

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 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

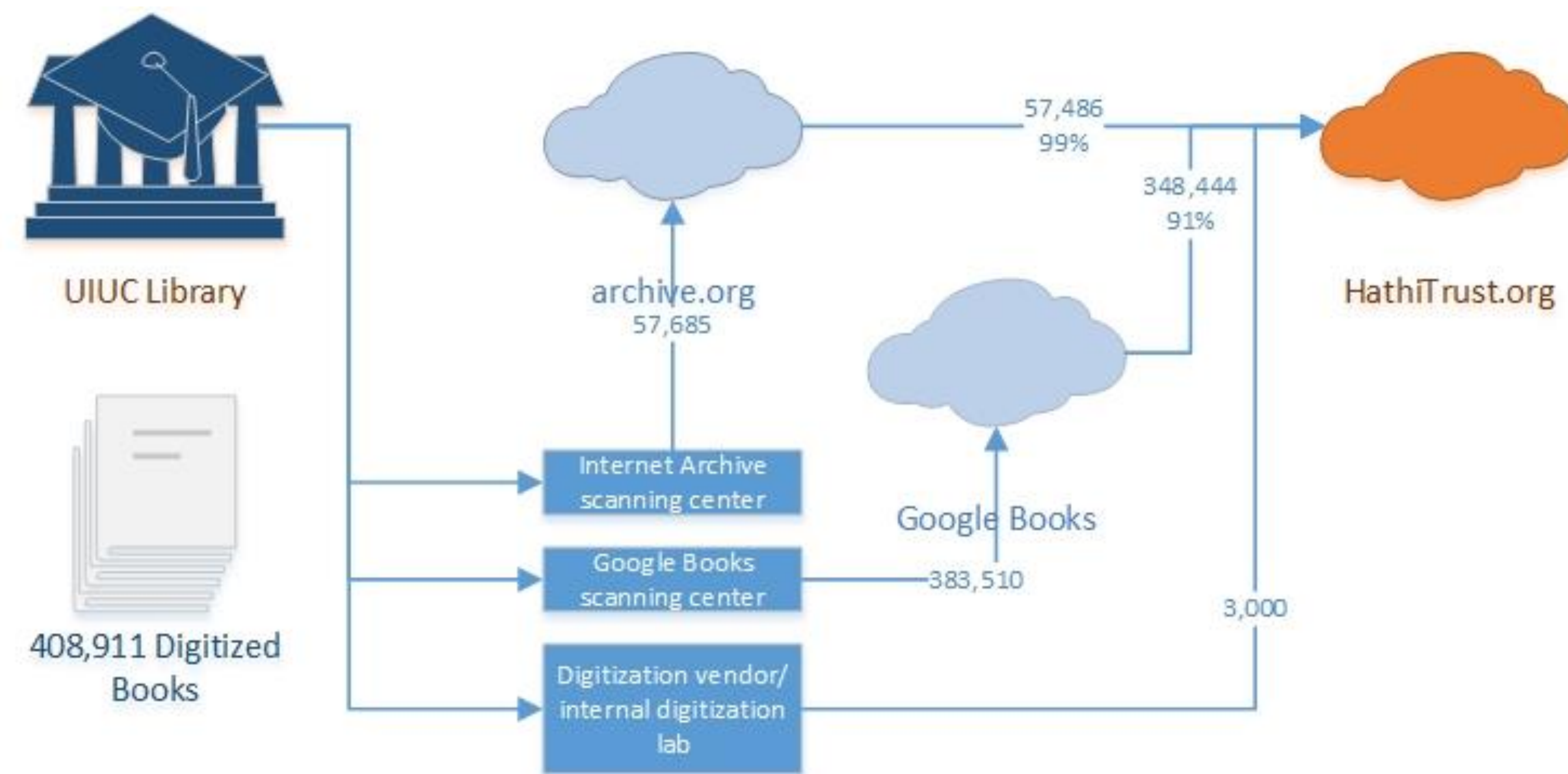
As part of this study, researchers collected Google Analytics web usage data from July 30, 2014 through August 1, 2015 for:

- **ContentDM** image collections: <http://imagesearchnew.library.illinois.edu/>
- **IDNC** newspaper collections: <http://idnc.library.illinois.edu/>
- **IDEALS** institutional repository: <https://www.ideals.illinois.edu/>
- **HathiTrust** digitized books: <https://www.hathitrust.org/>

