records describing the resources).

CH comprises very diverse, transnational communities, which results in scattered collections that use many resource description standards and data models that are

specific to a context (e.g. a country or a set of institutions). Discoverability of CH resources is typically addressed by forming networks, where a central organization (a CH institution or other kinds of organizations) provides search and access services based on collecting the metadata associated with these resources. Such central organizations are in a position to better enable wider discovery and reuse of the resources, by applying practices and technologies that cannot be applied sustainably at the level of each single digital collection.

Within CH, this collecting approach, called metadata aggregation, uses data aggregation technologies that are different from the technologies used in other domains, such as by internet search engines or in the Web of Data. The Open Archives Initiative Protocol for Metadata Harvesting [1] (OAI-PMH) has been the dominant technology, since it is specialized for the aggregation of datasets of metadata.

Business models for the scalable and sustainable metadata aggregation have been implemented in the CH domain by organizations and networks such as Europeana in Europe, DPLA in the United States of America, Trove in Australia, Digital Library of India, and DigitalNZ in New Zealand. The implementation of such networks is costly, however.

In the meantime, the CH community has spent a lot of effort on redefining the traditional metadata models for cultural heritage resources into novel models based on semantic technology. Nowadays, many digital libraries and online catalogs publish metadata about CH resources as linked data (LD).

Our research sets out to investigate the feasibility of using LD for aggregation, which may bring cost benefits. If aggregators were able to make use of the available LD, data providers already making available LD would more easily share their metadata with cultural heritage aggregators. And for providers that are not yet publishing LD, participation in CH heritage aggregation based on LD would make the adoption of LD publication more valuable. They would reuse the LD with other application cases than aggregation, and with other domains besides cultural heritage, as for example with Internet search engines (if Schema.org, the data model used by search engines for harvesting structured data on the web<sup>1</sup>, is also used for LD publication).

This paper presents a case study on the LD aggregation task conducted within the Europeana Common Culture project<sup>2</sup>. This task ran from May 2019 to June 2020, and involved eleven organizations, representing the three member roles of the Europeana network: data providers, intermediary aggregators, and the central aggregation hub, Europeana. We report on the challenges that were faced by data providers, the standards and specifications applied in the case study, and the aggregated LD that resulted at the end of the task.

We follow, in Section 2, by describing related work on LD aggregation. Section 3 presents the requirements that underlie the design of our method of LD aggregation in

---

<sup>1</sup> For more information about the role of Schema.org for structured data on the web, please consult: <http://schema.org/docs/about.html>

<sup>2</sup> For information about the Europeana Common Culture project, consult <http://pro.europeana.eu/project/europeana-common-culture>

Europeana, and Section 4 follows with an overview of our workflow. Section 5 presents the software toolset that was implemented. Section 6 presents the results and their analysis. Section 7 highlights the main conclusions and presents future work.

## 2 Related Work

LD has received attention from many CH researchers and practitioners. However, we find in most cases that the main focus of the literature concerns the publication of LD [2–4] and does not investigate in detail how the metadata aggregation can be deployed on top of available LD<sup>3</sup>. One noticeable exception is the Research and Education Space project (RES) that finished in 2017 and has successfully aggregated a considerable number of LD resources from the CH, education and academic data sources. The resulting aggregated dataset can be accessed online, but an evaluation of its aggregation procedures and results has not been published. The project’s available technical documentation [5] addresses some of the required functionality that is relevant to our work. Some tasks, however, were not fully documented in the final specifications published by the project.

Solutions have been proposed by others for aggregation of LD (for example [6]) that tried to tackle the issue with generic solutions. None of the work in this area resulted in a standardized approach, however, and we could not find any sustainable application within cultural heritage.

The work we present here was conducted in the context of a line of research within Europeana, which aims at improving the efficiency of its network and sustainability. Our earlier studies on this topic identified LD as a potential technical solution for these objectives in metadata aggregation [7].

