

DPLA Midwest Workstream Meeting Key Takeaways

October 11, 2012
Chicago, IL

On October 11, 2012, the six DPLA [Workstreams](#), the [Steering Committee](#), and the [Board of Directors](#), along with members of the public, met in a series of breakout sessions and working groups at the Chicago Public Library's Harold Washington Center to discuss the development of both the DPLA platform and the front end user interface, the establishment of the DPLA as an independent non-profit organization, and the launch of the DPLA's [Digital Hubs Pilot Project](#), among other topics. The goal of these meetings was to chart out agenda items for the next six months of work and beyond, as well as to provide a forum for participants and members of the public to share updates and ideas with one another.

This document is a synthesis of the key takeaways from the day's discussions, which focused on three major subjects: (1) The future of the DPLA; (2) its technical development; and (3) its content infrastructure and the Digital Hubs Pilot Project.

I. The Future of the DPLA

The DPLA has taken a number of key steps over the past six months with respect to governance. In August 2012, a [nominating committee](#) was formed to propose a slate of candidates for the inaugural DPLA Board of Directors. In September 2012, a five-person Board was approved by the DPLA Steering Committee, with two additional members to be added by April 2013. Additionally, in early October a new independent 501(c)(3) non-profit organization was incorporated with a basic charter and set of bylaws, upon which the Board, in consultation with the community, will expand in the coming months. The Secretariat and Board have engaged the services of Spencer Stuart, a leading executive search firm, to identify and recruit an Executive Director for the new organization. The law firm WilmerHale continues to provide legal counsel to the DPLA during the process of incorporation.

There are still some open questions facing the DPLA, including those around membership, which it will address in the coming months.

5-year plan/The new organization

Participants spent much of the day discussing these and other developments in light of the DPLA's [five-year plan](#), published as a draft in June 2012 and still under development. The plan attempts to lay out the DPLA's path forward as crisply as possible with current information and assumptions.

The DPLA's institutional location is to be determined and will remain open during the Executive Director search, though it is presently assumed that the DPLA will be centered in a location that will make staff recruiting possible. The DPLA is working to

acquire additional support for its core operations during what participants deemed its “start-up phase,” including funding for the Executive Director and support staff for the first few years of operation. While the DPLA has considerable support for the start-up phase, some participants have focused on identifying an alternative funding model in the near term, which the DPLA will require for supporting operations and key staff once the initial grants begin to expire.

Participants emphasized that the DPLA must continue to expand its community of supporters and participants over the next six months and beyond in order to prove successful; however, the organization must be structured in such a way so as to enable continual forward progress. Participants agreed that the project now stands at the intersection of those two approaches.

Participants discussed how to build relationships with local libraries, as well as ways in which the DPLA could support those public libraries facing pressure from local legislatures to cut costs. The six workstreams, which have worked to address the most principal components of the DPLA planning process, will continue their work between now and April 2013, though it is possible they will transition into new but similar bodies under the leadership of the new Board. One participant put forth the idea of DPLA “emissaries” that would go into local communities to engage them on issues directly related to the DPLA and digital access.

On the question of what would encourage people to retain interest in the DPLA five years from now, participants agreed that a demonstration of value to many distinct communities would be essential. Beginning with the DPLA use cases, both [institutional](#) and [end-user](#), the DPLA has given much thought to how best to connect with a broad and expanding group of individuals. The DPLA’s commitment to open data and its ease of use also will aid in retaining interest. Other items brought up in conversation highlighted efforts that have already been practiced with consistency throughout the planning initiative, including a commitment to an open source metadata platform; an open technical process and system of governance; and a desire to reach beyond the library community to teachers, publishers, activists, technologists, and others. Participants agreed that interest in the DPLA would remain strong as long as individuals could find personal meaning in its resources and services.

Membership, Partnership

The DPLA has not yet reached consensus on whether or not a membership model is desirable. One participant asked whether the naming of the initial DPLA Service Hubs signals a *de facto* move toward a particular type of membership model. While the answer to this question is not yet known, some participants felt that multiple types of potential partnerships might make sense for the DPLA.