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

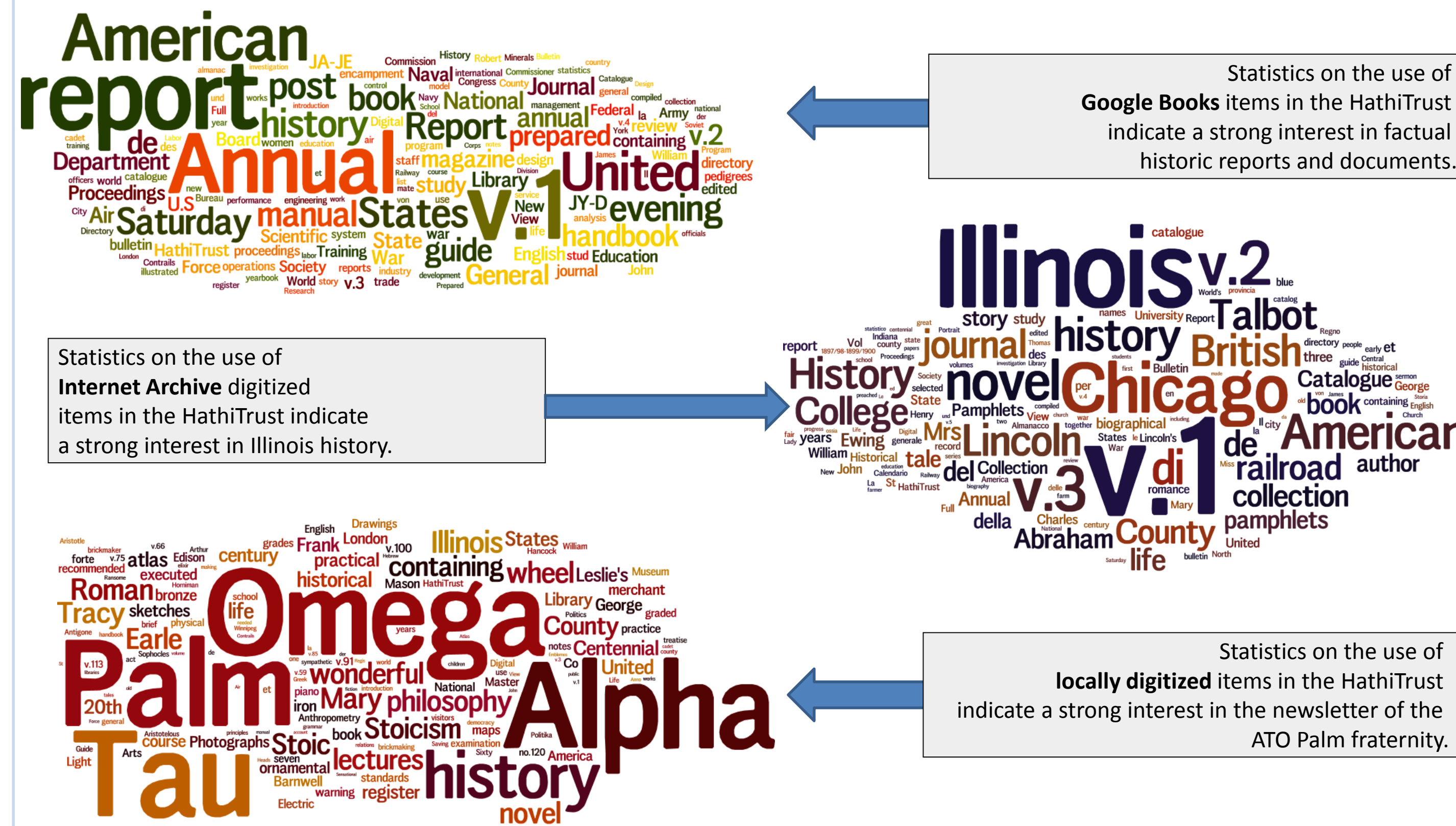
Digitized Books

The library has digitized over 400,00 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 this effect.



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.