An important point is that LD provides a technological basis for metadata aggregation, which brings new options for data synchronization and data modelling. Some of the participants of the Europeana Common Culture LD task have recently engaged in a smaller pilot (the National Library of the Netherlands, the Dutch Digital Heritage Network (NDE) and Europeana Foundation) [8], whose positive results have led to the setup of this larger scale LD aggregation that we describe in this paper. NDE, the technical lead in our current task, is a Dutch national program aiming to increase the social value of the collections maintained by libraries, archives and museums in the Netherlands by improving their visibility, usability and sustainability. Designing and implementing a discovery infrastructure for LD is one of NDE’s main efforts [9].

Our case study focused on a scenario of LD aggregation using mostly the Schema.org vocabulary [10] to represent cultural heritage object (CHO) metadata. Schema.org is an initiative which encourages the publication and consumption of structured data on the web. It is a cross-domain vocabulary originally created by the major Internet search engines, but nowadays evolves as a community-based effort. Europeana has researched the best practices for publication of Schema.org CH metadata [11] and makes available Schema.org metadata within the webpages of its

---

<sup>3</sup> In fact the related LD work here has not changed since our earlier papers.

portal. Schema.org has also been applied successfully in our earlier LD pilot [7]. Also directly related to our current work is our earlier evaluation of Schema.org usage in CH institutions for aggregation by Europeana [12], whose positive outcome has been a support for also researching Schema.org for LD aggregation.

### **3 Requirements of the Europeana network**

This task was conducted under the current requirements for metadata aggregation in the Europeana network, which are based on the metadata harvesting functionality of OAI-PMH, and the Europeana Data Model (EDM) [13].

Although LD, by definition, uses well-known standards, these standards are defined for establishing wide interoperability of data. When building a solution based on LD for a particular interoperability use case, such as the one of Europeana, additional specifications are required. These serve the purpose of specifying how particular functional and informational requirements should be fulfilled by LD publishers and consumers. In the case of Europeana, such specifications guide how data providers and aggregators should exchange metadata in a LD-based solution.

In this case study, we have further elaborated the specification for defining a LD dataset for Europeana, from earlier work [8]. In short, the revised specification [14] allows data providers to provide dataset-level metadata as a LD resource, and makes use of well-known vocabularies for this kind of metadata. A key aspect of this specification is that it includes the types of dataset distributions that data providers can use, including the required metadata for each kind of distribution.

The significant number of data providers interested in participating in the LD aggregation task, motivated the creation of another specification, which addresses the use of Schema.org LD for describing CHOs according to the requirements of Europeana and EDM [15]. Currently, Europeana only supports ingestion of metadata in EDM, but experiments on applying Schema.org to metadata descriptions of CHOs have shown that it can provide good quality data that is capable of fulfilling the requirements of Europeana. This specification provides a general level of guidance for usage of Schema.org metadata that, after conversion to EDM, will result in metadata that is suitable for aggregation by Europeana. This specification also guides the functionality of the Schema.org to EDM converter that is part of the toolset developed in the project. This converter is described in Section 5.

A non-functional requirement also drives the design of our solution. This line of research aims at improving the efficiency and sustainability of the Europeana network. Whenever possible, we applied mature software supported by a community of developers and users instead of developing new software.

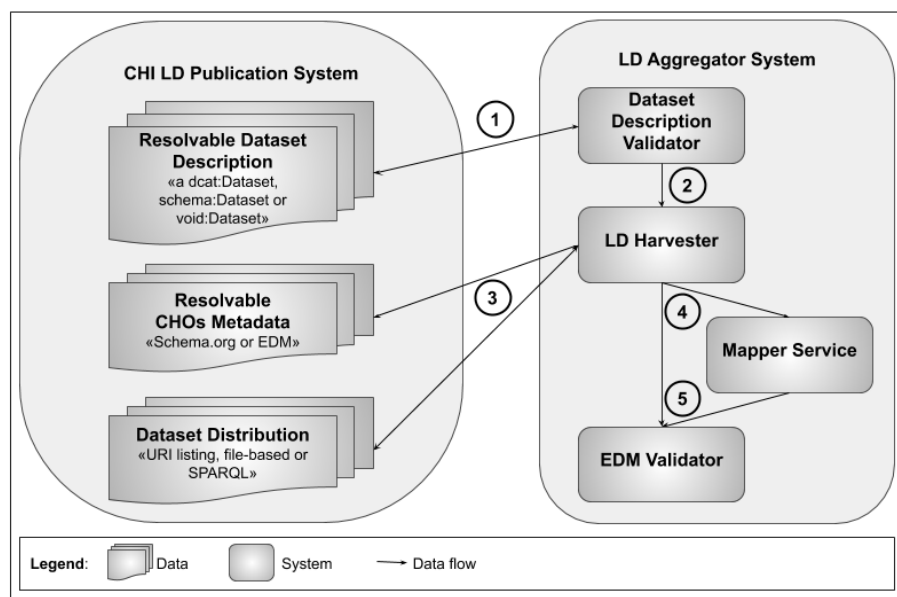
### **4 Overview of the approach to linked data aggregation**

