# Metadata Aggregation: Assessing the Application of IIIF and Sitemaps Within Cultural Heritage


Nuno Freire[1(✉)], Glen Robson[2], John B. Howard[3], Hugo Manguinhas[4], and Antoine Isaac[4]

[1] INESC-ID, Lisbon, Portugal
nuno.freire@tecnico.ulisboa.pt
[2] National Library of Wales, Aberystwyth, UK
glen.robson@llgc.org.uk
[3] University College Dublin, Dublin, Ireland
john.b.howard@ucd.ie
[4] Europeana Foundation, The Hague, The Netherlands
{hugo.manguinhas,antoine.isaac}@europeana.eu



**Abstract.** In the World Wide Web, a very large number of resources is made available through digital libraries. The existence of many individual digital libraries, maintained by different organizations, brings challenges to the discoverability and usage of the resources. A widely-used approach is metadata aggregation, where centralized efforts like Europeana facilitate the discoverability and use of the resources by collecting their associated metadata. This paper focuses on metadata aggregation in the domain of cultural heritage, where OAI-PMH has been the adopted solution. However, the technological landscape around us has changed. With recent technological accomplishments, the motivation for adopting OAI-PMH is not as clear as it used to be. In this paper, we present the first results in attempting to rethink Europeana's technological approach for metadata aggregation, to make the operation of the aggregation network more efficient and lower the technical barriers for data providers. We (Europeana and data providers) report on case studies that trialled the application of some of the most promising technologies, exploring several solutions based on the International Image Interoperability Framework (IIIF) and Sitemaps. The solutions were trialled successfully and leveraged on existing technology and knowledge in cultural heritage, with low implementation barriers. The future challenges lie in choosing among the several possibilities and standardize solution(s). Europeana will proceed with recommendations for its network and is actively working within the IIIF community to achieve this goal.



**Keywords:** Metadata · Cultural heritage · Metadata aggregation · Web technology · Standards




J. Kamps et al. (Eds.): TPDL 2017, LNCS 10450, pp. 220–232, 2017.
DOI: 10.1007/978-3-319-67008-9_18

# 1 Introduction

In the World Wide Web, a very large number of resources are made available through digital libraries. The existence of many individual digital libraries, maintained by different organizations, brings challenges to the discoverability and usage of the resources by potentially interested users.

An often-used approach is metadata aggregation, where a central organization takes the role of facilitating the discovery and use of the resources by collecting their associated metadata. Based on these aggregated datasets of metadata, the central organization (often called aggregator) is in a position to further promote the usage of the resources by means that cannot be efficiently undertaken by each digital library in isolation. This scenario is widely applied in the domain of cultural heritage (CH), where the number of organizations with their own digital libraries is very large. In Europe, Europeana has the role of facilitating the usage of CH resources from and about Europe, and although many European CH institutions do not yet have a presence in Europeana, it already holds metadata from over 3,500 providers[1].

In this paper, we present the first results of our work on rethinking Europeana's technological approach for metadata aggregation. Our goal is to make the operation of the aggregation network more efficient and lower the technical barriers for data providers to contribute to Europeana. Our approach was to undertake case studies with real CH collections and systems from data providers of the Europeana network. The case studies were based on promising technologies, which have been identified in our previous work. The results achieved make the following contribution to digital library research:

- A functional analysis for innovative use of state of the art technologies, based on a large network of data providers – the Europeana Network.
- A real-world application experience of open standards, thus contributing for their future improvement.

The paper will describe the technological approach to metadata aggregation most prevalent in CH in Sect. 2. Further details regarding metadata aggregation in CH are presented in Sect. 3. The series of case studies that were performed are presented in Sect. 4, along with their analysis. Section 5 summarizes the outcomes.

# 2 State of the Art: Metadata Aggregation in Cultural Heritage

In the CH domain, the technological approach to metadata aggregation has been mostly based on the OAI-PMH protocol, a technology initially designed in 1999 [1]. OAI-PMH was originally meant to address shortcomings in scholarly communication by providing a technical interoperability solution for discovery of e-prints, via metadata aggregation.

[1] Source: http://statistics.europeana.eu/europeana [consulted on 27th of April 2017].

