DPLA Financial/Business Models Workshop Notes
13 March 2012 – University of Michigan, Ann Arbor

Introduction
Paul Courant

The DPLA hasn't yet specified its exact scope or preferred audience. We have an opportunity to contribute to what the products, services, and attributes might be. Two different types of DPLA are possible: (1) a DPLA that would have at least some, perhaps a great deal of in-copyright work, or (2) a DPLA that does not make or does not try to make those works available and offers only those works in the public domain.

While, as a practical matter, in-copyright materials will likely not be part of DPLA anytime soon, there are several elements of what the DPLA does that will require financial/business model support. The DPLA will have an internal operational structure (utilities, pay people to turn lights on and answer mail, etc.), and may be responsible for developing collections and digitizing work not previously digitized.

Discussion
A brief discussion following this introduction centered around three things: (1) the functions, roles and infrastructure that will cost money, (2) potential sources for that money, and (3) what resources are already available.

Participants suggested a few functions, roles and infrastructure that would absolutely require funding streams. Covering the cost of central administration is critical. If there is to be an elaborate suite of middleware, the DPLA will need some number of developers and an executive director. Relating to funding, the question was raised as to whether DPLA would be embedded in another organization that may require overhead.

Existing grants are startup funds, which will be available for only 18 months. Stewardship of existing grant funds has to lead to a place where we are financially sustainable; the DPLA must be sustained in other ways. Many DPLA contributions may be in-kind donations (and we won't always want them). Many of the roles, functions and infrastructure will not be funded by selling a product, and the cost of these will matter (Half a million dollars? Three million dollars?).

Finally, participants noted that to understand the funding priorities already defined, the Finance/Business model workstream needs access to the current budget and grant proposal. It was suggested that if the workstream had a well-defined study to propose, the Berkman Center might allocate grant funds for that purpose. We may be able to bring in help and get support for research and consulting projects.

Updates from the Content & Scope workshop and the Governance workshop
Blane Dessy & Jeremy York

Content & Scope Update
The content & scope (C&S) workstream listed three tiers of content to discuss:
Tier 1: Metadata aggregation of existing, digitized, public domain and cultural heritage collections – a topic they discussed in the workshop
Tier 2: Metadata aggregation of licensed and/or copyrighted material, which will be discussed in the future
Tier 3: Content repository for user-generated content, which will be discussed in the future

For the metadata discussions, the workstream talked about policies around CC0 licensing (waiving all rights) and CC-BY licensing (attributing the work as specified by the author or licensor) and the node structure of DPLA and what makes a node (a content aggregator).

Two parts were relevant to the finance workstream from these discussions that we could begin to build a business model around: incentives to contribute metadata to the nodes, and incentives for groups to become nodes.

Governance Update
The Governance workstream spent time on open meeting laws, reviews of analogous models, and mission and vision goals. They plan to develop an elevator speech vision and a four-page mission statement. There were discussions around core functions of advocacy and lobbying. Participants suggested DPLA should be inclusive of all humans and not necessarily have an academic focus. Ultimately, whatever we are doing should be in collaboration with existing libraries and associations. We must not be out of sync with the larger library world. The conceived organizational structure is hoped to have minimal bureaucracy, but the DPLA should promulgate an organizational, executive structure in the near term.

Discussion
The discussion generated several questions for other DPLA entities, a central one being "What kinds of content (format, topic) will be in the DPLA?" The DPLA concept currently has no limits or edges, and without these it is difficult to shape the DPLA, market the DPLA, or create a business model for the DPLA. The workstream recognized that the C&S group needed to resolve questions around metadata and licensing first, but relevant questions about content remain. It was agreed that the overall lack of centralized coordination among workstreams, with each workstream a silo, causes confusion for financing.

The workstream suggested one suite of resources for the DPLA: government documents, which are works in the public domain but not available anywhere else centrally and digitally.

Regarding governance, all agreed that the DPLA needs an Executive Director sooner rather than later that is dedicated, paid, fulltime, and empowered to make decisions.

Discussion also focused on public libraries. There was concern that public and state libraries would view the DPLA as a competitor for scant resources. However, participants suggested asking those stakeholders what the DPLA could enable them to scale up and do that is in line with their missions but that they cannot do now.

