

## DPLA West Workstream Meetings

April 26, 2012

On April 26, 2012, the Berkman Center for Internet & Society, in conjunction with the San Francisco Public Library and the Internet Archive, convened a series of meetings of the six DPLA workstreams at the San Francisco Public Library to discuss key issues pertaining to the DPLA and its core objectives. The goal of these meetings was to chart out agenda items for the next six months of work, as well as to provide a forum for workstreams to share updates and ideas with one another.

While individual workstreams discussed topics specific to their objectives, a number of themes bubbled up across the different workstreams, including the roles and responsibilities of nodes/aggregators, workstream communication protocols, digitization projects, and incentives for participation. This document is a synthesis of the key takeaways from the day's discussions.

### Nodes

The idea of a distributed network of indexed metadata, similar in concept to Europeana, is the predominant model for the DPLA. In this model, a designated institution or set of institutions (a node) would aggregate metadata from multiple institutions and process it for harvesting by the DPLA. Nodes would in effect serve as a sort of intermediary or broker between local institutions and the DPLA. The overall collection of nodes would constitute a greater DPLA network. By virtue of this intermediary role, nodes would potentially need to assume a number of key responsibilities in order to ensure standards and best-practices across the network. Many DPLA workstream participants agree that leveraging already harvested metadata and content at the node level (e.g., Internet Archive, HathiTrust, California Digital Library, statewide digital collaboratives) would be the most efficient and effective content acquisition strategy.

The idea of a "node" was tentatively defined as a "content or metadata aggregator" based around a specific subject or geographic area. While a general definition of node was agreed upon, specifics for inclusion such as requisite size or level of capabilities (i.e., state-level aggregators vs. individual contributions from organizations) were not. Participants suggested that the Steering Committee craft and sign a note that formally enumerates a node-based structure for the DPLA. Participants agreed that finding 3–10 institutions to act as nodes in the DPLA prototype was a key priority moving forward. Thirty-two states already have statewide digital library collaborations and were suggested by several participants as a good starting place.

The process by which potential nodes are identified and recruited is currently underway: the DPLA Secretariat, working alongside Emily Gore, is conducting a survey of state libraries, collaborative digitization organizations, and aggregation models. The details of a

memorandum of understanding (MOU) formalizing the roles, responsibilities, and terms of relationship between nodes/data aggregators and the DPLA were discussed. Participants were invited to suggest edits to a [working MOU document online](#).

### **Communication & Decision Making**

Participants discussed a number of ways in which workstreams could better exchange information with one another. Group suggestions included a regularly scheduled conference call or similar method of communication between workstream co-chairs, as well as a regular conference call between the Content and Technical workstreams. Participants noted that a master calendar containing key deadlines and timeframes would help in synchronizing cross-workstream communication.

The group expressed desire for a focused Steering Committee planning session to prioritize and articulate a number of key decisions, such as a structure for the network of nodes/aggregators.

### **Audience & Participation**

Discussions concerning “paths for involvement” and incentives for participation popped up in a number of workstream meetings. Participants agreed that a set of clearly identified and marketable incentives for participation should be articulated. The benefits of participating in the DPLA may not be readily apparent to certain groups, and organizations that are presently willing to contribute or use the DPLA do not have a clear path for involvement. How does the DPLA, for instance, encourage small libraries lacking in significant infrastructure to participate?

To help articulate incentives, several participants suggested breaking potential DPLA partners down into tiers based on type:

- Tiers for content/metadata providers;
- Tiers of access (i.e., read-only, download, etc.);
- Tiers for libraries that want to use DPLA code;
- Tiers for individual users (i.e., scholar, casual reader, etc.);
- Tiers for organizations.

Participants pointed out that the clear definition of these “paths” and incentives will need to involve the Content & Scope, Audience & Participation, Technical, and Governance workstreams at the very least.

Meanwhile, preliminary work on defining types of individual participation has been taken up by Nate Hill, Audience & Participation Co-Chair, who has begun fleshing out a number of

[use cases for the DPLA prototype](#). Participants agreed that mapping out ostensible user types would help in providing structure to the issue of audience and participation. Future work includes figuring out a way to best capitalize on these use cases, as well as getting workstreams and other interested parties to contribute.