Our approach to LD aggregation is based on the publication of several LD resources by data providers and in aggregation systems run by aggregators as well as Europeana. Data providers must publish LD about their CHOs and also a LD

description of the dataset these CHOs belong to. The systems of aggregators must be able to access and interpret the providers' dataset descriptions and apply the appropriate LD harvesting mechanism to collect the metadata about the providers' CHOs.

Figure 1 presents additional details of this approach. The data provider's LD publication must comprise the following:

- Resolvable CHO metadata - the CHOs that comprise the dataset to be aggregated must have resolvable URIs that lead to RDF data according to HTTP content negotiation [16]. In this case study, three data models were used for the CHOs' metadata: EDM, Schema.org and a "flattened" subset of EDM.
- Resolvable Dataset description - the dataset to be aggregated must be described as a LD resource and have a resolvable URI. The dataset's metadata must have a title and specify the mechanism through which aggregators can harvest the dataset. For dataset's metadata, data providers may use three data models: DCAT [17], VoID [18] and Schema.org.
- Dataset Distribution - the distribution mechanism for the dataset, such as a data dump (file), a SPARQL endpoint, or a listing of resource URIs.



**Fig. 1.** High-level view of the linked data aggregation approach and its data flow.

The systems of an aggregator comprise the following:

- Dataset Description Validator - a software component able to validate dataset descriptions according to the Europeana requirements.
- LD Harvester - a software component to transfer the datasets from data providers into the aggregator's system. The harvester must be able to

interpret the providers' dataset descriptions and implement all the distribution mechanisms covered in the Europeana guidelines.

- Mapper Service - a software component that implements a conversion of CHO's metadata from Schema.org to EDM, according to the Europeana guidelines.
- EDM Validator - a software component able to validate CHO's metadata according to the Europeana requirements.

Next to these, aggregators generally use a data repository to maintain the aggregated dataset. Our effort however focuses only on the steps that precede the loading of datasets into such repositories.

The flow of data for the aggregation of a LD dataset is also shown of Figure 1. It consists of the following five steps:

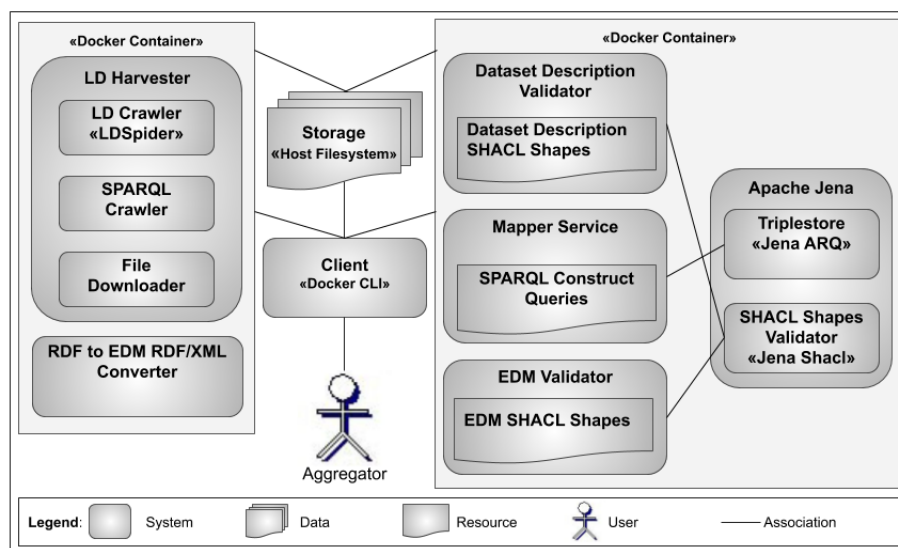
1. The aggregator validates the provider's dataset description for compliance with the Europeana guidelines.
2. If valid, the dataset description is passed on to the LD Harvester.
3. The LD Harvester reads the metadata about the distribution included in the dataset description and executes the appropriate harvesting mechanism for the type of distribution (described in Section 5).
4. When the CHO's metadata requires conversion to EDM, the Mapper Service converts it to EDM. When the metadata is in EDM, we proceed directly to the next processing task.
5. The EDM Validator verifies the compliance of the CHO's EDM metadata with the Europeana requirements.