Business models for working with existing collections & institutions
Tom Sanville

It costs nodes a lot of money to collect, ingest and enable adequate metadata, ensure links to content are correct, manage rights and execute take-down requests, and perhaps develop a sustainability plan. The higher the minimum metadata standards are, the more costly but the more usable the data may be. Participants also noted that in addition to one-time costs needed to clean up metadata, nodes or the DPLA may also need continuing monies for ongoing functions at the aggregation level. The design of the system can make this more or less expensive.

One business model to address this relies on the assumption that nodes may already perform some of these functions and thus may not need to seek outside funding to accomplish them. The alternative to this situation is that nodes do need funding to fulfill their duties. If they need one-time funding to clean up their metadata, this money can come from the DPLA or grant institutions or the node themselves. One possibility is that having metadata (or collections) meet the minimum requirements for the DPLA becomes a regular grant requirement. If nodes fund their metadata cleanup themselves, they will need an incentive to pay to participate in the DPLA. There would still be baseline costs involved, such as mapping metadata; but the cleaner the data, the lower the cost.

The major incentive discussed for nodes to incur a cost to participate was enhanced search and discoverability. However, the nodes with the least pristine metadata, and therefore a higher cost, are likely to be the poorer entities. In a case where they don't have resources, participating in the DPLA is not an option without funding.

One way to deal with this problem is to offer metadata templates for a low cost (several hundred dollars) so that less well-funded institutions can use the templates make their ingest of metadata more efficient. For example, the DPLA could provide an open source version of a city directory, or a map, or set of page images that smaller institutions can use when submitting metadata.

The theme of engaging with local entities (historical societies, public libraries, etc.) cropped up often. As participants started to look more broadly at what DPLA might become, they asked what does DPLA provide that local institutions care about but cannot do themselves? Search and discoverability is one answer to this. In addition though, participants drew on the Tier 3 concept by posing a citizen-science model where local institutions (Civil War Societies, Wild Flower Societies, etc.) contribute user-generated content to the DPLA through nodes.

Discussion moved to a model where at some level of involvement in DPLA there would be face-to-face meetings where members of local historical societies, for example, get together

in a region to learn together and then contribute to the same node. For example, trainings on minimum metadata requirements could happen this way. This type of model would add value and distinguish DPLA from organizations like the Internet Archive and HathiTrust. These stories are very marketable. Consequently, participants are very interested in bringing representatives of local entities to the table.

We do need a node structure (aggregators) to pull together contributions. One concern is that we may want to represent an underserved community, but their offerings may be in low demand. We may need to start with a broader mass appeal, proceeding first with those who are willing and able, then skip the Tier 2 copyrighted material, and go directly to the user-generated content of Tier 3.

Potential models for working with publishers
Peter Brantley

For first time in American history the vast majority of new works published and discussed (digitally) in the marketplace are not available at no dollar cost to every American through public libraries. Public comments about the DPLA seem to reference tackling this problem, as does emphasis given to it by the Content & Scope Workstream (Tier 2 content). So, there is some interest in working with publishers to capture these resources in the DPLA.

However, right now the situation with the largest publishers and their relationships with libraries is dire. Only two are currently providing their books in digital form, and those come with some serious constraints and costs. Realistically, the large publishers are fighting other games and are not interested in this market; they don't want to lose revenue on digital sales because of lending. A service like Amazon Prime Lending entirely disintermediates libraries by offering free shipping and free digital lending for a growing collection of books.   DPLA will not get books from the big 6 publishers anytime soon.

Publishers that are small or medium in size are still willing to have titles acquired by the Internet Archive. The small and medium publisher market is a market of potential interest to a national and digital library effort, but this too may not yield much acquisition. The premise of establishing a national lending service based on licensing agreements is not an easy thing to do. Legal costs would be very high.

There is a growing self-published market, and self-published authors represent a category of rights holders who are alive and are not caught up in larger entities. Academics, too, are increasingly publishing through these means to broaden their audiences. However, Amazon is looking to enter this market, and then that content would likely not be available through DPLA.

Journals have similar issues, but they are already made more accessible by aggregation. Open access journals could be receptive to providing access through DPLA.

