

Exploring DPLA's Content: A Discussion of its Strengths and Weaknesses

Authors: Megan Barney, Megan Woods, Sara Dean, Danay Vera, Katie Woods, and Patrick Johnston

The Digital Public Library of America (DPLA) is a nonprofit organization that began planning in 2010 and publicly launched in 2013. The aim of DPLA is to serve as an easily accessible digital center containing photos, books, maps, and more. On its homepage, DPLA boasts a conglomeration of more than 17 million items through which users may browse. As a repository of digital objects, DPLA is not limited to images of photographs or paintings, but also includes oral histories, newspaper articles, and political documents.

DPLA has worked with libraries, archives, and museums to assemble this digital library. DPLA's Strategic Plan 2015 through 2017 provides a breakdown of its more than 1,000 contributing institutions. Public libraries and university libraries provide the greatest share of DPLA's content, contributing 23.38% and 22.19% respectively.

Breakdown of contributing institutions, as of April 2014

(DPLA Strategic Plan, 2015-2017)

DPLA uses a "Hubs Model" to establish a network of digital collections. There are two types of hubs: content and service. According to DPLA, content hubs "provide more than 200,000 unique metadata records that resolve to digital objects" (DPLA "Hubs"). Some current content hubs include: HathiTrust Digital Library, Internet Archive, Library of Congress, National Archives and Records Administration, and The Smithsonian Institution. On the other hand, service hubs act on either the state or regional level to "bring together digital objects from libraries, archives, museums, and other cultural heritage institutions" that contribute to local, regional, and/or national history (DPLA "Hubs"). These service hubs, such as Caribbean Service Hub, Digital Maryland, and Mountain West Digital Library, assist DPLA in aggregating and maintaining the metadata of the content they provide (DPLA "Hubs"). Through this model, both smaller collections and larger institutions are able to contribute to DPLA. (See below for further information on DPLA's hubs and contributing institutions)

Visual representation of DPLA's Hubs model (<https://dp.la/info/wp-content/uploads/2013/04/HubNetwork-med.jpg>)

DPLA's "Exhibitions" page might be of particular interest to those studying public history. There are currently 34 different exhibitions listed, with titles varying from "American Empire" to "Recreational Tourism in the Mountain West." Each exhibition has six or seven sections, and each section consists of four or five images with curatorial interpretations. DPLA does not merely contain multitudes of images; DPLA staff, graduate students, and public librarians curate exhibits of images through the Public Library Partnerships Project. These exhibits allow visitors to explore images within certain topics and themes. Similarly, the site includes a collection of "Primary Source Sets." These small image collections revolve around interesting historical topics and provide links to primary sources that users can explore as they see fit. This function seems intended for teachers rather than students.

Curation of the entire database, however, would be a monumental task. The site's primary feature is its search function. The "Searching DPLA" page explains how to do an advanced search with quotations and other statements, but this research technique required some digging to find. Otherwise, users may view search results by subject, location, language, contributing institution, partner, "type" (for example, image versus moving image),

or date. The site's browse function initially seems to be limited; the homepage shows users that they may look at items by geographic location or time period, but users may not realize that they can browse by subject unless they check the footer at the very bottom of the page.

Given the site's magnitude, it is difficult to truly comprehend everything DPLA has to offer. However, there are ways for users to estimate its contents. DPLA has an Images app with a "surprise me" feature that allows users to simply survey examples of the kind of material the database contains. This returns a variety of results like "unicorn," "stuff," "mountain," "jazz," or "hippies." Yet the app does not represent the full breadth of DPLA's content, as it only displays 25 images per search. Despite its flaws, the search function on the main site is more useful in allowing users to view as many results as possible when exploring a subject. On its website, DPLA lists 120 different subject categories and displays the number of items within each category. Below are two lists: The first shows the "Top 10" subjects represented in all items, while the second shows the "Top 10" subjects represented in images only. Although these lists are similar, there are some discrepancies highlighting our earlier point about DPLA consisting not only of images but of digital "objects."

Top 10 General Subjects: All items	Top 10 General Subjects: Images
1. Plantae - 1010030	1. Plantae - 445243
2. Dicotyledonae - 674143	2. Dicotyledonae - 325082
3. United States - 659352	3. Ethnic Relations - 229913
4. Places - 575012	4. Holocaust, Jewish (1939-1945) - 227613
5. Texas - 569503	5. Asteraceae - 204471
6. Business, Economics and Finance - 517235	6. Asterales - 197650
7. Communications - 479236	7. Jews - 181694
8. Newspapers - 454807	8. Anthropology - 154886
9. Advertising - 446339	9. United States - 131911
10. Journalism - 435217	10. Places - 125243

Closer examination of the subject list raises an important concern. Since DPLA receives the metadata of each image from its hubs, the institutions or hubs assign subjects to the images, not DPLA. Therefore, institutions or hubs may classify subjects differently, resulting in inconsistencies in the organization of the database. For example, one institution might identify an image as "train" while another might choose to identify it as "railroad" instead. Due to these deviations, DPLA's subject search might not be a reliable means of accessing all available images.