After the last step the resulting (EDM) metadata can be passed to the aggregator's other systems for integration into its central aggregated dataset.

## 5 System implementation

Our task has produced a toolset that implements the workflow described in the previous section. The toolset was developed having in mind its usage by any aggregator in the Europeana network, therefore it makes use of mature software supported by a community of developers and users.

Figure 2 presents the high-level architecture of this toolset. The software components introduced in the previous section are implemented with a Docker configuration constituted by two independent Docker containers that host them. The exchange of data between the Docker containers happens via the file system (i.e., the file system of the Docker host is used by both containers). Docker containers preserve the technical independence of each of the tools, thus resulting in greater portability and flexibility of deployment. Aggregators can deploy only part of the toolset, according to their needs. Docker containers also provide greater scalability, enabling the toolset to be applied by aggregators with small and large collections.



**Fig. 2.** High-level architecture of the toolset for linked data aggregation in cultural heritage.

The components of the aggregation system are implemented as follows.

**Dataset Description Validator.** The validation of the dataset descriptions, in RDF, is performed via a SHACL Shapes [19] specification. For executing the validation, the Apache Jena SHACL validator is used.

**LD Harvester.** To transfer the datasets from data providers into the aggregation system, the LD Harvester has three subcomponents, each implementing one of the supported types of dataset distribution:

- **LD Crawler.** A LD dataset is harvested via HTTP content negotiation of the URIs specified in the dataset RDF description. The crawler may be configured to continue crawling any URIs found in the harvested data, and stop crawling when a configured depth<sup>4</sup> is reached. This component combines our own software for processing the dataset descriptions with the LDSpider [20] software for LD crawling.
- **SPARQL Crawler.** A LD dataset is harvested by first querying the SPARQL endpoint of the data provider with a query specified in the dataset description. The query outputs the list of URIs that are part of the dataset. This component combines software that interprets the dataset description, executes the SPARQL queries and commands LDSpider. The URIs obtained via the SPARQL query are used for the crawling process.
- **File Downloader.** A LD dataset is harvested via downloading of file-based distributions (also known as data dumps) that contain the complete dataset. This component was developed for this specific system. It interprets the

<sup>4</sup> The maximum depth is a parameter of crawlers that defines when they should stop following newly found links, based on distance from the resource where the crawl started.

dataset description, downloads and processes the files, uncompressing them if necessary, and transforms the RDF data into the RDF serialization that is used within the LD aggregation system (N-Triples).

**Mapper Service.** This component implements the conversion of CHOs' metadata from Schema.org to EDM. It follows the Europeana guidelines for this conversion, and allows for customization of the conversion for each collection. The conversions are specified as SPARQL construct queries [21], and the queries are executed on an Apache Jena triplestore, where the dataset has been imported, via the Jena ARQ API.

**EDM Validator.** The validation of the CHOs' metadata is performed via a SHACL Shapes specification. For executing the validation, the Apache Jena SHACL validator is used.

**RDF to EDM RDF/XML Converter.** The current data ingestion system of Europeana requires a particular RDF serialization, compatible with RDF/XML but guided by an XML Schema that Europeana applies to validate all datasets it receives<sup>5</sup>. Our converter transforms well-known RDF serializations into the RDF/XML required by Europeana.

