



The Importance of Interoperability: Lessons from the Digital Public Library of America

Heather Moulaison Sandy & Chris Freeland

To cite this article: Heather Moulaison Sandy & Chris Freeland (2016) The Importance of Interoperability: Lessons from the Digital Public Library of America, *International Information & Library Review*, 48:1, 45-50, DOI: [10.1080/10572317.2016.1146041](https://doi.org/10.1080/10572317.2016.1146041)

To link to this article: <http://dx.doi.org/10.1080/10572317.2016.1146041>



Accepted author version posted online: 04 Feb 2016.
Published online: 04 Feb 2016.



Submit your article to this journal [↗](#)



Article views: 119



View related articles [↗](#)



View Crossmark data [↗](#)

ADVANCES IN LIBRARY DATA AND ACCESS COLUMN

The Importance of Interoperability: Lessons from the Digital Public Library of America

Heather Moulaison Sandy^a and Chris Freeland^b

^aiSchool, University of Missouri, Columbia, Missouri, USA; ^bUniversity Libraries, Washington University in St. Louis, St. Louis, Missouri, USA

ABSTRACT

The Digital Public Library of America enables the discovery of digitized content held by the U.S. cultural heritage institutions by aggregating metadata contributed from participating organizations. Significant interoperability issues exist when gathering metadata at a national scale, which the Digital Public Library of America mitigates through use of its metadata application profile. In its latest release, the metadata application profile permits the storage of uniform resource identifiers, helping the Digital Public Library of America make the move to linked open data. This article examines these Digital Public Library of America developments through a case study of the Missouri Hub, which facilitates the aggregation of metadata from Missouri-based institutions into the Digital Public Library of America.

KEYWORDS

Digital Public Library of America (DPLA); interoperability; metadata; Missouri hub

Introduction

Libraries and other cultural heritage institutions in the United States, as elsewhere in the world, hold unique, digitized content. Since 2013, the Digital Public Library of America (DPLA; <http://dp.la/>) provides a single interface to search for a number of these items. For this single search interface to work, and for the many application programming interfaces (APIs) that have been built to function, there needs to be a certain degree of consistency in the underlying metadata. Getting individual institutions from across the United States to conform to a given set of requirements is no small feat, but once it is accomplished, everyone wins—from the users to the librarians promoting their collections, to the programmers building APIs to provide unique access. Some of the more successful institutions have shown a dramatic increase in usage of their collections since joining DPLA (Jaycox, 2015), meaning that DPLA member institutions are providing the right content, and users are consuming it.

DPLA is not the first trans-institutional resource of its kind. In particular, Europeana (<http://www.europeana.eu>) and the National Library of Australia's Trove (<http://trove.nla.gov.au>) both preceded DPLA, and in doing so, paved the way for the United States to follow. DPLA, Europeana, and Trove are all similar in that they are portals providing access to metadata

about digital content. With DPLA, metadata and some full-text content are stored to promote retrieval through a number of means, but to access the digital object itself, users leave the DPLA and visit the sponsoring institution's Website. Hits, therefore, ping institutional servers since the institutional servers are the ones maintaining the content; this allows individual institutions to know unambiguously how usage is increasing after joining DPLA and which usage is driven through the DPLA interface or through another mechanism associated with it. Because the contributions to DPLA must be open data, use and re-use has the potential to be promoted widely. Although DPLA's work has not been internationally unique, it is, nonetheless, an ambitious and much-needed approach to providing access to the U.S. digital cultural heritage.

Interoperability is essential for any large-scale project where data or metadata is supplied from a variety of sources. Interoperability is defined as “the ability of multiple systems with different hardware and software platforms, data structures, and interfaces to exchange data with minimal loss of content and functionality” (NISO, 2004). Interoperability and the standards that support it must be taken into consideration on a local level as metadata for content is prepared for ingest into the system.

