On March 1, 2011, the Berkman Center convened a group of participants from public and research libraries, government agencies, publishers, and private industry for a day-long workshop focused on the content and scope of a proposed Digital Public Library of America (DPLA). Participants were invited to explore intrinsic features of specific types of content and their interrelations and to consider questions of how to deal with vendors and materials under various types of restrictions. The goal of this initial meeting was to make an important contribution to the overarching goal of the initiative: to work towards a shared vision of a DPLA and a set of prioritized next steps. This document highlights a selection of central discussion points and questions; we hope that these takeaways will serve as input into future discussions about a potential DPLA.

I. Introduction

Steering Committee members John Palfrey and Doron Weber outlined a key set of questions for the day, including what content a potential DPLA might include and how it might be accessed; whether the DPLA should include the word “public” in its name; and how best to encourage broad involvement in this initiative, including efforts to learn from past and current projects. A video of these opening remarks is available on YouTube: http://www.youtube.com/watch?v=MtANodeeEl0.

II. A Digital Public Library of America—for whom?

In the workshop’s opening session participants began to define a spectrum of future DPLA users by examining the needs of public library users and research library users—a dichotomy that most participants ultimately rejected in favor of a framework focused more on a range of use cases, activities, and behaviors than on discrete user profiles or identities.

Some participants expressed concern over the use of the word “public” in the DPLA’s name, noting that the term “public library” has specific connotations involving the provision of services often targeted at non-academic users. Participants noted that if a proposed DPLA focuses exclusively on public domain content, it may not meet the needs of most public library users.

Participants familiar with the needs of research library users emphasized the need for reliable metadata, flexibility in search methods in order to accommodate different research needs, and an ongoing connection with the physical media despite the digital nature of the project. They also cautioned against a loss of the serendipity that currently exists when users browse through library stacks and suggested implementing a recommendation engine or a social tagging mechanism to help replicate this kind of discovery.

Several participants stressed the importance of APIs, open data, and open metadata to allow scholars and others to build new services on top of whatever form a future DPLA might take. This
point of view tied in with the “use case” framework mentioned above; these participants expressed a desire to identify possible uses and activities and develop a system that addresses these cases while leaving room for a host of possible future uses.

The discussion of user needs highlighted the fact that libraries are no longer able to provide access to all materials they previously could. Participants pointed out that the digital age has, in some ways, prevented broad access to popular materials. They raised the question of whether a DPLA should take on this problem, or instead focus on providing access to materials that are relatively easy to obtain. Related to this question is the question of whether a potential DPLA currently has the technical and legal tools it needs to operate. Some participants believed that current tools are sufficient and that a DPLA should begin with available materials, while others argued that a DPLA should play a stronger advocacy role in pushing for legal reforms that would help increase access, for example, to orphan works.

III. Characteristics of public domain collections and open business models

In the workshop’s second session, participants examined existing collections of public domain content, exploring issues of cooperation among existing providers, aggregation, possible financial models, and integration. The session focused largely on the importance of learning from past efforts, both domestically and outside of the United States, to make content available online. Participants discussed a number of possible models for aggregating and providing access to content, ranging from a completely centralized system to a federated search or portal that would rely on content hosted by independent initiatives.

While acknowledging the potential benefits of a centralized system, some participants raised concerns that such a system would threaten the identities of local libraries. Most participants agreed that a DPLA will likely need to incorporate a blend of the centralized and distributed models.

Some participants noted that a DPLA should consider how best to motivate current content providers and collections to participate, ensuring that individual institutions are given proper attention and exposure. Other participants expressed worries that approaching a DPLA with this as a primary concern would mean focusing less on the needs of users and argued that user needs should take priority over preserving the identities of individual collections.

Another major question raised was the relative importance of the content versus its container. Some participants argued that content is more important than its form and that a potential DPLA should focus on making its data open, rather than on preserving original physical container. Others argued that the container is a crucial part of content and that divorcing text from its physical form removes a substantial part of its context, possibly decreasing its value for future research needs.

With respect to financial and business concerns, participants noted that even public domain materials are expensive and time-consuming to digitize and make available online. Some suggested that using crowdsourcing to help collect and improve metadata may ease the burden somewhat.

IV. The more difficult questions: content with complex barriers to access
In the workshop’s third session, participants explored issues related to content with more complex barriers to access and use. Participants generally acknowledged that making fully- or partially-restricted content available through a DPLA would require significant levels of investment and curation, and that if a DPLA is to make this content available to users for free, it will need to find an appropriate place to which to shift the costs.