The software is open for reuse and its source code is publicly available in a GitHub repository. Its most relevant parts are the source code of the interpreter of dataset descriptions and the LD Crawler<sup>6</sup>, the SHACL Shapes for validation of dataset descriptions and CHOs in EDM<sup>7</sup> and the conversion of Schema.org to EDM as SPARQL Construct queries<sup>8</sup>. Reusers of the software may adapt the mapping and validation functionalities by providing their specific SPARQL Construct Queries or SHACL Shapes definitions.

The key specifications for the LD aggregation process are published in Zenodo: the specification for defining a LD dataset for Europeana [14] and the guidelines for preparation of Schema.org metadata about CHOs for Europeana providers [15].

## 6 Results

An LD aggregation task was presented to all partners of the Europeana Common Culture project. Ten partners expressed interest in participating as data providers and, at a later stage, one additional external organization also joined<sup>9</sup>. Out of these eleven data providers, seven were able to provide a LD dataset but two of the provided datasets were delivered with insufficient CHO metadata. These were provided in Schema.org, but they lacked the required data elements for a successful conversion to

<sup>5</sup> For details consult the section 'EDM XML Schema and EDM validation in Oxygen' at <https://pro.europeana.eu/page/edm-documentation>

<sup>6</sup> <https://github.com/netwerk-digitaal-erfgoed/lod-aggregator/tree/master/crawler>

<sup>7</sup> <https://github.com/netwerk-digitaal-erfgoed/lod-aggregator/tree/master/shapes>

<sup>8</sup> <https://github.com/netwerk-digitaal-erfgoed/lod-aggregator/tree/master/queries>

<sup>9</sup> The data providers of the LD aggregation task were the following: erfgoedplus.be (Belgium), National Library of Finland, German Digital Library, National Documentation Centre (Greece), DigiPhil (Hungary), CulturalItalia (Italy), National Library of Latvia, National Library of Portugal, National Library of the Netherlands, Nationaal Museum van Wereldculturen (The Netherlands), UMA information technology gmbh (Austria).



EDM. Five datasets were aggregated successfully. Out of these, three were provided in Schema.org, one in EDM, and one in the "flattened" subset of EDM<sup>10</sup> (Table 1).

**Table 1.** The outcome of the data providers participation in the LD aggregation task.

Total data providers	Outcome			
	Not delivered	With insufficient CHO metadata	Delivered in Schema.org	Delivered in EDM
<i>11</i>	4	2	3	2

Our analysis of the unsuccessful cases is that most data providers were publishers of LD, but some were not applying Schema.org, or were using it for describing other entity types than CHOs (for example, for contextual entities such as agents, places and concepts) and underestimated the effort of creating Schema.org metadata for CHOs. The two cases whose CHO metadata was not complete correspond to the Schema.org output provided out-of-the-box by the providers' digital library system - Fedora<sup>11</sup>. Their metadata in Schema.org describes generic repository items (i.e. digital representations), lacking the descriptive detail required for CHOs (i.e. the original object). Both organizations considered the effort of implementing all the requirements for Schema.org to be beyond their available resources for our time frame. In the six unsuccessful cases the participation of the organization was indeed purely voluntary: they did not receive any resources to prepare and publish their metadata as LD.

Regarding the five successful cases, four data providers had previous experience with Schema.org or EDM for describing CHOs. They focused their implementation efforts on any aspects their LD was lacking. The fifth data provider published its metadata as linked data for the first time. It published it in EDM during the course of this task by re-using its past work on EDM metadata mappings for delivering EDM to Europeana via OAI-PMH.

The resulting data, after being harvested and processed by the LD infrastructure<sup>12</sup>, amounted to approximately 16.5 million triples, describing around 875 thousand CHOs (Table 2). Regarding the quality of the linked data, we observed that data providers use mostly literal values (i.e. simple strings) to describe the CHOs. RDF resources representing fully-fledged contextual entities (persons, concepts, etc) were only consistently found in the data of two providers.

---

<sup>10</sup> The LD published by the Nationaal Museum van Wereldculturen is not 100% valid according to the Europeana requirements, conversion into valid EDM was achieved by using the data conversion mechanism of our LD toolset.