The DPLA's hub model: Managing the metadata

The DPLA aggregates content from two kinds of entities: content hubs and service hubs. Content hubs are large institutions such as the New York Public Library, the Biodiversity Heritage Library, and the HathiTrust Digital Library that submit at least roughly 200,000 items each. Smaller institutions are still able to participate; however, by banding together and forming and subsequently participating in service hubs. Service hubs unite disparate, smaller institutions who have organized, usually geographically, to submit content to DPLA. Examples of service hubs include the Connecticut Digital Archive, the Maine Hub, and Pennsylvania Digital Collections (Hubs, 2015). Although rights of individual resources may be protected, the metadata that is supplied to DPLA must be open and freely reusable. The combination of the two kinds of DPLA hubs results in rich, primary source multimedia collections that include content from over 1,000 partner institutions (Cohen, 2014) resulting in the inclusion of well over 11 million digital items, according to DPLA's Website. For this reason, interoperability is essential to allowing content to be made available through its interface.

One primary feature of the DPLA interface is that it allows for chronological and geographical exploration of aggregated contents. Assigning chronological and geographical metadata, however, is not necessarily within the purview of smaller institutions contributing content through the service hubs. The rest of this column will explore how interoperable metadata is created by service hubs to be loaded into DPLA, and will describe the experience from the point-of-view of one hub, the Missouri Hub (<http://data.mohistory.org/mohub/>), and will consider the potential impact of future DPLA initiatives.

Preparing metadata for ingest in DPLA

When cultural heritage institutions like libraries, archives, and museums decide to participate in DPLA and include their metadata, DPLA carries out a series of enhancements that initially support interoperability and findability (Freeland & Moulaison, 2015). Descriptive text strings in metadata records such as "St. Louis, Missouri" are turned into geographic coordinates following ingest by DPLA (Licht, 2013), providing map-based access on the DPLA map (<http://dp.la/map>)

for users where there was none previously. Users are subsequently able to find content based on its geography through the exploration of a map of the United States. Users are able to refine their explorations by elements such as format, date, language, or contributing institution, making it possible to find one sound recording in the Navajo Native American language from Bluff, Utah that was contributed by the University of Utah—American West Center in association, presumably, with its hub, the Mountain West Digital Library.

Enhancements to date-related information are also made by the DPLA when possible as metadata is loaded, allowing content subsequently to be explored by users on the DPLA Timeline (<http://dp.la/timeline>). Similar to the DPLA map, the timeline allows for exploration by a given date. Unlike the map which is limited to the United States, the timeline extends well before the founding of the United States and includes information about resources and artifacts relevant to the time period or date selected. Outliers that have a way of drawing attention to themselves may in fact be mis-encoded metadata. For example, the year 1002 contains metadata for two items: (1) Annual Holy Convocation, COGIC (94th: 2001), educational institute training manual, and (2) a French text about popular French music from 1992. Although it may not all be perfect, the metadata provides new and interesting ways to explore an institution's content.

In order for content to be ingested in DPLA in the first place, institutions with a variety of digital content and with varying levels of support and available resources must find a way to make their metadata conform to DPLA standards. In short, they must make their metadata interoperable. When institutions are small and the required standards do not align with previously used standards, a bit of a disconnect can be felt and the ability to conform quickly can be completely unrealistic. The service hub solution that DPLA has devised allows many more institutions to have access to DPLA exposure and enhancements if they are willing and able to make joining a priority.

Exposing metadata to DPLA APIs

DPLA's Website promotes a number of APIs that search its entire contents in ways that are similar to some of

the services offered by the major search engines, as well as some that are unique to the primary source documents housed in DPLA. The DPLA app library (<http://dp.la/apps>) is located on the DPLA Website. Because the data is open, developers are free to create APIs using the contents of DPLA. For example, the color browse API (<http://dp.la/apps/33>) by Chad Nelson allows users to explore the DPLA through color, not dissimilar to the way major search engine image searches permit users to explore collections. Another API, the term frequency map (<http://dp.la/apps/25>) by Dean Farrell allows users to see how term frequencies in DPLA's texts vary by state and over time. Such a functionality is not dissimilar to the Google Books Ngram Viewer (<https://books.google.com/ngrams>).