The CH domain embraced OAI-PMH, since discovery of resources was only feasible if based on metadata instead of full-text [2]. In Europe, OAI-PMH was the technological solution adopted by Europeana since its start, to aggregate metadata from its network of data providers and intermediary aggregators.

However, the technological landscape around our domain has changed. Nowadays, with the technological improvements accomplished by network communications, computational capacity, and Internet search engines, the discovery of e-prints is largely based on full-text processing, thus the newer technical advances, such as ResourceSync [7], are less focused on metadata. Within the CH domain metadata-based discovery remains the most widely adopted approach since a lot of material is not available as full-text. The adoption of OAI-PMH for this purpose is not as clear as it used to be, however. OAI-PMH was designed before the key founding concepts of the Web of Data [3]. By being centered on the concept of repository, instead of focusing on resources, the protocol is often misunderstood and its implementations fail, or are deployed with flaws that undermine its reliability [2]. Another important factor is that OAI-PMH predates REST [4]. Thus, it does not follow the REST principles, further bringing resistance and difficulties in its comprehension and implementation by developers in CH institutions.

An additional aspect relevant for our work, is that CH institutions are increasingly applying technologies designed for wider interoperability on the World Wide Web. Particularly relevant are those related with Internet search engine optimization and the IIIF (International Image Interoperability Framework[2]). Regardless of the metadata aggregation process for Europeana, CH institutions are already interested in developing their systems' capabilities in these areas. By exploring these technologies, the participation in Europeana may become much less demanding for these institutions.

## 3 Characterization of Metadata Aggregation in Cultural Heritage

The CH domain has specific characteristics that influence how metadata aggregation is done. We consider the following to be the most influential:

- Several sub domains compose the CH domain: Libraries, Archives, and Museums.
- Each sub-domain applies its specific resource description practices and data models.
- All sub-domains embrace the adoption and definition of standards based solutions addressing description of resources, but to different extents. A long-time standardization tradition has existed in libraries, while it is more recent in archives and museums.
- Interoperability of systems and data is scarce across sub-domains, but it is common within each sub-domain, both at the national and the international level.
- Adopted standards tend to use XML-based data models, while models based on relational data are rare.

[2] International Image Interoperability Framework - http://iiif.io/#international-image-interoperability-framework.

- Organizations typically have limited budgets to devote to information and communication technologies, thus the speed and extent of innovation and adoption of new technologies is slow.

In this context, a common practice has been to aggregate metadata using an agreed data model that allows the data heterogeneity between organizations and countries to be dealt with in a sustainable way. The models typically seek to meet two main requirements: (a) retaining the semantics of the original data from the source providers; (b) supporting the information needs of the services provided by the aggregator.

These two requirements are typically addressed in a way that keeps the model complexity low, with the intention of simplifying the understanding of the model by all kinds of providers and to keep a relatively low barrier for both providers and aggregators to implement data conversion solutions,

Another relevant aspect of metadata aggregation is the sharing of the sets of metadata from the providing organizations to the aggregator. The metadata is transferred to the aggregator, but it continues to evolve at the data provider side, thus the aggregator needs to periodically update its copy of the data. In this case, the needs for data sharing can be described as a data synchronization problem across organizations.

In the CH domain, OAI-PMH is also the most well-established solution to address the data synchronization problem. Since OAI-PMH is not restrictive in terms of the data model to be used, it enables sharing of metadata for aggregation according to the data model adopted for each aggregation case. The only restriction imposed by OAI-PMH is that the data must be represented in XML.

In the case of Europeana, the technological solutions around the Europeana Data Model (EDM) (Europeana, 2016) have always been under continuous improvement. However, the solution for data synchronization based on OAI-PMH has not been reassessed since its adoption. In the case studies presented in the following sections we address mainly the data synchronization problem. Since the aggregation solution of Europeana is based on EDM, the data synchronization can be addressed with a wide variety of technologies because EDM follows the principles of the Web of Data, and can be serialized in XML and in RDF formats.