One possibility is to have participating hubs organizations serve as DPLA members, an option that the Financial/Business models workstream has previously discussed. One participant pointed out that if the DPLA is membership governed to a certain extent,

this might have an effect on leadership and innovation in the DPLA network. Depending on the model, participants stressed the fact that membership can be inclusive, divisive, or both, and that a bottom-up affinity-based approach would perhaps be the most practical model, though a range of nuances complicate the understanding of “community” in such an arrangement. Participants agreed that an overall public-private partnership would be critical to help insulate the DPLA from any changes in the political climate.

II. Technical Development

The development of the DPLA’s technical infrastructure is being conducted in stages. The first stage involved the initial development of a back-end metadata platform—an alpha version of the API that was presented at [DPLA West](#) in April 2012. The [next stage](#) in the DPLA’s technical development involves integrating continued development of the back-end platform, complete with open APIs, with new work on a [prototype front-end](#). For April 2013, the “content” within the platform is metadata. Developers will be able to access this content via a search API, and users can output data in JSON, XML, and, eventually, RDF. An initial version of a qualified [Dublin Core schema](#) has been drafted, and the technical development team is currently building an ingestion engine that harvests, enriches, and normalizes metadata from those institutions providing data—the initial Service Hubs in the [Digital Hubs Pilot Project](#). The Service Hubs, whose data provider agreements spell out some minimum standards for metadata quality, will manage the work involved in preparing and maintaining harvestable metadata, as the DPLA does not have the organizational resources to review all of the metadata that it aggregates.

The API is a REST API written in Ruby. The metadata back-end is a CouchDB database, and the search index is powered by ElasticSearch. There is currently no federated searching; all searching is done against the DPLA database. The ingestion engine is built in Python, with code from the Library of Congress ingestion process.

In October 2012, the [DPLA partnered with the Boston-based web design firm iFactory](#) to design and develop a front-end website. This website will serve as one possible front-end interface to access the data contained within the platform. The DPLA is being designed in such a way that will allow developers to build other front-ends using the API.

Metadata discussion

Participants spent a portion of the day conducting an open discussion on the DPLA’s metadata approach, as well as ideas for future consideration. The DPLA is presently harvesting metadata from five of the seven initial Service Hubs using OAI-PMH. Small libraries, museums and archives of all types can contribute their metadata to the DPLA via regional or state Service Hubs; participants discussed the possibility of creating

some kind of tool that would help small organizations organize their data for an aggregator.

The DPLA will have a body of “dark metadata,” or the rich, unmapped original metadata from service hubs, in addition to the metadata mapped to the DPLA schema. This richer metadata—MARC21 records, for instance, have a number of fields not required in the Dublin Core DPLA schema—will also be held, apart from the mapped metadata, and will ensure that the more complex original data is not lost in the process of harvesting and ingestion. Participants recommended that this set of dark metadata be searchable via the API.

In terms of data standardization, as much metadata as possible will be modified at the local/creation level. Participants pointed out the tendency for metadata to get “dumbed down” the higher up it moves in an aggregation chain, and participants advocated for standardization at the point of creation, or at the local level, so as to ensure minimal loss of information in the mapping process. Participants were unsure how to compensate for issues of bad metadata quality, though some suggested that institutions ought to be held accountable to a certain degree for low-quality metadata, such as exclusion from the master DPLA data repository. The DPLA technical development team is currently conducting an analysis of a subset of the metadata aggregated thus far. The team is operating under the assumption that the Service Hubs are accomplished in the field of creating good metadata.

Front-end/Apps discussion

Participants also discussed the prototype DPLA front-end currently under development by iFactory. They brainstormed possible apps that could be built on top of the DPLA platform. Participants thought about how to include, feature, and engage local content in the DPLA prototype front-end and other interfaces.

In advance of the November 8-9, 2012 [Appfest](#) hackathon at the Chattanooga Public Library in Chattanooga, TN, participants discussed potential apps for the DPLA platform. These apps would ostensibly be built by hackers at DPLA hackathons, tech-savvy researchers, post-graduate students, library/information vendors, library developers, corporations, university classes, and more. The variety of special collections available through the DPLA API will serve as the primary “carrot” to encourage the development of apps, participants suggested.