Other APIs support DPLA's work in other ways, beyond that of standard Web search engines. Eric Phetplace has made available an API called the WikipeDPLA (<http://dp.la/apps/22>); it shows results from DPLA as part of a Wikipedia article once the end user's Google Chrome extension is downloaded. When Wikipedia is subsequently viewed in Chrome, a link appears at the top of the Wikipedia page prompting users to search DPLA for related content. A verbal summary of topical results of DPLA items is then given in Wikipedia and a link permits users to hyperlink to a DPLA page with individual search results listed. Similarly, Dean Farrell's Metadata Explorer (<http://dp.la/apps/32>) permits users to view and visualize terms in DPLA, Europeana, or Digital New Zealand. Another API, the DPLA Depositors Visualization (<http://dp.la/apps/27>), also by Dean Farrell, allows DPLA aficionados to see how many items a service hub or a content hub has contributed. None of these products, however, would be possible without open data that is interoperable.

DPLA's metadata application profile (MAP)

To support interoperability, DPLA uses a MAP, an application profile or a set of metadata elements taken from multiple schemas for a particular local use. In this case, the DPLA MAP is used as a lingua franca for DPLA metadata. The DPLA MAP is based on the European data model (EDM) used by Europeana (Metadata Application Profile, 2015). Metadata schemas included in the DPLA MAP are available through their namespaces on the Web. Depending on the institution, the schemas used will be more or less

familiar. According to the latest version of the MAP, the following namespaces are included:

dpla: <http://dp.la/about/map/>
 cnt: <http://www.w3.org/2011/content#>
 dc: <http://purl.org/dc/elements/1.1/>
 dcterms: <http://purl.org/dc/terms/>
 dcmitype: <http://purl.org/dc/dcmitype/>
 edm: <http://www.europeana.eu/schemas/edm/>
 gn: <http://www.geonames.org/ontology#>
 oa: <http://www.w3.org/ns/oa#>
 ore: <http://www.openarchives.org/ore/terms/>
 rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
 rdfs: <http://www.w3.org/2000/01/rdf-schema#>
 skos: <http://www.w3.org/2004/02/skos/core#>
 wgs84: http://www.w3.org/2003/01/geo/wgs84_pos# (Metadata Application Profile, version 4.0, 2015, p. 2).

Of the schema listed here, many librarians will already be familiar with the somewhat limited Dublin Core Metadata Element Set (DCMES) and the much larger number of elements offered as DC terms that are also linked-data friendly. Specialists will likely be familiar with some of the other schemas listed; by using known and recognized schemas, DPLA does not have to find itself in the business of creating a new schema or trying to enforce one that it not currently in use.

DPLA's MAP not only describes content, but also can show the relationships between different elements in an item's metadata record. "This means it is more robust and abstract than a metadata schema like Dublin Core or MODS in that it describes entities and the relationships between them" (Introduction to the DPLA Metadata Model, 2015, p. 1). Describing not only entities but also their relationships is a goal of the conceptual model to describe the bibliographic universe, *Functional Requirements for Bibliographic Standards* (FRBR). FRBR was originally published by International Federation of Library Associations and Institutions (IFLA) in 1998 and is currently under revision to harmonize it with the other two FR family documents: FRAD to address its authority data, and FRSAD for the creation of subject data. Current efforts are underway at IFLA to combine these three into a coherent single model, the FRBR-Library Reference Model (FRBR-LRM). The user tasks identified for FRBR-LRM are (1) find, (2) identify, (3) select, (4) obtain, and (5) explore (Riva & Žumer, 2015). The final user task, explore, is perhaps the most noteworthy because it is the newest and the one most dependent on relationships. Like

the digital library counterparts interested in seamlessly showing relationships, bibliographic data conceptualizers are interested in permitting “exploration,” and to do so, relationships need to be explicit in the data being searched.

