Introduction
On February 1-2, 2012 a group convened at the College of Physicians in Philadelphia, PA. This group consisted of Content & Scope workstream conveners, experts in the field, conveners from the Legal, Technical, and Financial/Business Models workstreams, and members of the public. There was also remote participation through a conference line by other conveners and community members. The goal of the meeting was to discuss the first tier of potential content—existing digitized cultural heritage materials for the Digital Public Library of America, as well as methods of aggregation, standards, and policies.

Key topics to be discussed were stated in communications previous to the meeting and reiterated at the outset:
- Metadata aggregation of existing digitized cultural heritage collections
- Metadata aggregation of licensed and/or in-copyright material
- Content repository for user generated and/or other content

Summary of Discussion
The meeting began on February 1 with an introductory talk by Maura Marx, Director of the DPLA Secretariat at the Berkman Center for Internet & Society at Harvard. She discussed the overall goals of the DPLA and the specific task of the Content & Scope Workstream. Marx noted that there are several elements to be discussed/planned: (1) open source code, including a possible “app” store, (2) metadata, (3) content and the issue of copyrighted/non-copyrighted material, (4) tools and services including a decentralized approach, light applications, and a community of developers, and (5) community.

This first workshop also included a number of presentations from experts in the field. Lori Jahnke, digital services librarian at the College of Physicians, spoke about her work with the Medical Heritage Library (MHL). Jahnke presented on how the MHL functions and ways in which discipline-specific libraries could contribute to the DPLA.

On February 2, Terry Reese of Oregon State University Library kicked off the day by presenting on the lessons learned from building of “Libraries of Oregon” (LOO), an umbrella portal for public and university libraries in Oregon and a successor to the Oregon Digital Library. LOO is a collaboration between public universities and public libraries to: (1) help Oregonians find resources, (2) help Oregonians access resources, (3) help Oregonians connect to libraries, and (4) help Oregonians advocate for libraries in their communities. Extrapolating from that experience, Reese noted that the DPLA must not be seen as a substitute for public libraries and must position itself very carefully in order to
not jeopardize public library funding. He noted that there was and is still suspicion among the Oregon public library community about the goals/plans of the DPLA. Lessons learned from the LOO project include (1) all problems are local, (2) advocacy is key, (3) university partners need to get some value from the project, and (4) it's difficult to get different hierarchies to play together well. LOO is now in soft launch and will be officially rolled out in March 2012. To view the presentation slides, visit: http://www.slideshare.net/reese_terry/building-the-libraries-of-oregon-and-lessons-learned.

Sarah Shreeves of the University of Illinois at Urbana-Champaign presented next regarding the current NISO/IMLS Framework of Guidance for Building Good Digital Collections (http://framework.niso.org/) and how this effort intersects with the DPLA. The NISO/IMLS framework was first developed in 2000 and then revised in 2003 and 2007. It is hosted by NISO and is currently seeking funding from IMLS to update it. An update would include a section on aggregation. This section would talk about principles for aggregator services, as well as recommended specifications for digital library vendors to facilitate data harvesting. Another section would be on APIs, touching on why digital libraries need to provide this alternative means of access.

Martin R. Kalfatovic of the Smithsonian Institution and Biodiversity Heritage Library (BHL) reported from the DPLA Technical Aspects Workstream and presented information about the DPLA platform, which is now being developed by the interim technical development team. View slides from his presentation at: http://www.slideshare.net/Kalfatovic/technical-aspects-workstream-update.

The point of these presentations was to give a different perspective on the DPLA project and provided workshop attendees with a context from which to begin discussing the various issues and policies to be considered for the DPLA’s content & scope.

Workstream chair Rachel Frick thanked the speakers and reviewed the agenda. In order to generate ideas and allow for more discussion, the workshop then divided into three small groups. Sarah Shreeves, Robin Dale and Emily Gore served as the small groups’ facilitators.

Each group was charged with drawing up tangible solutions regarding the following:
- What is the language we need to draw the picture of a community of aggregators?
- If a library or content holder wants to contribute, what do they need to declare about copyrights?
- What about home institutional branding?
- Where is the metadata?
- What are those main topics and policy areas that are important in describing this national network of aggregators?
- How do we start developing the instructions?