For the purposes of this class, we are interested in a particular subset of DPLA's images: We would like to determine how many faces are contained in the database. In order to estimate this number, we looked at the amount of results that came up when we searched the subjects "people," "men," "women," "children," "group portraits," and "portraits." While looking at the pictures on this database we noticed that some pictures only fall under one subject while others fall under multiple subjects. Luckily, most of the subjects listed above do not appear on the same images. There are roughly 413,840 images that have people in them, though we could potentially be missing pictures that have people in them but are not listed under any of these categories. To compensate for this, we are including in our count the available subjects categories "Native American," "Jews," and "African Americans." Not every image within these subjects contains people, but many do, and this helps us make up for the difference created by group pictures only being counted under a subject once. These considerations brought us to a grand total of 754,326 faces. It is important to remember that this is a rough estimate only, and that there are probably more faces in DPLA than we were able to account for.

During our exploration of DPLA, we discovered gaps on an institutional as well as a content level. While DPLA already has many institutions helping with its mission, it would be wonderful to see more institutions contribute to the database in the future. Potential additions could include the US Holocaust Memorial Museum and The Center for Creative Photography at University of Arizona. DPLA could also benefit from acquiring the resources of another major digital collection, AP Images. These three previously mentioned institutions were not found to be contributing institutions on DPLA's website. Finally, The National Museum of African American History and the American Indian Museum currently do not contribute to the DPLA through the Smithsonian, while the other museums affiliated with the Smithsonian have some sort of contribution to DPLA. It would be interesting to conduct further research as to why these particular museums have not contributed digital items through the Smithsonian, as they would help to fill the gaps in DPLA's collection.

The DPLA could also be improved by greater specificity in its subjects. As of now, the list of subjects is plentiful but broad. Users can attempt to narrow down results by selecting multiple subjects, but this still requires them to work within the constraints of the categories that are currently offered by DPLA's contributors. This makes it more difficult to narrow results down to a specific subset of data. Take, for example, the subject "WWII;" when selected, this subject will bring up all the information regarding WWII in the database. A user can narrow down results by location, language, institution, medium, date, or additional subjects. Yet there are many different subtopics within the field of World War II studies that users cannot easily access because they are not included in DPLA's subject lists. DPLA could potentially remedy this by identifying the terms that are searched most frequently and adding those subtopics to the list for users to explore.

Alternatively, examining the most commonly searched terms could also inspire the curators of this vast collection to create exhibitions and primary source sets that cater to what people are most interested in. The current collection of exhibitions touch upon interesting histories, but many of them are niche subjects and may not help people who are looking to find content and information about more popular subjects.

Creating larger image and content collections based on the current exhibition models may be to DPLA's advantage. After all, their store of content is so great that dividing it into larger subject-centric databases, all with their own search functions, could help people who have trouble finding specific content among the millions of items in the site's collection. This could also serve as an entry point for people who are interested in browsing the larger collection, but don't know where to start.

For example, a user might be looking for an image for a project on Jim Crow laws. While some of the exhibits involve adjacent subjects, these might not satisfy the user's needs. Meanwhile, the database-wide search function would provide an overwhelming amount of content that the user might have difficulty narrowing down. While there would likely be useful content in the collection, sifting through hundreds (or likely thousands) of images would not be an efficient strategy for the user. If DPLA had a Post-Reconstruction Civil Rights sub-collection or database, it would be of great help to users in such situations. This would be a large undertaking, but the curators of this vast store clearly know how to construct subject-specific collections. This is just one solution to the problem, however. Undoubtedly, there are other ways of making this incredible collection of information more accessible to its user base.

In the four years that the DPLA has been active, its actions in forming a digital database combining various collections around the nation has been unprecedented. After exploring this site, it is clear that there is still a lot of work to be done to make DPLA the most efficient digital center for digital objects. Despite its shortcomings, the DPLA is advancing the digital humanities in a way that is beneficial to people worldwide. Moving forward, the DPLA can (and most-likely will) evolve as financial means and new technology become available.

Extra Information:

For full list of Content and Service Hubs: <https://dp.la/info/hubs/>

*For visual representation of Hubs: http://www.deanfarr.com/viz/partners_alt.php

*For visual representation of contributing institutions to each Service Hub: <http://www.deanfarr.com/viz/partners.php>

*Links found on DPLA's App page