DPLA initiatives

Although DPLA does not serve as a content repository, it does function as a steward of metadata contributed by libraries and cultural heritage institutions. The newest version of the DPLA MAP has the potential to permit the demonstration of these kinds of relationships. In particular, version 4.0 of the MAP permits “storing Universal Resource Identifiers (or URIs) from Linked Open Data (LOD) sources” (Metadata Application Profile, version 4.0, 2015). Because LOD requires the use of URIs, this next step will allow DPLA to tap into the existing semantic Web of meaningful linked data relationships. The DPLA has also said that it is looking to enhance records it receives by including URIs for place names. URIs will allow for semantic Web interoperability; their addition goes beyond the geographic coordinates for place names are currently provided as enhancements. DPLA will also enhance metadata to be compatible with the linked data Web by adding linked data vocabularies such as the ones maintained by the Library of Congress.

After that we can begin to consider other kinds of linked open data possibilities like topic analysis or disambiguation of terms, transliteration, enrichment of existing records with more metadata from other sources (a la Wikipedia, for example), and other exciting possibilities. (Gueguen, 2015, para. 4).

Case study: The Missouri hub experience

Armed with the knowledge of current and future benefits to joining with DPLA, a small group of Missouri cultural heritage institutions met in St. Louis in September, 2013, to discuss the possibilities and challenges that submitting their content to DPLA represented. This group was particularly eager to take decisive steps toward broader interoperability and anxious to have made strides by the close of 2014 to celebrate the 250th anniversary of the founding of the City of St. Louis, Missouri, USA, and the 100-year anniversary of the Federal Reserve Bank of St. Louis. Unlike other states in the United States, such as North

Carolina (Gregory & Williams, 2014), or regions such as Mountain West (<http://mwdl.org/>), Missouri did not have a statewide network able to organize the administrative and technical resources necessary for participation; to carry out the required work, the initial group of institutions was on its own (Freeland & Moulaison, 2015).

In December 2013, an administrative group was established, and the technological information necessary was negotiated. REPOX (<http://rebox.sysresearch.org/>), a self-proclaimed “data aggregation and interoperability manager” was selected to (1) harvest contributor metadata records, (2) normalize aggregated metadata, and (3) ultimately contribute to the DPLA. In choosing this software, the group decided to replicate the strategy used by successful hubs such as the North Carolina Digital Heritage Center, and in April 2014, the Missouri History Museum formalized the Data Exchange Agreement with DPLA (Freeland & Moulaison, 2015).

Participating Missouri institutions, initially primarily located in St. Louis, signed a memoranda of understanding and contributed to the development of the necessary technical infrastructures. These infrastructures included those necessary for the normalization of contributed data from multiple institutions. In order to participate effectively, institutions found themselves needing to reassign assisting staff in order to carry out their part, meaning that support had to be at the level of the parent institution and administration.

The first ingest took place in October, 2014, meeting the goals originally established by the founding institutions (Freeland & Moulaison, 2015). Since that time, additional hub members from outside of St. Louis have joined, including the Linda Hall Library in Kansas City. Already, the State Historical Society of Missouri had joined; it is located in Columbia, Missouri. The Missouri Hub has grown administratively since it began. For example, as a cohesive unit, members of the hub have worked together to present information about their experience around the state (e.g., Jaycox, 2015), individuals have presented research based on their work with the hub (e.g., Freeland & Moulaison, 2015; Podleski, 2014), and most interestingly, perhaps, the hub formalized its procedures in October 2015 and voted on bylaws and a governance structure, including the formation of an executive committee.

Table 1. Institutions participating in Missouri Hub and corresponding records contributed to DPLA, as of June 2015 (based on Freeland & Moulaison, 2015, p. 3).