In general, participants thought that the DPLA should wait on this Tier 2 initiative and see how the landscape plays out before proceeding. This is an important issue, but it is not on the critical path for a launch or first stage of DPLA. It could be funded separately as needs

arise. It was determined that for now our role with regard to publishers is advocacy and not action. Whether the DPLA should fund advocacy issues at this point is in question.

Inventory of DPLA "things to be financed"
Kevin Guthrie

There are several major areas of expenditures that are ongoing. The first area of infrastructure is **staffing**: The DPLA will need an Executive Director by April 2013 if not sooner. The organization will also eventually need a program manager, and liaisons or advisors. Liaisons may reach out to local communities or interface with the nodes. Advisors may help make technical decisions moving forward while taking into account the ideas stemming from the workstreams. Software developers will also need to be involved, even if some DPLA functions are crowd-sourced.

Other things to be financed:
- **Technology Development and Operations**: There are central and non-central aspects, and this is the biggest area financially.
- **Content Development and Curation**: This affects most other items.
- **Digitization**: For new content, perhaps government documents.
- **Purchasing Collections**: Compensating owners of collections to take those into the public domain.
- **Legal Costs**: Some of these may be pro bono at first, or donated in kind, but it's best to expect a lawsuit at some point. We may also want to offer rights consultations.
- **Advocacy**: We may want to hire a lobbying firm for advocacy.
- **Help Desk/User Services**: With aggregation of data, questions likely can't be answered by a central resource. This is a new network-enabled problem.
- **Marketing and Outreach**: We need market research for both the supply and demand sides.
- **Building Social Networks**: of academics, the public, school libraries, and face-to-face meetings to support these connections among local entities

All of these functions that must be funded will cost different amounts depending on the strategies adopted by other workstreams. Thus, we developed a matrix to show how different strategies will alter the funding needs landscape for the DPLA (see **Appendix B**).

Group discussion

Discussion first generated a list of questions for other workstreams, the steering committee and the secretariat (see **Appendix A**). Then, two issues discussed earlier in the day cropped up again: First, the workstream re-visited the idea of engaging local entities in new and value-added ways. This workstream suggests the DPLA work with local historical societies and public libraries to become nodes, perhaps in an aggregated way. Second, a participant proposed that the workstream use government documents as a collection topic when launching the DPLA.

Participants noted two types of entities to reach out to in anticipation of the April plenary: representatives from both local organizations and major funding agencies to keep both sets of groups abreast of progress and to solicit their advice.

There was a candid discussion and some dismay at the way DPLA is currently developing. The organic method of development lacks a central decision-maker and that makes participants uncomfortable. Decisions must be made before this workstream can move forward. This workstream strongly urges the DPLA to determine a central decision-maker, perhaps the steering committee. Furthermore, the workstream agreed that now that all workstreams have had their initial workshops, the steering committee should take all of the input and re-express the direction and vision of the DPLA.

Finally, discussion looked forward to the April plenary. The general sentiment was that the plenary must produce something substantive to retain people's faith in the project. To this end, participants universally wanted to meet with other workstreams the day before the plenary. However, there was concern that this could be unproductive if it were an open meeting, so it was suggested that only the chairs meet.

Two significant documents to come out of this meeting are:

Appendix A: This list of questions for other workstreams, the steering committee, and the secretariat can help narrow the focus of the DPLA and help each workstream as they go forward.

Appendix B: The workstream created a matrix to help guide the other workstreams as they produce deliverables. The financial and business models workstream needs help from other workstreams in narrowing down this matrix and hopes that they take this into account when crafting agendas for future meetings.

Meeting Attendees
Paul Courant, University of Michigan (Chair)
Sarah Barbrow, University of Michigan
Peter Brantley, Internet Archive
Gregory Crane, Tufts University
Laura DeBonis, formerly of Google Books
Blane Dessy, Library of Congress
Kevin Guthrie, ITHAKA
Donna Hayward, University of Michigan
Shana Kimball, University of Michigan
Mark Kurtz, BioOne
Mary Morris, University of Michigan (Coordinator)
Jon Murley, Berkman Center
Tom Sanville, LYRASIS
Jeremy York, HathiTrust