After the breakout sessions, each small group presented on what was discussed. From here, the group created a large list of all topics and policy areas that needed to be covered. Each participant was asked to indicate the topics they found most important, allowing the group
to rank which ideas were the most pressing or would influence other decisions as driving principles. The top three were:

- Metadata Policy (CC0 vs. CC-BY)
- What makes a node (content aggregator) in a national network of aggregations?
- How to leverage existing distinct vocabularies and ontologies

These ideas were discussed thoroughly, in order to nail down specific principles and policy recommendations for the DPLA. Summaries contributed by individual workstream conveners are in the Key Takeaways section below.

The metadata conversation led into the related issue of having a low enough barrier for organizations to contribute data, but having rich enough data to support the services as envisioned by a possible DPLA “app store.” The group ran out of time to discuss the third topic of reuse of ontologies and existing vocabularies; it was decided to take the topic to the listserv for broader discussion.

Key Takeaways
Workshop participants identified three key areas for further research by Content & Scope Workstream members. These areas have been drawn up into proposals below; these proposals will be submitted to the Steering Committee for consideration after a public comment period.

1. DPLA Metadata Rights Management Considerations (developed by Betsy Kruger)
At its February 2, 2012 meeting in Philadelphia, the DPLA Content & Scope Workstream identified rights management for contributed metadata as one of its top issues. While metadata records that are purely factual are not copyrightable, creators of descriptive metadata records (particularly in the case of images where the line between descriptive data and scholarship blur) would, in some cases, assert that they hold intellectual rights to such descriptions.

Guiding our discussion was the goal of connecting as many users to as much content for as much reuse as possible. To this end, the workstream discussed two options: A) That metadata providers agree to authorize DPLA to release all contributed descriptive metadata into the public domain using the CC0 public domain dedication; or B) Implementing a tiered approach whereby metadata providers can choose between either the CC0 license or the CC-BY license, which permits others to distribute, remix, tweak, and use commercially metadata contributed to the DPLA, as long as attribution to the creator is made.

A. CC0 Option (http://creativecommons.org/about/cc0)
CC0 would enable metadata contributors to the DPLA to waive any interests in their metadata and thereby place it “as completely as possible in the public domain, so that others may freely build upon, enhance and reuse the works for any purposes without
restriction under copyright or database law."1 The British Library, some other national libraries, and the DPLA’s European equivalent—Europeana—have all released their descriptive metadata as CC0.2 Europeana has also implemented non-binding guidelines requesting that users of metadata contributed to Europeana “actively acknowledge and give attribution to all the sources” of the metadata.

The Content & Scope Workstream leans towards requiring metadata providers to use the CC0 license, primarily because this would allow the DPLA to use a linked open data approach to leverage and improve access to digital collections around the country. The linked open data approach requires that metadata be published under a license that allows it to be freely used and reused for any purpose.3 If DPLA chooses CC0 as the only option for rights management of descriptive metadata, our website should clearly explain why. We would also recommend implementing similarly non-binding usage guidelines, such as those employed by Europeana,4 which encourage all metadata users to acknowledge and give attribution to the metadata provider.

B. Two tiered option: CC0 or CC-BY (http://creativecommons.org/licenses/by/3.0/)
The workstream discussed the fact that art museums are frequently concerned about maintaining rights to the descriptive metadata they create, and so having CC0 as the only option for metadata contributors may result in a loss of valuable content providers. As a possible solution to this problem, we looked at a two-tiered option whereby metadata contributors could choose between CC0 and CC-BY. The CC-BY license permits others to distribute, remix, tweak, and use commercially metadata contributed to the DPLA, as long as attribution to the creator/distributor is made.

If the DPLA were to implement this two-tiered approach, the workstream strongly recommends that our website clearly states that CC0 is strongly preferred; outlines the reasons why; and also outlines the limitations that would be placed on descriptive metadata covered with a CC-BY license. Our hope would be that by clearly articulating the limitations imposed by CC-BY, most metadata providers would choose the CC0 license.