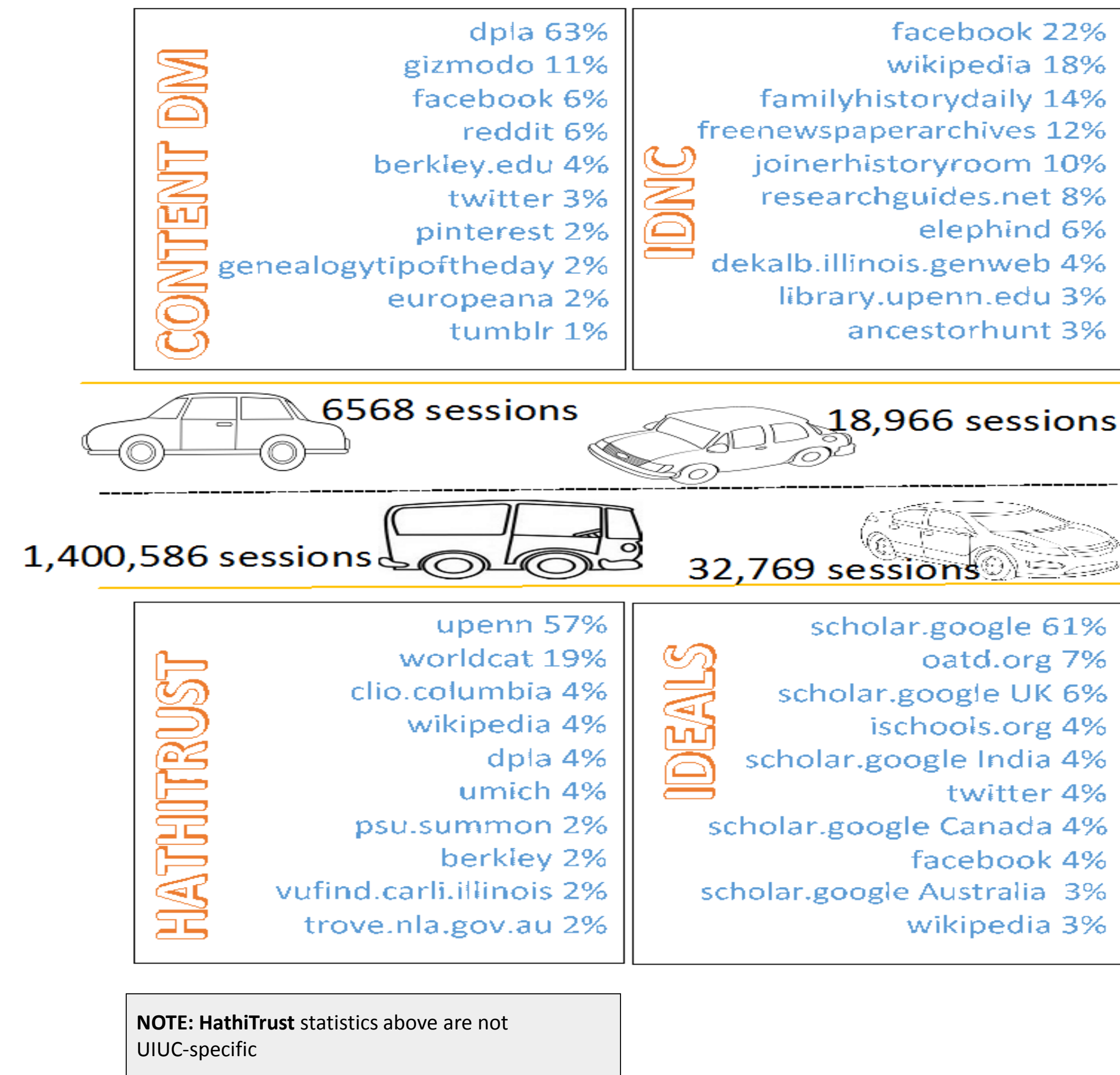
Statistics on the use of Internet Archive digitized items in the HathiTrust indicate a strong interest in Illinois history.

Statistics on the use of Google Books items in the HathiTrust indicate a strong interest in factual historic reports and documents.

Statistics on the use of locally digitized items in the HathiTrust indicate a strong interest in the newsletter of the ATO Palm fraternity.

