Strategic Digital Collection Development in Academic Libraries
Kyle Rimkus, Kelly Applegate, and Jennifer Hain Teper
Preservation Services, University Library, University of Illinois at Urbana- Champaign

Introduction
This 2015 study uses web analytics, subject term analysis, and download statistics to gauge the presence, visibility, and popularity of University of Illinois digital collections on the web. With a grounding in current best practices, it provides a pragmatic methodology for the institutional analysis of digital collections with an eye to strategic digital collection development.

Goals and Methods
Collection characteristics. What are the common subject or format characteristics of the library's most-used collections? That is, what are the strengths of the current Illinois digital presence?

Collection access. By what paths are patrons most frequently arriving to the library's digital collections? What conclusions can be drawn from the way patrons discover this content vis-à-vis the way the content is accessible?

Research proceeded in three distinct steps:
1. Identify collections for analysis
2. Define collection properties for analysis
3. Gather and analyze data

Web Analytics
As part of this study, researchers collected Google Analytics web usage data from July 30, 2014 through August 1, 2015 for:
- ContentDM image collections: https://imagesearchnew.library.illinois.edu/
- HathiTrust: https://www.hathitrust.org
- IDEALS institutional repository: https://www.ideals.illinois.edu/
- IDEALS newspaper collections: http://idnc.library.illinois.edu/
- Digital Public Library of America: https://dpb.americanmemory.org

Data analysis is still underway, and the results presented in this poster are preliminary.

Digitized Books
The library has digitized over 400,000 books since 2007, locally, and in partnership with scanning centers managed by Google and the Internet Archive. The library's goal is to deposit all public domain books it scans into the HathiTrust Digital Library, and has adopted workflows to effect this.

HathiTrust Web Statistics
The library analyzed Google Analytics web statistics on the usage of its collections. In these logs, the library was able to differentiate between content that originated from local deposit, Google Books, and Internet Archive digitization of UIUC items. The word clouds below are formed from frequently recurring words in the titles of items accessed during the analysis period.

Referral Traffic
The study reviewed several thousand sessions and pageviews in Google Analytics to determine where web traffic comes from, in order to determine which referral channels are the collections' greatest assets. The numbers show that sharing metadata in aggregated systems like the Digital Public Library of America increases access, as does making metadata more readily discoverable via search engines.

Trends in Newspapers
Newspaper Collections
Statistics show a high interest in the library's digitized vaudeville newspapers, and newspapers from local communities, and less interest in its extensive digital holdings of farm community newspapers.

Top Downloads in IDEALS
This table presents the top fifteen document downloads in IDEALS for the year. By comparison, approximately 40,000 items were downloaded; the median number of downloads per item was 27 and the mean was 87.

Status of Project
This poster presents a preliminary visualization of selected data. The study plans to produce a paper for publication, dissemination, and presentation for the library community in early 2016.