<sup>11</sup> <https://duraspace.org/fedora/about/>

<sup>12</sup> Additional information regarding the computational resources applied and some measurements of the performance results can be consulted at: <https://github.com/netwerk-digitaal-erfgoed/lod-aggregator/tree/master/tests>

**Table 2.** Amount of aggregated data for the LD aggregation task.

Data provider	Number of Triples	Number of CHO instances	Number of Contextual Entities
<i>National Library of the Netherlands</i>	27,488	1,255	0
<i>National Documentation Center</i>	1,882	79	87
<i>National Library of Portugal</i>	2,458,660	105,365	2
<i>National Library of Finland</i>	1,205,754	37,195	19,157
<i>Nationaal Museum van Wereldculturen</i>	12,878,876	731,780	0
<b>Total</b>	<b>16,572,660</b>	<b>875,674</b>	<b>19,246</b>

## 7 Conclusion and future work

We have presented a case study where we have applied an approach for metadata aggregation via LD, which had previously been successful for a small-scale pilot. During this LD aggregation task of the Europeana Common Culture project, we have built a toolset for LD aggregation that is designed for deployment by aggregators of the Europeana network or other similar networks. Although the toolset includes functionality that is tailored for EDM (especially the specifications for conversion from Schema.org and data validation), aggregators using other data models may add their own conversions and validations using the standards implemented by the toolset - SPARQL Construct queries and SHACL Shapes.

The toolset is based on Docker containers, allowing to preserve the technical independence of its tools, making the solution more portable to different environments and more scalable, giving the possibility to apply the toolset to small or large collections.

The participation of data providers was voluntary, but many have shown interest in participating. Eleven data providers joined the LD aggregation task at the beginning but not all of them were fully aware of the technical challenges that this novel approach would bring. Four of the providers were not able to deliver a dataset as LD, and two other providers delivered datasets with insufficient data for aggregation into Europeana. In the five successful cases, four providers already had in-house knowledge or an existing implementation of LD, and for one, it was its first effort in publishing LD. Our overall conclusion is that there is much interest in implementing LD among data providers. Nevertheless, it requires a significant level of resources when the organization does not have any previous experience.

We have identified three areas for future work. First, we believe that data providers need supporting tools for preparing their LD. The validation tools implemented in the toolset may also be used in the creation of services for data providers, allowing them to check the validity of their data at earlier stages of LD publication. A second line of work should focus on components for interoperability and integration of the toolset into the aggregators' systems, so that our components may be called from within these systems. A third one could explore the publication of Schema.org data within

the web pages of the data providers' digital libraries. This approach is not based on LD content negotiation but on other technologies employed by Internet search engines, namely Sitemaps and Microformats<sup>13</sup>. Our toolset could be applied to these kinds of sources if it is complemented with a Sitemaps crawler and an HTML Microdata extractor. For data providers, this would bring the benefit of a common solution for making their metadata available for indexing by search engines and for aggregation by Europeana and other CH networks.

## Acknowledgements

We would like to acknowledge the support given by the participants of the pilot: Helena Patrício, Tetiana Kysla, Minna Rönkä, Osma Suominen, Haris Georgiadis, Agathi Papanoti, and Palko Gabor, as well as Erwin Verbruggen for his coordination. We also would like to acknowledge Adina Ciocoiu from the Europeana Foundation for the analysis of data samples collected during the case study.

