DPLA | DIGITAL PUBLIC LIBRARY OF AMERICA

Technical Working Meeting
December 9, 2011

Introduction

On December 9, 2011, the Digital Public Library of America Secretariat convened a small working meeting with a representative of each Beta Sprint project presented at the October 2011 plenary meeting, members of the DPLA Technical Aspects workstream, several key industry players, and a small group of technical experts. The goal of the meeting was to bring together a group of talented colleagues in libraries and non-profits with the best technical minds working on commercial products in a collective brainstorming session to determine a rough outline for a generative DPLA platform.

Key considerations stated at the outset of the meeting include the desire to release a platform upon which developers can innovate, well in advance of the April 2013 deadline. This platform should be not only a place to share existing content, but also to support new content and applications. DPLA Steering Committee Chair John Palfrey emphasized the need for the DPLA to remain a highly distributed effort, with key partners across the country, while at the same time accomplishing a rapid development process between now and April 2013.

Principles for Technical Development

DPLA Technical Aspects Workstream Co-Chair Martin Kalfatovic re-introduced the DPLA Principles for Technical Development, drafted at the June 2011 technical workshop in Washington, DC. In brief, these principles cover four main areas:

- *Metadata:* All metadata in the DPLA should be in the public domain; the DPLA should contemplate content from outside of libraries and museums.
- *Code*: Code created and used by the DPLA should be open, sharable, and reusable.
- *Content*: Content will come from partners and should be made available with no new restrictions.
- *Participation:* The DPLA will not put in place any new gatekeepers or restrictions, but rather will encourage and actively support the community of developers and others who want to reuse and extend its code, content, data, and metadata.

Content and Metadata

Participants briefly discussed content to be included in the DPLA, noting that much of the initial content will likely come from partners including libraries, museums, and archives. Participants requested an environmental scan of state and local heritage efforts and suggested that such a scan be conducted in part by working with the Content and Scope Workstream and crowdsourcing information via the DPLA wiki.

Participants also discussed whether the DPLA will focus primarily on metadata or if it will also offer content hosting or storage and preservation services. Participants noted that hosting content could be helpful in some cases, but that relying largely on lightweight processes to ingest metadata, rather than hosting a mass of content within the DPLA, will enable more widespread participation from libraries and other partners.

With respect to metadata specifically, participants noted that the DPLA should avoid taking a "lowest common denominator" approach to metadata at the risk of effectively downgrading high-quality metadata. At the same time, the DPLA should not exclude partners whose metadata does not meet rigorous, rich standards. Some participants argued that the DPLA should not suggest any standards that limit participation (for example, only accepting certain formats); others felt that the DPLA should simply support a set number of metadata formats and then allow contributors and implementers to use the tools they feel best fit their needs.

As a corollary to this discussion, participants suggested that the DPLA or various DPLA partners might hold events to engage the community of people who enjoy editing and cleaning metadata as a way of enabling the broadest possible participation from various metadata contributors.

Participants also pointed out that in an environment that encourages user-generated content, metadata and data can be confused easily. The DPLA Principles for Technical Development currently state that all metadata will be available for bulk download and reuse; some participants questioned whether this would include data like user reviews.

Platform and Initial Interfaces

David Weinberger (Harvard Library Innovation Lab) opened this session with several questions/comments:

- *Who are the users and beneficiaries of the DPLA platform?* The platform will be open to all, but we cannot plan for every eventuality, and as such, the technical development team needs some sense of who the important users will be in order to set priorities.
- *How can those who build upon the platform contribute their work back to the DPLA? How can the DPLA absorb and make use of this work?*
- *In addition to APIs, is there a set of tools the DPLA should provide to developers in order to enable them to do their work more quickly (even in the next few months)?*

- *Should the DPLA develop an "app store"—some location or capability for both developers and users to find code and tools?*
- *What shared services (user profiles/logins, social graphs, etc.) should the platform provide?*
- *Upon which standards should the DPLA be built?* The DPLA is in a position where it potentially may drive standards; the technical development team should consider this carefully.

In a brainstorming session, individual participants identified the DPLA platform as including or being:

- Modular
- Open, versioned APIs
- Intertwined with and/or driving standards
- Fundamental level services
- User of existing web standards
- A specifier of formats and/or schemas
- Discovery services/name resolution services (in order to make it more discoverable in the open Web)
- A catalog (to help with authority problems: control, classification, versioning)
- A crawler service
- Links that are open data friendly
- Thesaurus classification, rather than keyword search only
- Connected to existing projects such as Freebase and the Open Library Project
- A URI that takes API requests
- A value add to the Web
- A service that helps with or supports vocabulary control
- A reference implementation
- A set of open code/metadata/services/schemas /standards/protocols
- A scaffold

- Support for "social infrastructure" (inside or outside the core?)
- Facing developers, and ultimately end users
- Including an "incubator"
- Unique IDs, including a collections ID system
- Data processing/clustering services for items, perhaps through a combination of human intelligence and computer algorithms
- Prepared to handle hierarchies of materials in order to enable inclusiveness
- License manager (if required)
- Metadata, including search and usage data
- A way to add value to the Web by defining media formats and standards for presenting information
- A collection of well-running, well-respected, well-fed bots and scripts in order to reduce the burden on developers and services that interact with the platform
- Training and support for humans who use the platform

Development Process and Next Steps

Participants expressed the need for fairly rapid prototyping, given the April 2013 deadline, as well as a mechanism for fast code review and response for those who would like to contribute. Multiple Beta Sprinters offered their code, metadata, and/or APIs to the group and have begun posting these to the Community Portal on the DPLA wiki (http://cyber.law.harvard.edu/dpla/Community_Portal).

The development process will mostly likely involve a small team of developers, with the Technical Aspects Workstream serving an advisory role and acting as a channel for public input.

Participants agreed that key next steps include sharing existing code, metadata, and APIs; developing a set of the minimum requirements for a viable DPLA platform in order to enable rapid, focused development; and defining a clear data model (based on existing models) for the DPLA.