Some of the apps brainstormed included: Omeka plug-ins; embeddable widgets that seem like they’re part of another site and are grabbing data from the DPLA; exhibition apps that allow users to create their own curated exhibitions; content creation apps; apps for teachers to pre-select content for students; Digg-like functionality for rating and ranking user-created content; apps that provide attribution when right-clicking/“saving as”; games, points, or badges that incentivize contributions and interactions with the DPLA; and visualization tools from the DPLA.

Participants also discussed other functions, services, or tools that the DPLA might eventually provide users. Libraries are historically good at buying things that are too expensive for their users (journal subscriptions, etc.), one participant pointed out. Could the DPLA lend expensive software, such as Final Cut Pro or Adobe Photoshop, to its users? Others asked how the DPLA could self-generate feedback about its services, such as popularity charts, click stream data, and/or analyses of items or services that might be missing.

III. Content Infrastructure and the Digital Hubs Pilot Project

Under the recently launched [Digital Hubs Pilot Project](#),¹ the DPLA will undertake the first effort to establish a national network out of the over forty state/regional digital libraries, numerous digital projects and large content repositories, and other promising initiatives across the United States, bringing together myriad digitized content from across the country into a single access point for end users. The approach is to work with seven states or regions (Service Hubs) and an equal number of content providers (Content Hubs) to aggregate content on a pilot basis. The goal is to demonstrate how the DPLA, on behalf of the American people, can make vastly more of a whole from the sum of the parts of the United States' digitization efforts.

The Service Hubs initially identified for the pilot are:

- [Mountain West Digital Library](#) (Utah, Nevada and Arizona)
- [Digital Commonwealth](#) (Massachusetts)
- [Digital Library of Georgia](#)
- [Kentucky Digital Library](#)
- [Minnesota Digital Library](#)
- [South Carolina Digital Library](#)
- [Oregon Digital Library](#)

The DPLA's vision is to strengthen and connect existing state and/or regional infrastructure to create a system of Service Hubs from which aggregated data from libraries, museums and archives is harvested. Each Service Hub will offer a full menu of standardized digital services to local institutions, including digitization, metadata, data aggregation and storage services, as well as locally hosted community outreach programs bringing users in contact with digital content of local relevance. Service Hubs will aggregate metadata, digitize new content, and work with partners throughout their state or region to identify new content for digitization and description. Service Hubs will also provide storage for this digitized content. Community programming is also a critical component of the pilot. The two-year Hubs Pilot aims to help existing state programs offer these services to all institutions in their state or region. Service Hubs will serve as

¹ For a detailed summary of the Digital Hubs Pilot Project and the seven initial service hubs, see: Notes from the September 20, 2012 Digital Hubs Pilot Project Kick-off Workshop, http://blogs.law.harvard.edu/dplalpha/files/2012/10/DPLA_DigitalHubsProject_KickoffMeetingNotes_10-18-12_FINAL1.pdf.

an on-ramp for every institution in a pilot state or region to participate in the DPLA network. The DPLA hopes to replicate this model in other states and regions if it proves successful.

Content Hubs are large content holders— such as the National Archives and Record Administration, the Smithsonian, and Harvard University—that will share a 1:1 relationship with the DPLA, offering their data up for direct harvest by DPLA aggregation systems.²

At this stage, the Hubs Pilot is harvesting metadata and previews of “green-lighted” content, or content that has no rights restrictions placed on its use, though advocacy roles for “amber” and “red light” content are a possibility in the future. Based on conversations about the availability of green-lighted content that resolves to digital objects, the Hubs Pilot will start with existing digitized content that has been supported by federal, state, and local funds. In joining together these pools of cultural heritage content, the Hubs Pilot Project aims to encourage conversations amongst community members across the country, both locally and in a national context. Toward that end, the seven initial Service Hubs have agreed to create virtual exhibitions that employ and highlight content in their collections.