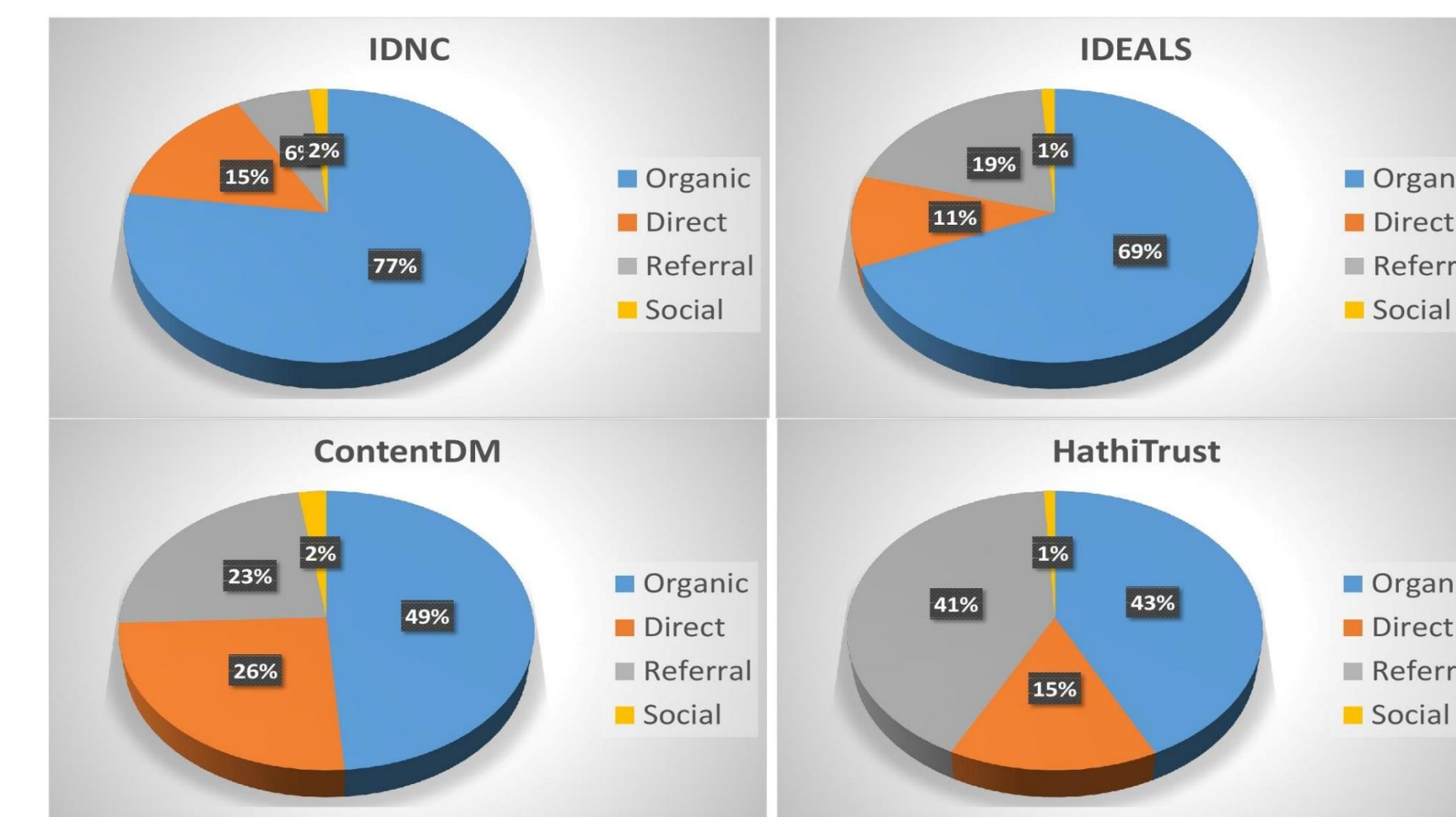
Referral Traffic

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.



