

# **The Digital Public Library of America Initiative: Considering Content and Scope**

Carole Palmer & Katrina Fenlon

E-Research Roundtable

GSLIS

30 March 2011

# Steering Committee

**Paul Courant**, Dean of Libraries at the University of Michigan

**Robert Darnton**, Director of the Harvard University Library

**Charles Henry**, President, Council on Library and Information Resources (CLIR)

**Susan Hildreth**, Director of the Institute for Museum and Library Services

**Brewster Kahle**, Founder of the Internet Archive

**Michael A. Keller**, Director of Academic Information Resources, Stanford

**Carl Malamud**, President, Public.Resource.Org

**Deanna Marcum**, Associate Librarian for Library Services, Library of Congress

**Maura Marx**, Berkman Fellow and Executive Director, Open Knowledge Commons

**Jerome McGann**, Professor at the University of Virginia

**John Palfrey**, Faculty Co-Director, Berkman Center; Vice Dean of Library and Information Resources at Harvard Law School

**Peggy Rudd**, Executive Director/State Librarian, Texas State Library and Archives

**Donald Waters**, Program Officer, Andrew W. Mellon Foundation

**Doron Weber**, Vice President, Programs at the Alfred P. Sloan Foundation

# “Playing Catch-Up in a Digital Library Race”

- New York Times, January 8, 2011:

Opens with America’s public library movement in 1731 with Benjamin Franklin’s Library Company of Philadelphia.

*Lending libraries may have been the newfangled democratizing factor of their day. Centuries later, though, the United States finds itself trailing Europe and Japan in creating the modern equivalent: a national digital Library that would serve as an electronic repository for the nation’s cultural heritage.*

# Prompted, in part, by Google Books

Google came along and woke everyone up and showed the world what could be done in a short period of time.

– Maura Marx, Berkman fellow

There's a conflict between Google's aims to make money for its shareholders and the goal of libraries to make books available to readers.

But DPLA has better chance of success if includes out-of-copyright books owned by member libraries that Google has digitized.

– Robert Darton, Director Harvard University Library

# Working Group Meeting, 1 March 2011

## **A Digital Public Library of America—for whom?**

- Public library users - Josie Parker, Ann Arbor Public Library
- Academic users - Dan Cohen, George Mason University

## **Public domain collections and open business models**

- Digitized public domain books, Chris Freeland, Biodiversity Heritage Library
- Cultural heritage collections, the Europeana experience, Louise Edwards
- Digitized newspapers, Bob Horton, Minnesota Historical Society

## Content with complex barriers to access

- James Shulman, ARTstor
- Science as a public resource, Peter Jerram, Public Library of Science
- Access to scholarly books, Kevin Guthrie, Books at JSTOR
- In-copyright books, Steve Potash, Overdrive
- Google Books and access to orphan works, Dan Clancy, Google
- Working with educators and end users, Jason Ediger, Apple

# Highlights from March Concept Note

## Users:

... for students of all ages, from grades K-12 to postdoctoral researchers and anyone seeking self-instruction; it will be a deep resource for community colleges, vocational schools, colleges, universities, and adult education programs;

it will supplement the services of public libraries in every corner of the country; and it will satisfy other needs as well—the need for data related to employment, for practical information of all kinds, and for enrichment in the use of leisure.

... the DPLA cannot be everything to everyone. For it to fulfill its mission, its scope must be carefully defined, and it must be erected incrementally, according to a realistic plan.

# Governance

... cannot hold together without adherence to common practices, and those coordinated modes of behavior cannot be sustained unless there is an adequate administration to govern the whole system.

One could imagine grafting the system onto a structure that already exists, like the Library of Congress, CLIR, or ALA. But it will be necessary also to explore the advisability of creating a new entity.

... design of any governing body should be the work of a commission composed of representatives from the worlds of libraries (public as well as private), information technology, publishing, and the general public. It probably should include deputies from the research libraries whose holdings will be integrated into the system.

In order to be protected from political pressures, it should be a free and autonomous body, perhaps something like the BBC.



# Access

... must be open to all Americans, free of charge.

openness should extend to everyone on the globe where possible, subject to legal constraints that may arise.

should be coordinated with those of digital libraries in other countries.

use should be unrestricted, unless exemptions from copyright requirements may exclude commercial applications.

# Areas where LIS, and DCC in particular, may contribute

- Content
- Metadata
- Architecture

# Content

... it should concentrate at first on the written record—books, pamphlets, periodicals, manuscripts, and digital texts.

It will not neglect audio-visual materials, but it will coordinate its growth with LoC, NARA, the Smithsonian Institution, and other national repositories.

should begin with works in the public domain that have already been digitized and are accessible through the Internet Archive, HathiTrust, a broad range of government material, and possibly private-sector initiatives such as the Google Books Project.

These can be supplemented by digital collections of research libraries and amalgamated holdings such as the digitized newspapers from the fifty states

A new program of scanning collections should then be undertaken with the goal of including all printed material up to 1923 from all of the major research libraries.

# Architecture & Metadata:

... build an approach that draws its lessons from the architectures and large-scale systems of the Web, and particularly from existing and ambitious digital library initiatives, such as Europeana. Such a system would most likely take the form of an open, distributed network of online resources enriched by semantic data and thus made discoverable on the Internet.

Legacy digitized collections require enriched semantic metadata, and current scanning operations need updated tools to create such metadata

Does DCC really fit in the DPLA vision?

Is there a strong argument for inclusion of special collections?

What is the particular value of public library and museum collections, not yet articulated by DPLA?

How valuable is our experience with national aggregation?

What are the most important DCC lessons for DPLA?  
Conceptual, metadata, infrastructure?

Overall what LIS expertise is needed in this kind of initiative?

# Controversial topics

- Should the DPLA initiative drop “public” from its name?
- How can public libraries be actively involved?
- How best to work with Google, if at all?