## 4 Case Studies

Our earlier work addressed the data synchronization problem by reviewing the state of the art and emerging Web technologies [5]. In the continuation of this work, we identified two key technologies, on which we based the case studies presented in this paper: IIIF (International Image Interoperability Framework) and Sitemaps[3].

IIIF is a family of specifications that were conceived to facilitate systematic reuse of image resources in digital repositories maintained by CH institutions. It specifies several HTTP based web services [6] covering access to images, the presentation and structure of complex digital objects composed of one or more images, and searching within their content. IIIF's strength resides in the presentation possibilities it provides for end-users.

[3] Sitemaps XML format: https://www.sitemaps.org/protocol.html.

From the perspective of data acquisition, however, none of the IIIF APIs was specifically designed to support metadata aggregation. Nevertheless, within the output given by the IIIF APIs, there may exist enough information to allow HTTP robots to crawl IIIF endpoints and harvest the links to the digital resources and associated metadata.

Sitemaps allow webmasters to inform search engines about pages on their sites that are available for crawling by search engine's robots. A Sitemap is an XML file that lists URLs of the pages within a website along with additional metadata about each URL (i.e. when it was last updated, how often it usually changes, and how important it is in comparison to other URLs within the same site) so that search engines can more efficiently crawl the site. Sitemaps is a widely-adopted technology, supported by all major search engines. Many content management systems support Sitemaps out-of-the-box and Sitemaps are simple enough to be manually built by webmasters when necessary. Moreover, there are Sitemaps extensions, like Google's Image Sitemaps[4] and Video Sitemaps[5], which have potential usage in metadata aggregation.

We have identified other promising Web technologies (for example, technologies related with the social web or the web of data [5]). We have chosen IIIF because it is getting increasingly traction in CH. Moreover, it is a community developed, open framework. Our requirements and suggestions for metadata aggregation may thus be incorporated into future versions. The choice for Sitemaps was motivated by its wide usage within the Europeana data providers. In addition, Sitemaps also provides a very simple technological solution, with a very low implementations barrier.

We have undertaken several case studies to investigate the feasibility of performing metadata aggregation via IIIF and/or Sitemaps. These studies were conducted in cooperation with data providers of the Europeana Network[6], which were actively deploying these two technologies within their own information systems.

IIIF played a double role in our work. It was used as the data source from where the source metadata was aggregated from providers, and as a technology that can be used with other suitable web-based technologies to facilitate aggregation processes. We have specifically studied how the functionality available in the IIIF Presentation API could be used to provide similar aggregation functionality as OAI-PMH and Sitemaps.

The following subsections will describe the case studies and how the two technologies were used for metadata aggregation.

### 4.1 Crawling IIIF Services from IIIF Inventories Using the Presentation API

The first case study was exploratory and targeted at IIIF in general. It was performed solely by Europeana with the objective of evaluating the functional capabilities of IIIF, the amount of data sources available, the maturity and compliance of the IIIF

[4] https://www.google.com/schemas/sitemap-image/1.1/.

[5] https://developers.google.com/webmasters/videosearch/sitemaps.

[6] The source code of the prototypes developed in the case studies is openly available at https://github.com/nfreire/Open-Data-Acquisition-Framework.

implementations and the quality of the available metadata. For this purpose, a IIIF aware Web Crawler was prototyped and the results of crawling several IIIF services were evaluated. All source code, collected samples and results of this case study may be consulted online[7].

To find available IIIF services to crawl, we have identified two crowdsourced listings of existing IIIF services, both provided in machine readable ways: IIIF Top Level Collections[8] and the iiif-universe[9]. From these listings, we have chosen 13 collections containing CH resources. The crawler was pointed to these IIIF collections and attempted to fully harvest them.

Alongside the full harvest, the crawler extracted a sample of IIIF Manifests[10] from each service, for later analysis. The extracted samples were manually inspected for the availability of (references to) descriptive metadata, the semantic granularity of the model in which they are available, and the availability of machine readable licensing information for re-use.

