DATA-DRIVEN DECISION MAKING FOR DIGITIZATION

Ryan K. Lee, Cory L. Nimer, J. Gordon Daines III, Sam Dearden

Brigham Young University

INTRODUCTION

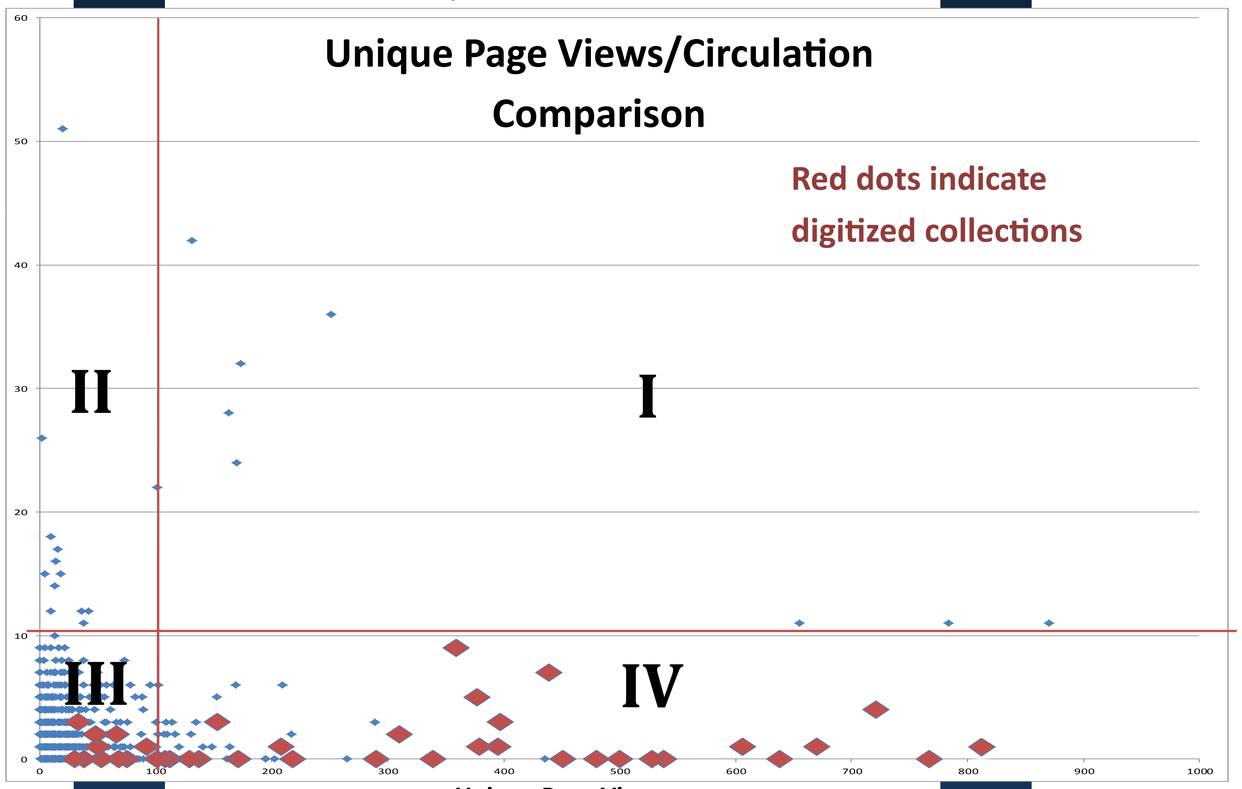
Curators in the L. Tom Perry Special Collections currently make digitization decisions based on personal choices and criteria, with little to no backing from data. This project is an attempt to compile and use two sets of statistical data—reading room circulation and Web analytics—to help curators make data-driven digitization decisions.

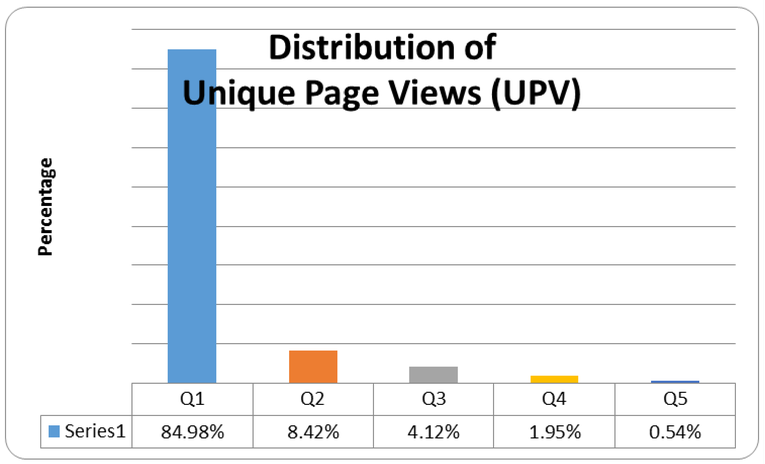
METHODOLOGY

- Retrieve and aggregate Web statistics (Google Analytics) for Unique Page Views for 2012-2014
- Combine two separate reading room circulation reports:
  1) Manually compiled by reading room staff
  2) Circulation information from library catalog
- Compare circulation statistics with Google Analytics for correlation of use in both areas
- Identify candidates for digitization based on high use and gauge patterns of patron interest

RESULTS

- Graph "Distribution of Unique Page Views" shows collections divided into quintiles (20% each) based on use, Quintile 1 being highest use.
  - Nearly 85 percent of online traffic comes from the first 20 percent of the collections.
- Additional findings from correlating page views with circulation statistics **(**see online document via URL or QR code provided for further details**):
  - Online use is not a good predictor of in-person use:
    - Quintile 5 (Q5) accounted for only 0.54 percent of the UPVs but nearly a quarter of the total material circulations.
    - Q1 collections had on average circulated twice during the time period measured, while only 35 percent of the collections had been circulated.
    - Q5 had circulated on average once, with nearly 49 percent of its collections accessed in the reading room.
- Determined need to use two metrics to provide recommendations for digitization
- Scatter graph "Unique Page Views/Circulation Comparison" shows collections plotted in one of four quadrants, divided at 100 unique Web views and 10 circulations to determine high use.
- Most digitized content (indicated in red) fell into Quadrant IV (collections with high Web visibility and low circulation) or in Quadrant III (low Web visibility and low circulation).





CONCLUSIONS

- Using page views with circulation stats provides a more accurate sense of the usefulness of the collections.
- However, page views does not predict in-person use.
- High online use may be more due to the level of description of the finding aid (item-level equals more page views) than interest in the content.
- Quadrant 1 collections identified as being of greater interest both online and in-house, thus likely the best candidates for digitization.

QUESTIONS FOR FURTHER RESEARCH

- Are collections in Quadrant II poorly described, since they are attracting in-house use but not Web access?
- Why are our patrons visiting the finding aids for materials that are not accessed in person, as charted in Quadrant IV?
- Does digitization decrease the in-person use in the reading room?
- What needs to be done to improve access to the materials in Quadrant III, which currently are not being accessed either remotely or in the reading room?

For more information, see http://net.lib.byu.edu/inventories/researchforum2014.pdf