Some participants raised the possibility of approaching publishers with a “moving wall” or embargo proposal—effectively waiting to make copyrighted content available for distribution through a DPLA for a set time period after publication. The term of this agreement might be six months for newspapers and several years for other publishers.

Another model raised is the open access model, which focuses largely on making academic journal articles freely available online. Some participants suggested that a DPLA work with universities to encourage or perhaps require faculty to participate in open access publishing, while others argued that this is impractical and that the open access model does not necessarily work in all fields.

Some participants pointed out that current pricing and licensing schemes may change once publishers have greater access to data about the level of demand for digital versions of their content.

One of the major questions raised in this session was the question of tiered access, which could take a number of forms. One possibility suggested by participants is to completely avoid tiering, making all content available with the same restrictions to all users. Another possibility is to implement a detailed tiering model that may allow a DPLA to include a much larger set of content while making different pieces and collections available to users under different restrictions. A third possibility is to work with publishers so that users are unaware of the license and/or business model behind particular pieces or collections, but that a DPLA includes a mechanism to ensure payment for copyrighted content to different publishers.

The question of whether a DPLA should act as an advocate for new copyright legislation was raised again in this session. Many participants argued that work on a DPLA should not wait until new legislation is passed, instead striving to overcome some of the existing issues in order to help make legislative change easier.

V. Summary and key questions

A DPLA has multiple stakeholders
Participants emphasized that a DPLA includes multiple stakeholders, not just users—existing libraries, publishers, and authors all have a stake in a potential DPLA and in navigating digital content needs and uses in the future. Librarians often know the most about content trends; publishers and authors stand to benefit from access to the audience of a potential DPLA; and both libraries and publishers may find value in the metrics a DPLA could offer.

A DPLA should be focused on a mission, rather than a blueprint
Participants expressed hopes that a DPLA will be defined by its mission rather than by a set of predefined uses for specific user groups. It should be open to encourage innovation and creativity and to allow new services to be built on top of its existing layers.
**A DPLA has multiple uses, not distinct users**
Rather than defining public and academic users as separate entities, a DPLA should operate under a framework that allows for a range of different use cases.

**A DPLA is not only about the content**
Throughout the workshop, participants stressed the need for open data, open metadata, and APIs. Building these generative tools into a DPLA will allow for as-yet-unknown uses of the content within; they will enable users to build services on top of what a DPLA itself can provide, enhancing its value.

**The word “public” should be used carefully**
Many participants noted that the word “public” may prove challenging for a DPLA. Some expressed concerns that a DPLA may inadvertently take public funding away from existing public libraries, while others pointed out that a DPLA could help drive attention to public libraries. Many participants emphasized that a DPLA will support, not replace, existing public libraries.

**Copyright problems are the biggest issue facing a DPLA**
In order to get a DPLA off the ground, work may need to focus on public domain content first and approach trickier content incrementally. Other participants, however, noted that content in the “yellow zone” of copyright—out of print and orphan works—outnumbers both public domain and in-copyright works, making legal reforms necessary for the success of a DPLA. Most participants agreed that a DPLA that includes only public domain content is not sufficient.

Within this question is the question of tiering: some participants rejected the idea of tiered access, arguing that all content should be universally free to all users; others were more open to exploring tiering models.

**VI. Next steps**
The workshop concluded with a discussion of what participants and the wider community can do to work toward the establishment of a DPLA.

*Workstreams*
This initial workshop was organized around a single topic—content and scope—while raising questions that fit within a number of other areas. The Steering Committee has defined five workstreams, outlined on the [public wiki](#); community members have identified two additional workstreams (Audience and Interactivity). An important next step is to organize efforts around each workstream, collecting information on the wiki:

- Audience
- Content and Scope
- Financial/Business Models
- Governance
- Interactivity
- Legal Issues
- Technical Aspects

*Workplan*
Within these workstreams, community members should work toward rough consensus that can be
incorporated into a workplan for future action. This workplan is under development by the Steering Committee and will be released in the late spring.

**Future Workshops**
The hope is that each workstream will have working meetings once the project is fully up and running, and also that the DPLA planning initiative might be able to host a large, public meeting (in essence, a “plenary” type meeting where all the workstreams come together) later in the year. Various organizations have offered to support meetings in the near future, which we much welcome and encourage others to propose, as well.