This early exploration revealed that IIIF contains all the necessary elements for automatic metadata aggregation. Some of these elements are, however, not mandatory for implementation, thus they will not be available in every IIIF service. The following optional elements of IIIF APIs must be provided by data providers, to enable metadata harvesting for Europeana:

- Structured metadata: the metadata available in the output of IIIF (manifests) is intended for end-user presentation, thus it cannot fulfill Europeana's ingestion requirements. But this can be overcome by using the optional links to structured metadata as specified in IIIF (using a seeAlso property[11]). When these are correctly populated (which our study confirmed to be not always the case), they enable crawlers to obtain structured metadata, such as EDM, Dublin Core, etc.
- IIIF Collection indicating the objects for Europeana: In IIIF, it is not required that the endpoint implements a mechanism to make publicly known all the digital objects that it makes available.

The implications for data providers in the provision of these two aspects for Europeana was addressed in the case studies that follow.

### 4.2 Crawling IIIF Services via IIIF Presentation API Collections Made Available by Europeana Providers

The IIIF Presentation API offers a Collection construct to represent groups of objects[12]. Although not all IIIF services make Collections available, they are often provided.

7 https://github.com/nfreire/IIIF-Manifest-Metadata-Harvesting.

8 https://docs.google.com/spreadsheets/d/1apQKFkfBV89BvycaBPN6v-LjeaKaVVMaMUsY6L4KRJo/edit#gid=0.

9 https://github.com/ryanfb/iiif-universe.

10 http://iiif.io/api/presentation/2.1/#manifest.

11 http://iiif.io/api/presentation/2.1/#seealso.

12 http://iiif.io/api/presentation/2.1/#collection.

Even when Collections are not available, the implementation effort would be very low for the provider. By making a IIIF collection known to Europeana, all the resources it references can be crawled and their metadata harvested by Europeana.

In this case study, Europeana worked with two data providers from the Europeana Network: The National Library of Wales and University College Dublin. At the starting point of the cases studies, both organizations had IIIF services available, but neither of them had EDM metadata available through IIIF and neither had IIIF Collections available for Europeana.

Implementation of a IIIF Collection was easily achieved in both cases.

Regarding the implementation of EDM metadata, it was a straightforward task for the National Library of Wales, since the library had recently deployed an EDM conversion of their MODS[13] metadata for other purposes. University College Dublin was also successful in implementing an EDM conversion and including it in their IIIF endpoint output. However, in the case of University College Dublin, the support of the Europeana ingestion team was required for implementing the EDM conversion and obtaining valid EDM metadata.

We identified an additional issue for metadata aggregation from IIIF services - IIIF collections do not provide the modification timestamp of resources. This aspect has an impact in the efficiency of the harvesting process. It becomes relevant in very large collections with hundreds of thousands of resources, where re-harvesting of resources that have not changed should be avoided. Given the importance of efficiency for the aggregation of large datasets, this problem still needs to be addressed by Europeana (and the IIIF community in general). To overcome it, other technologies may be used in conjunction with IIIF. Examples are Sitemaps and HTTP headers, which we have evaluated in the use cases described in the remainder of this section.

### 4.3 Crawling IIIF Services Referenced by Sitemaps

The issue of harvesting efficiency identified in the previous case study has been brought to the attention of the IIIF community. Discussions have been started for achieving a standard mechanism or recommendations to address it within the IIIF framework, in the context of a new IIIF Discovery Technical Specification Group[14]. The general opinion among the IIIF community was that IIIF Collections were designed for different purposes, i.e., to support use cases of end-users interacting with the IIIF viewers. The use of Collections for metadata harvesting purposes is therefore not an optimal solution from the point of view of the design of the IIIF framework.

In this context of these discussions, we conducted a supporting case study, where we experimented with Sitemaps-based solutions – both standard Sitemaps and Sitemaps with extensions.

[13] Metadata Object Descriptive Schema (MODS) is a schema for a bibliographic element set: http://www.loc.gov/standards/mods/.

[14] http://iiif.io/community/groups/discovery/.

#### 4.3.1 Standard Sitemaps

The providers have created Sitemaps listing the specific resources that should be aggregated by Europeana. Europeana implemented a prototype for a IIIF harvester based on these Sitemaps, and the solutions were deployed and tested successfully with real datasets.