A DPLA discussion that focuses on the question “what is the optimal use case?” might help to prioritize how tough-minded we are about the CC0 option. If, for example, the most important use case is the curious person on the street, we might want more content that comes from sensitive or reticent owners that such a user will be glad to have represented

1 CC0 Option (http://creativecommons.org/about/cc0).
2 See for example “Europeana adopts new data exchange agreement, all metadata to be published under CC0.” (http://creativecommons.org/blog/entry/29133).
in DPLA. If enabling this type of use is our priority for the DPLA, then a two-tiered option might be best. But if our highest priority is remix/reuse and the advantages offered by a linked open data approach, then we really need to have CC0 only content or (as the two-tier plan allows) different filters for different uses. One direction that Europeana is pursuing is to provide linked open data capacity with the aim of generating spillover economic gain. Some European museums have concluded that they will contribute only sparse records to Europeana; while users lose something in terms of search, the museums can retain as much of what they perceive to be their intellectual property as possible.

2. DPLA Minimal Metadata Requirements for Consideration (developed by Emily Gore and Sarah Shreeves)
At the February 2, 2012 meeting in Philadelphia, the DPLA Content & Scope Workstream discussed minimal metadata requirements for aggregation. The landscape of current digital projects was discussed, and the conclusion reached was that most projects are using Dublin Core or Qualified Dublin Core as a metadata standard, usually viewed as minimal metadata standards in the library community. Others are using MODS/METS and other richer content standards. It is obvious that the more robust metadata provided, the more sophisticated search facets can be offered. With balance in mind and a goal of broad participation, the group suggested the use of the DLF Aquifer metadata guidelines as a framework to guide the DPLA metadata requirements. Below, find a brief, modified version of the this framework for DPLA aggregations:

A. Minimum for participation: allows user to cite resource
The minimum for participation level defines the information necessary for the most basic indexing/search/retrieval of records. The Content & Scope Workstream recommends that the DPLA not harvest records that do not contain this information. This information should include:

- title
- creator (if known)
- date
- name and contact information of owning institution
- rights information
- URLs (permanent URLs are encouraged)

B. Richer metadata means richer search and retrieval
Metadata schemas tend to have required elements as well as elements that are not required but that aid in the discovery of an item. Examples of these elements include language information, subject headings, detailed descriptions of contents, physical descriptions of source material, permanent URLs, object types, etc. The inclusion of this metadata allows for better discovery and reuse of the digital objects.

3. Statement on Contributing to DPLA via Aggregations (developed by Robin Dale)
At the February 2, 2012 meeting in Philadelphia, the DPLA Content & Scope Workstream discussed the most viable way to contribute content to and build the DPLA between now and 2013. The unanimous response was that the most efficient and effective way would be through leveraging existing aggregations of content such as state collaboratives, large-scale
subject or format-based aggregations (i.e., HathiTrust, IMLS-DCC, ARTStor, JSTOR, Internet Archive), federal repositories, etc.

The Content & Scope Workstream recognizes that there is a strong advantage for the DPLA to harvest metadata and/or content from these “nodes” on a greater DPLA network rather than with individual libraries of varying sizes and competencies. First and foremost, working with aggregators will allow the DPLA to leverage the “one-to-many” relationship that will be vital when addressing the number and range of potential DPLA content contributors. The benefits of working with aggregators include: the standards and best practices implemented by the aggregators, especially related to metadata creation and rights management; staffing and support at aggregators that may enable metadata correction, normalization and enrichment not found at the local organization or project level; and, the responsibility for rights management of the content within aggregator policies and systems. Recognizing that not all aggregators utilize the same guidelines or have the same capabilities, the group decided that the DPLA should create a framework outlining content contribution via aggregations including a set of minimal criteria for aggregators wishing to be a node on the DPLA contributors’ network. Below is a brief outline of a framework for DPLA content acquisition through aggregations:

A. Harvested Content, where possible, will be done at the Aggregator/Node level
The Content & Scope Workstream recommends that the DPLA investigate harvesting metadata and content from aggregators and clearly define the role of these aggregators (“nodes”). Related recommendations include:
- Target and consider contributions from Nodes first and include subject-specific, format-specific, collaborative repositories and other opportunities to optimize content contributions while minimizing the initial number of contributor relationships to be managed.
- Include contributions from individual organizations at a later date
- Ensure that Nodes are able to conform to DPLA Node Responsibilities (see below)

B. Incentivizing Nodes
DPLA may need to offer incentives to aggregators in order to gain participation. The Content and Scope Workstream will further discuss this concept to ensure that target content can be included in the DPLA.
- Investigate whether funding can be identified, leveraged and forwarded to Nodes to aid in these responsibilities (e.g., LSTA)

C. Node Responsibilities
Under this proposal, DPLA Nodes (aggregators) will have certain responsibilities to fulfill as DPLA-approved aggregators. These are not necessarily added responsibilities for some, but will be key contributions made to the greater DPLA network. The workstream proposes that Nodes be distinguished based on levels of service they can provide and be reviewed based on technical requirements, availability, security, sustainability. The workstream also recommends that Nodes be responsible for meeting DPLA Requirements for Partner Nodes (brief checklist in process). These responsibilities include:
• Collect, ingest, and enable minimal level metadata to be harvested by DPLA (as outlined by *DPLA Minimal Metadata Requirements for Consideration*):
  o Where metadata is normalized/corrected/enriched, allow DPLA to harvest this metadata, per the DPLA statement on *Metadata Rights Management Considerations*.
  o Should nodes be able to host/provide content as well as metadata?
• Ensure links (URIs) to content and metadata are correct and make corrections, as necessary
• Assume responsibility for implementing rights management metadata from content owners
• Nodes may provide metadata enrichment services; the resulting content would be utilized by the DPLA
• Assume responsibility for following up on any “take-down” requests made directly to Nodes or made to the DPLA and subsequently forwarded by the DPLA
• Have an identifiable plan for sustainability for the data, including data handoff, if necessary.
• Nodes may not necessarily be mandated to be “preservation services” for content they have in their possession.

Participants

<table>
<thead>
<tr>
<th>Name</th>
<th>Institution/University</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marguerite</td>
<td>MIT Press</td>
</tr>
<tr>
<td>Robin</td>
<td>Lyrasis</td>
</tr>
<tr>
<td>Blane</td>
<td>Library of Congress</td>
</tr>
<tr>
<td>Rachel</td>
<td>CLIR/DLF</td>
</tr>
<tr>
<td>Emily</td>
<td>Florida State University</td>
</tr>
<tr>
<td>Dave</td>
<td>UC Berkeley School of Law</td>
</tr>
<tr>
<td>Rebekah</td>
<td>Berkman Center for Internet &amp; Society</td>
</tr>
<tr>
<td>Glen</td>
<td>Lightbeam Group</td>
</tr>
<tr>
<td>Robert</td>
<td>IMLS</td>
</tr>
<tr>
<td>Lori</td>
<td>The College of Physicians/Medical Heritage Library</td>
</tr>
<tr>
<td>Martin</td>
<td>Smithsonian Institution Libraries</td>
</tr>
<tr>
<td>Betsy</td>
<td>UIUC</td>
</tr>
<tr>
<td>Maura</td>
<td>Berkman Center for Internet &amp; Society</td>
</tr>
<tr>
<td>Jim</td>
<td>Columbia University</td>
</tr>
<tr>
<td>John Mark</td>
<td>University of Pennsylvania</td>
</tr>
<tr>
<td>Terry</td>
<td>Oregon State University</td>
</tr>
<tr>
<td>Amy</td>
<td>Boston Public Library</td>
</tr>
<tr>
<td>Sarah</td>
<td>UIUC/IDEALS</td>
</tr>
<tr>
<td>James</td>
<td>ARTstor</td>
</tr>
<tr>
<td>Julie</td>
<td>Georgetown County Library</td>
</tr>
<tr>
<td>Jena</td>
<td>CLIR/DLF</td>
</tr>
</tbody>
</table>