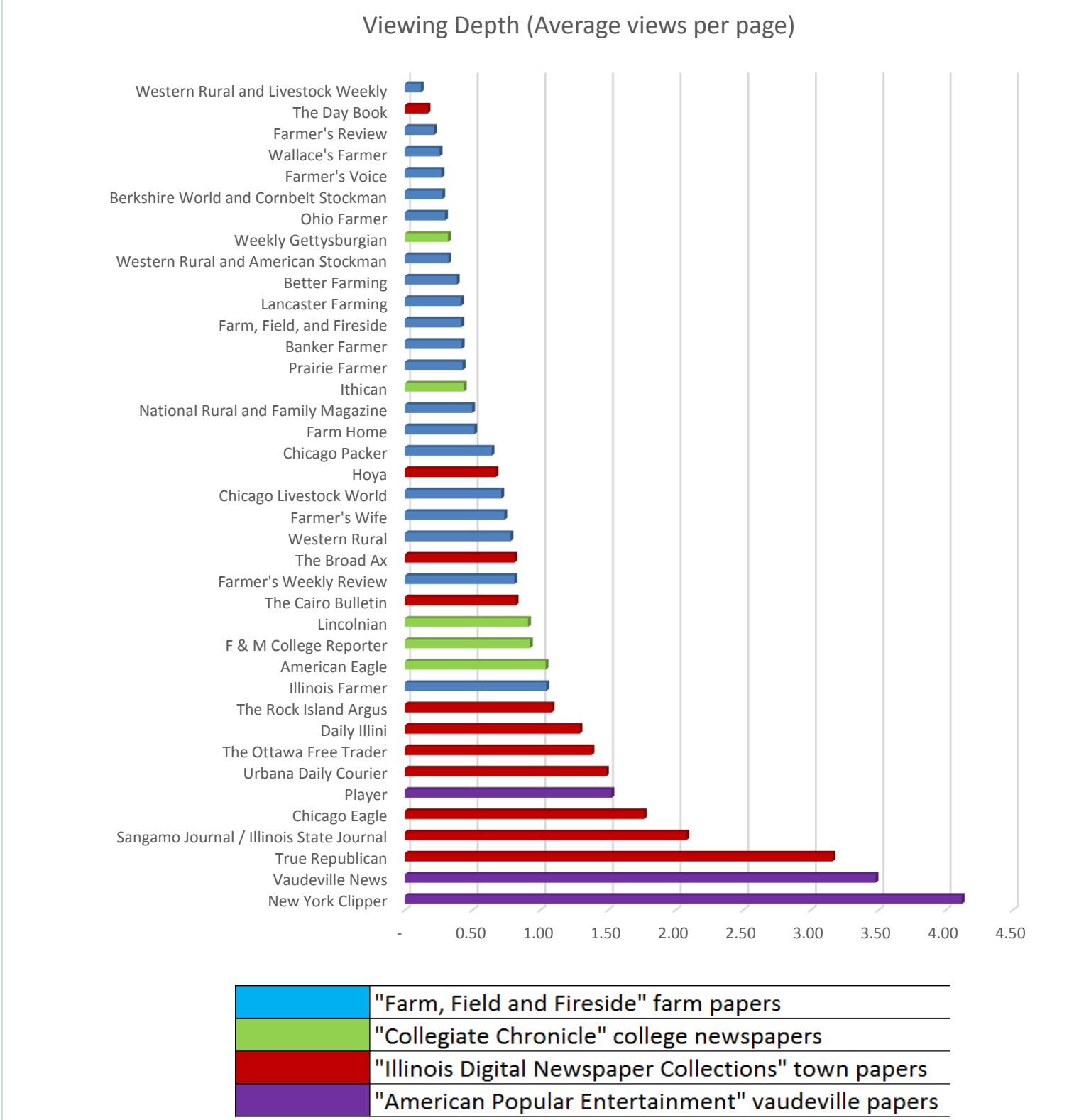
Organic Searching

Organic search traffic is defined as traffic that arrives at the UIUC collections based only on a keyword search in a typical search engine online, and our collection appearing in the list of retrieved results. From the results, the user clicks through to our site. Overwhelmingly, the digital collections in this study get the largest majority of traffic from organic web searching.



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

Title	Downloads
Estrategias para Motivar el Aprendizaje Colaborativo en Cursos a Distancia	21,444
The product life cycle : its role in marketing strategy/some evolving observations about the life cycle	19,452
El aula virtual: Usos y elementos que la componen	18,574
Heat Transfer in Refrigerator Condensers and Evaporators	16,279
Gender Differences in Leadership	16,117
Managing Resistance to Change	12,083
From love at first sight to soul mate: Romantic ideals in popular films and their association with young people's beliefs about relationships	10,928
Research on teaching reading comprehension	9,014
Analyses of differences between written and oral language	8,841
Smartphones in the U.S.: Market Analysis	8,553
Riveted And Bolted Structural Joints	8,490
Teaching the process of reading comprehension	7,538
Motion and time study (Institute of Labor and Industrial Relations. Bulletin No. 24)	6,960
Vocabulary instruction and reading comprehension	6,756
Motivation-need theories and consumer behavior	5,774

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.