When using standard Sitemaps, the identifiers of the IIIF Manifests are present in the Sitemap as would any other URL pointing to a web page. That is, the Sitemap XML directly references the IIIF Manifest in the <loc> element, as shown in Fig. 1.

```
<url>
   <loc>https://data.ucd.ie/api/img/collection/ivrla:3573</loc>
   <lastmod>2014-08-24T04:09:09.716Z</lastmod>
</url>
```

**Fig. 1.** Example of URL data in a Sitemap from University College Dublin. The *loc* element references a IIIF Manifest.

This solution presents some shortcomings, however. The critical issue is that such Sitemap cannot be used for efficiently representing web pages and IIIF Manifests at the same time. The crawler would need to fetch the content of every URL in order to verify when it refers to a IIIF resource or to a webpage. We thus carried out further experiments with two alternative solutions based on extensions of Sitemaps: one with elements from the IIIF namespace and another with elements from the ResourceSync namespace [7].

#### 4.3.2 Sitemaps Extended with Elements from the IIIF Namespace

Our goal is to extend Sitemaps to better contextualize and relate the IIIF resource with the end-user access webpages of the digital library.

In our first extension, we make explicit the availability of the resource via IIIF, in the Sitemap, and make it possible to relate the resource with IIIF Collections and end-user webpages. The example in Fig. 2 contains the end-user access location from the digital library in the <loc> element. The link to the IIIF Manifest is made via a

```
<url>
   <loc>http://newspapers.library.wales/view/3679651</loc>
   <iiif:Manifest
xmlns:iiif="http://iiif.io/api/presentation/2/">http://dams.llgc.org.uk/iiif/newspaper/issue/36796
51/manifest.json</iiif:Manifest>
   <dcterms:isPartOf>http://dams.llgc.org.uk/iiif/newspapers/3679650.json<dcterms:isPartOf>
   <lastmod>2014-11-08</lastmod>
</url>
```

**Fig. 2.** Example of URL data in a Sitemap from the National Library of Wales, with references to the webpage of the resource, the IIIF Manifest and its IIIF Collection.

<iiif:Manifest> element[15] and a link to a IIIF collection which the resource belongs to) is made via a <dcterms:isPartOf> element.

As shown by the ResourceSync related research [8], these extended Sitemaps remain compatible with Internet search engines.

#### 4.3.3 Sitemaps Extended with Elements from the ResourceSync Namespace

Our last case study uses extensions from the ResourceSync namespace [7]. The end-user access location from the digital library in still referenced in the <loc> element. The link to the IIIF Manifest is made by a <rs:ln> element with the attribute 'rel' set to 'alternate', and a link to a IIIF collection made by a <rs:ln> element with the attribute 'rel' set to 'collection'. The values of the 'rel' attribute is defined in the ResourceSync specification. To explicitly state that these links lead to IIIF resources, the attribute 'conformsTo' from Dublin Core is also included in the <rs:ln> elements.

This extension provides the same expressive capabilities of the extension based on IIIF elements for relating all the relevant resources. Its main motivation is that by using elements from ResourceSync, we expect that the Sitemap can be better interpreted when used beyond the context of IIIF (Fig. 3).

```
<url>
   <loc>https://digital.ucd.ie/view/ucdlib:38491</loc>
   <rs:ln rel="alternate" href="https://data.ucd.ie/api/img/manifests/ucdlib:38491"
type="application/json" dcterms:conformsTo="http://iiif.io/api/presentation/2.1/"/>
   <rs:ln rel="collection href="https://digital.ucd.ie/view/ucdlib:38488”
type="application/json"  dcterms:conformsTo="http://iiif.io/api/presentation/2.1/"/>
   <lastmod>2014-08-24T04:09:09.716Z</lastmod>
</url>
```

**Fig. 3.** Example of URL data in a Sitemap from University College Dublin, with references to the webpage of the resource, the IIIF Manifest and its IIIF Collection, and the indication of the IIIF API version in use.

### 4.4 Crawling IIIF Services via the IIIF Presentation API and HTTP Cache Headers

Another option to solve the issue of harvesting efficiency was experimented between Europeana and the National Library of Wales. The idea is to extend the solution based on IIIF Collections by using HTTP cache control [9]. Here, the IIIF service is required to have the implementation of some HTTP cache headers for the URLs that provide access to the IIIF resources.