## References

1. Lagoze, C.; Van de Sompel, H.; Nelson, M.L.; Warner, S.: The Open Archives Initiative Protocol for Metadata Harvesting, Version 2.0. Open Archives Initiative. (2002).
2. Jones, E.; Seikel, M. (eds): *Linked Data for Cultural Heritage*. Facet Publishing. 2016. ISBN: 9781783301621
3. Szekely, P.; Knoblock, C.A.; Yang, F.; Zhu, X.; Fink, E. E.; Allen, R.; Goodlander, G.: Connecting the Smithsonian American Art Museum to the Linked Data Cloud. *The Semantic Web: Semantics and Big Data*, pp. 593-607. (2013). doi:10.1007/978-3-642-38288-8\_40
4. Dragoni, M.; Tonelli, S.; Moretti, G.: A Knowledge Management Architecture for Digital Cultural Heritage. *Journal on Computing and Cultural Heritage (JOCCH) - Special Issue on Digital Infrastructure for Cultural Heritage, Part 2*, vol. 10, issue 3. (2017). doi:10.1145/3012289
5. McRoberts, M. (eds): *A guide to the Research & Education Space for contributors and developers*. BBC. (2016). Available online: <https://bbcarchdev.github.io/inside-acropolis/> (accessed on 29th of May 2019)
6. Vander Sande, M. ; Verborgh, R.; Hochstenbach, P.; Van de Sompel, H.. Towards sustainable publishing and querying of distributed Linked Data archives. *Journal of Documentation*, vol. 74, n. 1. (2018). doi: 10.1108/JD-03-2017-0040
7. Freire, N.; Robson, G.; Howard, J. B.; Manguinhas, H.; Isaac, A.: Metadata Aggregation: Assessing the Application of IIIF and Sitemaps Within Cultural Heritage. In: Kamps J., Tsakonas G., Manolopoulos Y., Iliadis L., Karydis I. (eds) *Research and Advanced Technology for Digital Libraries. TPDL 2017. Lecture Notes in Computer Science*, vol 10450. Springer, Cham. (2017). doi: 10.1007/978-3-319-67008-9\_18

---

<sup>13</sup> A Sitemap lists URLs of the pages within a website along with additional metadata about each URL so that search engines robots can efficiently crawl the site. Microformats are standards used to embed semantics and structured data in HTML.

8. Freire, N., Meijers, E., Voorburg, R., Cornelissen, R., Isaac, A., de Valk, S.: Aggregation of Linked Data: A case study in the cultural heritage domain. In: *Information*, 10(8), MPDI. (2019). doi:10.3390/info10080252
9. Meijers, E., de Valk, S., Helmus, W.: A distributed network of heritage information. White paper. 2018. (2018). Available online: <https://zenodo.org/record/1185401> (accessed on 15th of June 2019)
10. Google Inc., Yahoo Inc., Microsoft Corporation and Yandex; "About Schema.org", n.d., Available online: <http://schema.org/docs/about.html> (accessed on 29th of May 2019)
11. Wallis, R., Isaac, A., Charles, V., Manguinhas, H.: Recommendations for the application of Schema.org to aggregated Cultural Heritage metadata to increase relevance and visibility to search engines: the case of Europeana. *Code4Lib Journal*, Issue 36. (2017). ISSN 1940-5758
12. Freire N., Charles V., Isaac A.: Evaluation of Schema.org for Aggregation of Cultural Heritage Metadata. In: Gangemi A. et al. (eds) *The Semantic Web. ESWC 2018. Lecture Notes in Computer Science*, vol 10843. Springer, Cham. (2018). doi: 10.1007/978-3-319-93417-4\_15
13. Europeana Foundation: Definition of the Europeana Data Model v5.2.8. (2017). <http://pro.europeana.eu/edm-documentation>
14. Freire, N.: Specifying a linked data dataset for Europeana and aggregators (Version 0.2). Zenodo. (2020). doi:10.5281/zenodo.3817314
15. Freire, N.: Guidelines for providing and handling Schema.org metadata in compliance with Europeana (Version 0.1). Zenodo. (2020). doi:10.5281/zenodo.3817236
16. Sauermann, L., Cyganiak, R.: Cool URIs for the Semantic Web. W3C Interest Group Note. (2008)
17. Albertoni, R. Browning, D., Cox, S., Beltran, A.G., Perego, A., Winstanley, P.: Data Catalog Vocabulary (DCAT) - version 2.. W3C Recommendation. (2020)
18. Alexander, K., Cyganiak, R., Hausenblas, M., Zhao, J.: Describing Linked Datasets with the VoID Vocabulary. W3C Interest Group Note. (2011)
19. Knublauch, Holger; Kontokostas, Dimitris (eds.): Shapes Constraint Language (SHACL). W3C Recommendation. W3C. (2017).
20. Isele, Robert; Umbrich, Jurgen; Bizer, Christian; Harth, Andreas: LDSpider: An open-source crawling framework for the Web of Linked Data. In *Proceedings of 9th International Semantic Web Conference (ISWC 2010) Posters and Demos*, (2010).
21. Prud'hommeaux, Eric; Seaborne, Andy: SPARQL Query Language for RDF. W3C Recommendation. W3C. (2008).