Institution	Date Joined	Items Contributed
Federal Reserve Bank of St. Louis	Sep. 2014	29
Kansas City Public Library	Sep. 2014	1947
Missouri History Museum	Sep. 2014	29796
St. Louis Public Library	Sep. 2014	Not yet harvested as of June 2015
The State Historical Society of Missouri	Sep. 2014	18697
Washington University in St. Louis	Sep. 2014	4536
St. Louis University	Oct. 2014	12052
Linda Hall Library	Dec. 2014	12352
St. Louis Mercantile Library – University of Missouri-St. Louis	Apr. 2015	3041

Working out the technical rough spots

Once the first Missouri Hub metadata was published in DPLA, any inevitable troubleshooting could begin. Issues with data fidelity were identified, including problems such as missing thumbnail images, and mismatches in data feeds from participating institutions were identified. Additionally, problems with the “normalization of aggregated records against the DPLA MAP were both identified and addressed within the Missouri Hub technical infrastructure” (Freeland & Moulaison, 2015). These initiatives were overseen at the Missouri History Museum located in St. Louis, Missouri.

However, technical difficulties were not the only ones that needed to be addressed, administrative concerns surfaced as well. Operational routines needed to be formalized in order for subsequent ingests to go smoothly and for new members to be able to begin participation. With the potential for inclusion of new member institutions, however, the creation of formalized bylaws and governance structures was required. At a meeting in Kansas City, Missouri, in October, 2015, the necessary administrative structure for the maintenance and healthy continuation of the Missouri hub was put into place. As members across the state have joined the Missouri Hub, they have begun promoting their participation to their constituencies and on their Websites.

Benefits to participation

The DPLA API for visualizing contributions confirms that the Missouri Hub has contributed 41,557 items by the end of 2015. A formal hub-wide evaluation has not yet taken place among members of the Missouri hub, but the sense among participants informally is that their collections are receiving additional nationwide

attention and that participation has been a positive experience for them on the whole (e.g., Jaycox, 2015). Overall, participants are pleased to have made the decision to allocate resources in this way and to open their collections to the enhancements and broader partnerships that DPLA permits. Although for many of these institutions, for example, it would be unthinkable to be able to program an API using some of their own content while also linking to other relevant content. DPLA participation, therefore, requires some additional investment, but the benefits outweigh the costs.

Future interoperability and linked data

The question of interoperability continues to loom large among smaller institutions, especially if they have limited resources. Going forward, the benefits of DPLA participation will allow even small institutions to become part of the linked data Web. The members of the Missouri Hub are already looking forward to joining the linked data Web and exposing their contents as LOD in accordance with the DPLA's stated goals. In fact, this was one of the main reasons the group came together in the first place: to find a low-cost way to make their data linked data enabled (Freeland & Moulaison, 2015). All of the Missouri Hub member institutions listed in Table 1, however, already possess the necessary resources and administrative support to take this first step in joining the hub and in promoting their contents on DPLA. Not all institutions have these resources, but that does not mean that their contents should not be on the linked data Web going forward. As an example, in Missouri, a state-wide digital collection called the Missouri Digital Heritage (<http://www.sos.mo.gov/mdh/>) already contains more than 9 million records through the Missouri State Archives, the Missouri State Library,

and other institutions from across the state. Collections are contributed from a number of different kinds of institutions such as archives, universities, museums, and public libraries. Missouri Digital Heritage collections include contents from organizations as diverse as the Old Stagecoach Stop Foundation, the Harry S. Truman Library and Museum, the Museum of Osteopathic MedicineSM and the Hannibal Free Public Library. Although not robust enough to have been able to lead the Missouri hub initially, the Missouri Digital Heritage has announced plans to make its data conform to DPLA standards, thereby potentially allowing the shared, digitized contents of even the smallest Missouri institutions to become enhanced and discoverable through the DPLA's interface (Jaycox, 2015).