## **Content & Scope**

The Content & Scope workstream largely discussed digitization models and projects, including the Internet Archive's [Million Books Project](#) and [Open Library](#). Government documents seem a likely target for inclusion in the DPLA. Participants noted that if the DPLA were to digitize new materials, metadata guidelines for describing them would need to be written, or perhaps modeled closely off another institution's metadata guidelines. Participants were undecided on how to secure original ebook content for the DPLA.

## **Finance & Governance**

In preparation for the eventual formation of an independent 501(c)(3) for the DPLA, the Governance workstream in conjunction with the Steering Committee has begun the process of setting up a search committee for a future Executive Director and a nominating committee for a future Board. Questions arose about the size and shape of the 501(c)(3), although a survey of different governance models scheduled to be completed by the DPLA Secretariat in June 2012 should offer guidance by way of comparison.

Participants suggested a number of deliverables, including a tighter and more clearly articulated mission and vision statement, a 5-year plan for the DPLA as an independent organization, and a statement on the DPLA's funding priorities.

## **Legal**

The Berkeley Center for Law and Technology is currently working on a number of publications covering the issues of orphan works, copyright, and legal issues pertaining to mass digitization. The Secretariat is working with the Cyberlaw Clinic and the Berkman Center on two additional memos, one on accessibility issues and one on privacy.

## **Technical**

Participants discussed at length the [DPLA prototype](#). While the prototype exists as a metadata platform at present, the interim technical development team said that the DPLA would develop a "shiny" front-end user interface as one option for users, with bulk

download being another. The possibility of another [Beta Sprint](#) for apps was suggested, perhaps sometime between October 2012 and April 2013.

Participants agreed that a cleaner, more human-readable scope and workplan for technical development was a priority moving forward. Unresolved questions revolved around how the DPLA would ingest new content and metadata, and what models for metadata/data aggregation already exist that the DPLA might be able to use.

## Next Steps

### *Survey of state libraries and collaborative digitization organizations*

Participants agreed that finding 3–10 institutions to serve as nodes/aggregators between now and April 2013 is a key priority. The DPLA Secretariat, working in tandem with Emily Gore, launched two surveys in early May, one of state libraries and one of collaborative digitization organizations, to help map the landscape of digitized content both in terms of what content is available and what kinds of funding exist at various levels for digitizing content.

Additionally, a draft memorandum of understanding (MOU) for data providers/aggregators is currently open for comment at <https://docs.google.com/document/d/1gEHZA8WQeLiXwpvMrQ4LnO2KRNTnxNeJPdeerR7jKfc/edit#heading=h.wrxx9o4gjf5e>.

### *Communication*

A number of suggestions were proposed to better facilitate cross-workstream and Steering Committee communication. These suggestions included regularly scheduled calls between workstream co-chairs and between the Content and Technical workstream co-chairs. The Content & Technical workstreams held an initial call in May and are scheduling another for late June/July.

### *Audience & Participation*

Participants emphasized the need for a clear articulation of incentives for participation, perhaps based on tiers or types, as well as different “paths for involvement.” Toward that end, Nate Hill, Audience and Participation Co-Chair, has undertaken a series of use cases that will seek to understand the needs of future users, thereby demonstrating solutions and limitations of hypothetical DPLA applications. These DPLA use cases can be found online at <http://www.natehill.net/dplausecases/index.html>.

### *Governance*

The work of the Governance workstream and the Steering Committee in the coming months will focus primarily on laying the groundwork to establish the DPLA as an independent organization. Such a process includes setting up a search committee for a

future Executive Director and a nominating committee for a future Board. Other priorities include articulating a tighter and more clearly articulated mission and vision statement.

*Five-year plan*

Workstream participants called for a five-year plan for the DPLA. The DPLA Secretariat is currently working on a draft plan.

*Technical Development*

Work is to continue on the DPLA prototype into the summer. Another [Beta Sprint](#) for apps is a distinct possibility, occurring tentatively sometime between October 2012 and April 2013.