The subjects of these exhibitions are:

- Activism and Civil Rights
- Prohibition
- The Great Depression and the Civilian Conservation Corps
- Native Americans
- “Our Town USA”/“Small Town Life”
- Natural Disasters
- National Parks

In seeking to clarify terms, participants reiterated the fact that content, within the context of the DPLA, means metadata. Because of this, the term “datastore” was put forth as a clearer description than “collection” in describing the DPLA’s aggregated data, as collection suggests repositories of content rather than metadata, though the project is still working to define terms during the start-up phase. All metadata incorporated into the DPLA will use a CCo license.

² Since the October 11, 2012 workstream meetings, the Harvard Library has officially announced that it will serve as the first DPLA Content Hub. See: “Harvard to Contribute Special Collections Materials to Digital Public Library of America,” October 31, 2012, <http://library.harvard.edu/harvard-contribute-special-collections-materials-digital-public-library-america>.

DCC Research

Katrina Fenlon, a researcher for the Digital Collections and Content Registry (DCC), presented findings from an IMLS-funded national study³ that surveyed reference service providers to determine what they want from a cultural heritage aggregation service such as the DPLA. Of the 1,200 institution surveyed, 416 responded. Fenlon and her colleagues found that local history was the topic of greatest interest for reference service providers. Participants commented that this finding aligned nicely with the DPLA's initial focus on cultural heritage material.

Exhibition themes, topics, and strategies

Participants broke off into groups to brainstorm potential exhibition themes for future DPLA projects, based on institutional type and content type. These included:

- Themes that have general, universal appeal, such as genealogical materials
- Sheet music and other musical projects
- Strategies that tie into existing local theme development or existing successful community engagement projects
- Teacher guides, resource guides
- Menus, tickets, cookbooks, and other local materials related to culinary arts
- Themes that would speak to a younger demographic (e.g., music and video items, sports, beauty products and cosmetics, election coverage, kid's books, Native American historical materials, etc.)
- Government documents (participants noted that this could be a cross-theme focus)
- Oral histories
- Content from contemporary/current events (e.g., material from the Occupy Wall Street movement)

Engaging a broad community

Participants discussed ways to engage institutions that are interested in contributing to the DPLA. Participants suggested that the DPLA could use collection development as both a recruitment strategy and advocacy platform, promoting a variety of local content to a national context for the former, and spotlighting collections that present copyright issues in the latter. Advocacy informs collection development, and collection development informs advocacy, participants said. Some participants suggested a toolkit that would facilitate the process by which smaller institutions could submit metadata directly to the DPLA, though Hubs Pilot project leaders advocated for a workflow that would see smaller institutions contributing data to a larger aggregator in their state or region who would in turn offer up that data to the DPLA for harvest.

³ See: Katrina Fenlon, "National Evaluations of IMLS DCC & Implications for DPLA," https://www.ideals.illinois.edu/bitstream/handle/2142/34752/surveySlides_DPLA_current.pdf?sequence=2.

IV. Next steps

Technical Development

The [DPLA Appfest](#) is an informal, open call for both ideas and functional examples of creative and engaging ways to use the content and metadata in the DPLA back-end platform. Participants will build apps on the DPLA API, which will include metadata from the Minnesota Digital Library, the South Carolina Digital Library, and the Digital Library of Georgia (the technical development team is currently ingesting data from the remaining Hubs). Between now and December 2012, the technical development team will continue work on the platform, ingest additional metadata, and, in conjunction with the DPLA Secretariat, approve visual design and user experience elements for the prototype front end. Between December and April 2013, work will involve building out the prototype and integrating it with the platform.

The Future of the DPLA

Between now and April 2013, the DPLA will identify an Executive Director to lead DPLA, Inc. The Board of Directors will work with the Secretariat and the DPLA community to refine the new organization's charter and bylaws; the Board will expand to include two additional members, for a total of seven. Some open questions remain for the DPLA, including those around membership, which it will address, in consultation with the community, in the coming months.

Content Infrastructure and the Digital Hubs Pilot Project

The DPLA will develop data provider agreements with the seven initial Service Hubs and will work to identify future hubs. Between now and April 2013, the DPLA aims to formalize agreements with Content Hubs in addition to its recent partnership with Harvard, as well as to roll out initial service offerings at the DPLA's Service Hub pilot sites.