In this solution, the Europeana IIIF crawler must include, in all the requests for IIIF manifests, the HTTP header *If-Modified-Since,* which will contain the timestamp of the last time the resource was harvested. The IIIF service then only needs to send the IIIF

[15] IIIF Presentation Ontology: http://iiif.io/api/presentation/2.

```
<url>.
   <loc>https://digital.ucd.ie/view/ucdlib:38509</loc>
    <rs:ln rel="describedby"          href="https://data.ucd.ie/api/edm/v1/ucdlib:38509"
dcterms:conformsTo="http://www.europeana.eu/schemas/edm/"/>
   <rs:ln rel="collection" href="https://data.ucd.ie/api/img/collection/ucdlib:38488"/>
   <video:video>
        <video:thumbnail_loc>https://digital.ucd.ie/get/ucdlib:38509/thumbnail
        </video:thumbnail_loc>
          <video:description>Irish poet Catherine Ann Cullen reads her poem 'Meeting at the
Chester Beatty' in UCD Library's Special Collections.</video:description>
          <video:player_loc allow_embed="yes">
https://player.vimeo.com/video/111413587</video:player_loc>
         <video:duration>00:02:51.04</video:duration>
         <video:family_friendly>yes</video:family_friendly>
         <video:live>no</video:live>
   </video:video>
   <lastmod>2015-09-10T17:14:26.523Z</lastmod>
</url>
```


**Fig. 4.** Example of URL data using the Sitemaps Video extension from University College Dublin. The Sitemap was extended to allow the association of EDM metadata.

manifest if an update has happened since that time in the manifest, the metadata or the resource itself. In case of deletion of the resource, the IIIF service returns a response with the HTTP Status code 404 *Not Found*. The assumption behind this solution is that the IIIF service can efficiently query the timestamp of the resources it serves and, in this way, save time and processing resources by not having to assemble and transmit the IIIF manifest back to the IIIF crawler.

We measured the performance of the crawling process for a collection of 500 resources in National Library of Wales's IIIF implementation. The outcome was a reduction around of 50% in the total time for crawling the 500 resources when they were not modified. The measurements were made at several points in time, different time of the week and the day, to prevent strong measurement bias due to variations in the user load of the IIIF service (our experiment was *in vivo* in the sense that it used the production service of the National Library of Wales and thus had to share bandwidth with "real" users).

### 4.5 Crawling (Non-IIIF) Resources Referenced by Sitemaps Extensions: Video and Image

An additional case study was performed with Sitemaps extensions used for better retrieval of image[16] and video[17] content within the Internet search engines that focus on

[16] https://www.google.com/schemas/sitemap-image/1.1/.

[17] https://developers.google.com/webmasters/videosearch/sitemaps.

**Table 1.** Summary of the main conclusions taken from the case studies

| Method | Summary |
|---|---|
| Crawling IIIF Universe services | No contact or intervention required from the data provider. Only a very limited number of registered IIIF services provide the components of the IIIF API required for the metadata harvesting process |
| Crawling IIIF services based on IIIF Collections | Very simple to implement by data providers. No resource modification timestamp available in IIIF, thus it is applicable only to small and medium sized datasets. IIIF Collections are intended for usage by IIIF viewers, thus it's use for harvesting may appear to be a deviation from its purpose |
| Standard Sitemaps | Simple to implement for data providers. (and simpler if resource modification timestamps are not implemented).<br>Reuse of Sitemaps originally created for search engines may be impractical, since there is no information available to distinguish Web pages from IIIF resources |
| Sitemaps extended with elements from the IIIF namespace | Simple to implement for data providers. Better contextualization of the IIIF resources, webpages, and their relations |
| Sitemaps extended with elements from the ResourceSync namespace | Simple to implement for data providers. Better contextualization of the IIIF resources, webpages and their relations<br>ResourceSync elements provide semantics associated with harvesting purposes |
| Crawling IIIF services based on IIIF Collections and HTTP cache headers | Enables a more efficient harvesting process, applicable to large datasets, but the IIIF specifications do not cover the use of HTTP headers. Implementation may not be possible for providers that are unable to modify their IIIF service |
| Crawling by Sitemaps Image and Video extensions | Enables reuse of the image and video Sitemaps made by data providers for Internet search engines. However, the metadata available may not fulfil the minimum requirements for making resources available in Europeana. For our case, an extension is required for linking to EDM metadata |

these kinds of media. Just like search engines, metadata aggregators may also use the media specific metadata for their purposes.

