An Exploratory Analysis of Subject Metadata in the Digital Public Library of America

Dublin Core International Conference
Wednesday, September 2, 2015
São Paulo, Brazil
Content Hubs
• ARTstor
• Biodiversity Heritage Library
• David Rumsey
• Harvard Library
• HathiTrust
• Internet Archive
• J. Paul Getty Trust
• National Archives and Records Administration
• Smithsonian Institute
• New York Public Library
• United States Government Publishing Office
• University of Illinois at Urbana-Champaign
• University of Southern California Libraries
• University of Virginia Library
• Digital Commonwealth
• Digital Library of Georgia
• Kentucky Digital Library
• Minnesota Digital Library
• Missouri Hub

• Mountain West Digital Library
• North Carolina Digital Heritage Center
• South Carolina Digital Library
• The Portal to Texas History
February 2015:
23 hubs
8,012,390 metadata records
Why subjects?
Average Number of Subjects Per Record

- ARLster
- Arizona State University Library
- Harvard Library
- HathiTrust
- Internet Archive
- J. Paul Getty Trust
- Smithsonian Institution
- University of Illinois at Urbana-Champaign
- University of Southern California Libraries
- Digital Commonwealth
- Digital Library of Georgia
- Minnesota Digital Library
- Missouri Hub
- Mountain West Digital Heritage Center
- North Carolina Digital Library
- South Carolina Digital Library
- The Portal to Texas History

Content Hubs vs. Service Hubs
Maximum Number of Subjects in Any Record
Percentage of a Hub’s Records Containing Zero Subjects

![Percentage of a Hub’s Records Containing Zero Subjects](chart.png)
<table>
<thead>
<tr>
<th>Number of Total Subjects and Unique Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td>ARTstor</td>
</tr>
<tr>
<td>Biodiversity Heritage Library</td>
</tr>
<tr>
<td>David Rumsey</td>
</tr>
<tr>
<td>Harvard Library</td>
</tr>
<tr>
<td>HathiTrust</td>
</tr>
<tr>
<td>Internet Archive</td>
</tr>
<tr>
<td>J. Paul Getty Trust</td>
</tr>
<tr>
<td>Smithsonian Institution</td>
</tr>
<tr>
<td>The New York Public Library</td>
</tr>
<tr>
<td>United States Government Publishing Office</td>
</tr>
<tr>
<td>University of Illinois at Urbana-Champaign</td>
</tr>
<tr>
<td>University of Southern California Libraries</td>
</tr>
<tr>
<td>University of Virginia Library</td>
</tr>
<tr>
<td>Digital Commonwealth</td>
</tr>
<tr>
<td>Digital Library of Georgia</td>
</tr>
<tr>
<td>Kentucky Digital Library</td>
</tr>
<tr>
<td>Minnesota Digital Library</td>
</tr>
<tr>
<td>Missouri Hub</td>
</tr>
<tr>
<td>Mountain West Digital Heritage Center</td>
</tr>
<tr>
<td>North Carolina Digital Library</td>
</tr>
<tr>
<td>South Carolina Digital Library</td>
</tr>
<tr>
<td>The Portal to Texas History</td>
</tr>
</tbody>
</table>

**Content Hubs**

**Service Hubs**
So what?
Questions?

Hannah Tarver  
hannah.tarver@unt.edu

Mark Phillips  
mark.phillips@unt.edu

Oksana Zavalina  
oksana.zavalina@unt.edu

Priya Kizhakkethil  
priyakizhakkethil@my.unt.edu