Conclusion

DPLA represents a growing body of metadata records highlighting some the crown jewels of United States digital libraries' collections. Although the barrier to entry is steep for small institutions wishing to participate, when partnering with a service hub where an infrastructure is already in place, the barrier to entry is lessened considerably. In providing access via the service hubs, DPLA allows an increasing number of digital objects to be made available, supporting both use and users of U.S.-based libraries and other cultural heritage institutions. For smaller libraries and institutions who otherwise would not be able to participate in such a large-scale initiative, partnering with the DPLA represents a compelling use of time and resources. Providing access to the contents of smaller collections in a way that is egalitarian is just one way that the DPLA is democratizing access to U.S.-based resources. Further support for enhancements with the potential to prepare data to be compatible with future changes to Web architecture such as LOD is a great step, indeed, toward promoting primary source documents that support cultural heritage on the Web.

As libraries in the United States consider the value of exposing their digital collections in this way, the importance not only of local administrative support, but also of local decisions supporting wide-scale interoperability becomes increasingly apparent. The DPLA MAP is one step in promoting interoperability between disparate libraries housing vastly different kinds of resources. As well, service hubs like the Missouri Hub

are instrumental in making sure that smaller U.S.-based institutions with great collections can contribute. DPLA has been growing quickly since its 2013 inception, and its potential at this point seems virtually unlimited.

References

- An Introduction to the DPLA Metadata Model. (2015). DPLA. Retrieved from http://dp.la/info/wp-content/uploads/2015/03/Intro_to_DPLA_metadata_model.pdf
- Cohen, D. (2014). *Inside the DPLA*. OCLC. Retrieved from <http://www.oclc.org/research/events/2014/03-07.html>
- Freeland, C., & Moulaison, H. L. (2015). Development of the Missouri hub: Preparing for linked open data by contributing to the Digital Public Library of America. [Poster] ASIS&T 2015 Annual Meeting, Association of Information Science & Technology, St. Louis, Missouri, November 5–10, 2015.
- Gregory, L., & Williams, S. (2014). On being a hub: Some details behind providing metadata for the Digital Public Library of America. *D-Lib Magazine*, 20(7/8). doi:10.1045/july2014-gregory
- Gueguen, G. (2015, March 5). DPLA MAP, version 4.0. DPLA. <http://dp.la/info/2015/03/05/dpla-map-version-4-0/>
- Hubs. (2015). DPLA. Retrieved from <http://dp.la/info/hubs>
- IFLA Study Group on the Functional Requirements for Bibliographic Records. (1998, 2009 rev.). *Functional requirements for bibliographic records*. IFLA. Retrieved from http://www.ifla.org/files/assets/cataloguing/frbr/frbr_2008.pdf
- Jaycox, E. (2015). DPLA: What is it? How will it benefit my patrons? How can my library participate? MLA/KLA Joint Conference. Kansas City, Missouri, September 30–October 2, 2015.
- Licht, J. (2013). *dpla-geocode-check*. *GitHub repository*. Retrieved from <https://github.com/jlicht/dpla-geocode-check>
- Metadata Application Profile. (2015). DPLA. Retrieved from <http://dp.la/info/developers/map/>
- Metadata Application Profile, version 4.0. (2015). DPLA. Retrieved from <http://dp.la/info/wp-content/uploads/2015/03/MApv4.pdf>
- NISO (National Information Standards Organization). (2004). *Understanding metadata*. Bethesda, MD: NISO Press. Retrieved from <http://www.niso.org/standards/resources/UnderstandingMetadata.pdf>
- Podleski, G. (2014). Choosing metadata for discovery and description in a mixed-content single-subject library: A case study. *Proceedings of the American Society for Information Science and Technology*, 51(1), 1–3. Retrieved from <http://dx.doi.org/10.1002/meet.2014.14505101100>
- Riva, P., & Žumer, M. (2015). Introducing the FRBR library reference model. Paper presented at: IFLA WLIC 2015—Cape Town, South Africa in Session 207—Cataloguing, August 20, 2015. Retrieved from <http://library.ifla.org/1084/1/207-riva-en.pdf>