Although we do not have much information about the usage of these types of Sitemaps in CH institutions, some cases are known to exist. One of them is the University College Dublin, which uses both extensions for images and videos.

Since images were being addressed in the context of IIIF metadata aggregation, we focused this case study on the analysis of the video extension.

From our metadata aggregation perspective, the main issue is that the metadata available through these extensions does not fulfil the minimum data requirements for making the resources available into Europeana. The solution adopted with University College Dublin was to further extend the Video Sitemaps with elements from ResourceSync that allow for the association of the EDM metadata, as shown in Fig. 4.

## 5 Future Work and Conclusion

In this paper, we presented the first results of our work on innovating Europeana's technological approach for CH metadata aggregation. Our primary goal is to find a solution that will make the continuous operation of the aggregation network more efficient and lower the technical barriers for data providers to contribute to Europeana. We conducted successful case studies with several technological options using deployed technologies and existing knowledge in CH institutions. A summary of the main conclusions from the case studies, is shown in Table 1.

Now, the challenge is to choose one of the several possibilities and work on establishing a best practice within the community. To achieve this, Europeana is working with the IIIF community in the context of the IIIF Discovery Technical Specification group and will proceed with recommendations targeted at its own partner network.

**Acknowledgments.** We would like to acknowledge the supporting work by Valentine Charles, from the Europeana Foundation, in the analysis of metadata samples collected during the case studies, and the IIIF community in general for the criticism and discussions of our work. This work was partially supported by Portuguese national funds through Fundação para a Ciência e a Tecnologia (FCT) with reference UID/CEC/50021/2013, and by the European Commission under the Connecting Europe Facility, telecommunications sector, grant agreement number CEF-TC-2015-1-01.

## References

1. Lagoze, C., Van de Sompel, H., Nelson, M., Warner, S.: The Open Archives Initiative Protocol for Metadata Harvesting, Version 2.0 (2002), http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm
2. Van de Sompel, H., Nelson, M.L.: Reminiscing About 15 Years of Interoperability Efforts. D-Lib Mag. **21**(11/12) (2015). doi:10.1045/november2015-vandesompel, http://www.dlib.org/dlib/november15/vandesompel/11vandesompel.html
3. Berners-Lee, T.: Linked Data Design Issues. W3C-Internal Document (2006), http://www.w3.org/DesignIssues/LinkedData.html
4. Kenneth, L., Richardson, L., Ruby, S.: Restful Web Services. O'Reilly (2007)
5. Freire, N., Manguinhas, H., Isaac, A., Robson, G., Howard, J.B.: Web technologies: a survey of their applicability to metadata aggregation in cultural heritage. In: 21st International Conference on Electronic Publishing (2017)

6. Snydman, S., Sanderson, R., Cramer, T.: The International Image Interoperability Framework (IIIF): A community & technology approach for web-based images. In: Archiving 2015 (2015), http://purl.stanford.edu/df650pk4327
7. NISO. "ResourceSync Framework Specification", National Information Standards Organization (2014), http://www.niso.org/apps/group_public/download.php/12904/z39-99-2014_resourcesync.pdf
8. Klein, M., Van de Sompel, H.: Extending sitemaps for ResourceSync. In: Proceedings of the 13th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL 2013), pp. 277–280. ACM, New York (2013). Doi:http://dx.doi.org/10.1145/2467696.2467733
9. Fielding, R., Gettys, J., Modul, J., Frystyk, H., Masinter, L., Leach, P., Berners-Lee, T.: Hypertext Transfer Protocol—HTTP/1.1. Network Working Group. The Internet Society. Request for Comments: 2616 (1999) https://www.w3.org/Protocols/rfc2616/rfc2616.html