



A WINDOW INTO USE AND USERS OF ARCHIVAL CONTENT IN AN INSTITUTIONAL DIGITAL REPOSITORY

BETH KAPLAN - HEAD, UNIVERSITY ARCHIVES & CO-DIRECTOR, UNIVERSITY DIGITAL CONSERVANCY | JASON ROY - HEAD, DIGITAL COLLECTIONS & INTERIM CO-DIRECTOR, DIGITAL LIBRARY DEVELOPMENT LAB

BACKGROUND

Colleges and universities are increasingly implementing digital institutional repositories (IRs) to gather and share the intellectual output of the campus community. Typically IRs contain scholarly works of faculty and staff (1). More recently, archivists have begun to explore the potential of the IR as a tool for capture, long-term storage, and enhanced access to born-digital or scanned core administrative content that has typically come to the University Archives in print form (2). In August 2007 the University of Minnesota Libraries launched the University Digital Conservancy, a campus-wide IR to serve as the "digital arm" of the University Archives, as well as the more traditional IR purpose as a venue for faculty works (3).

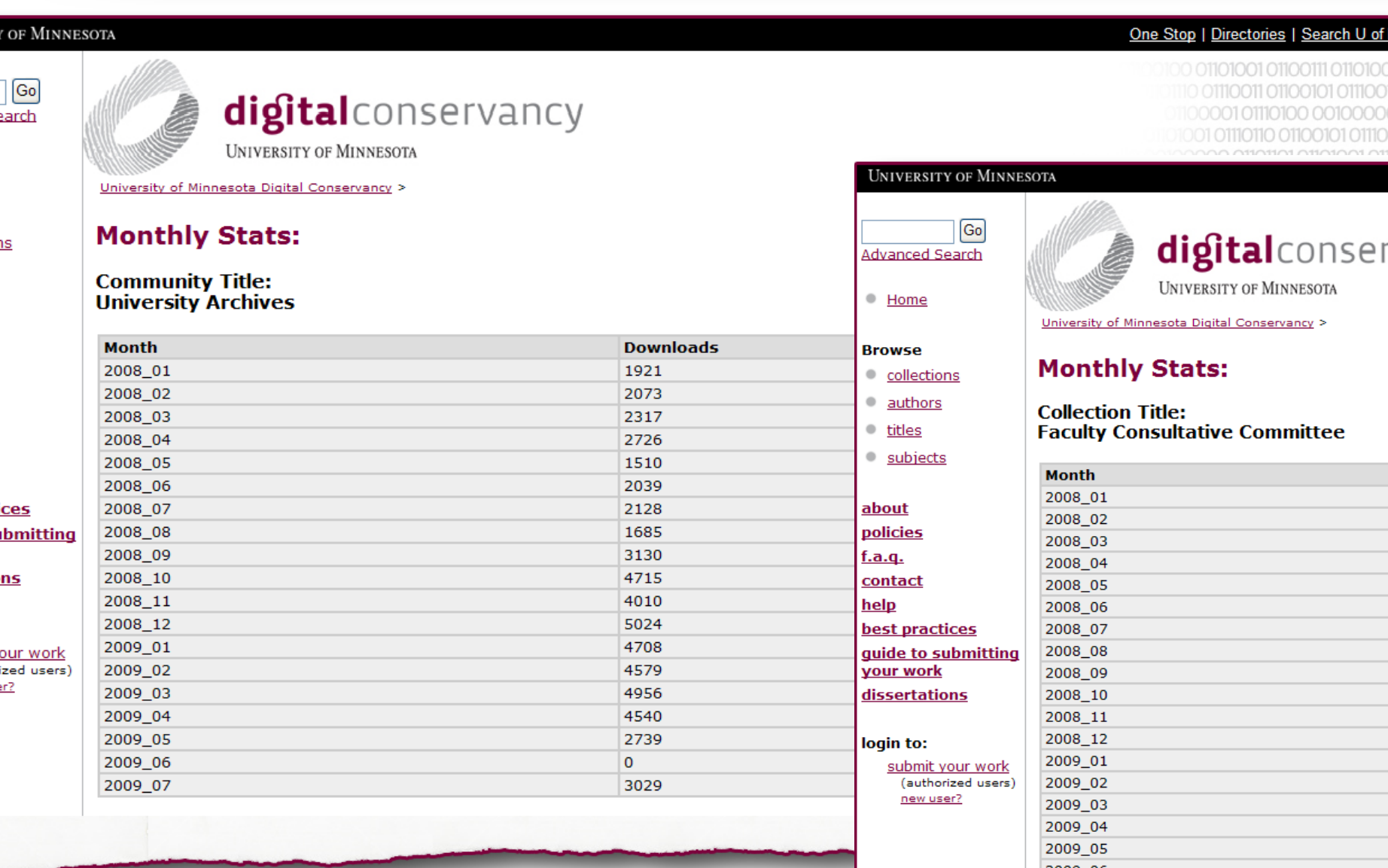
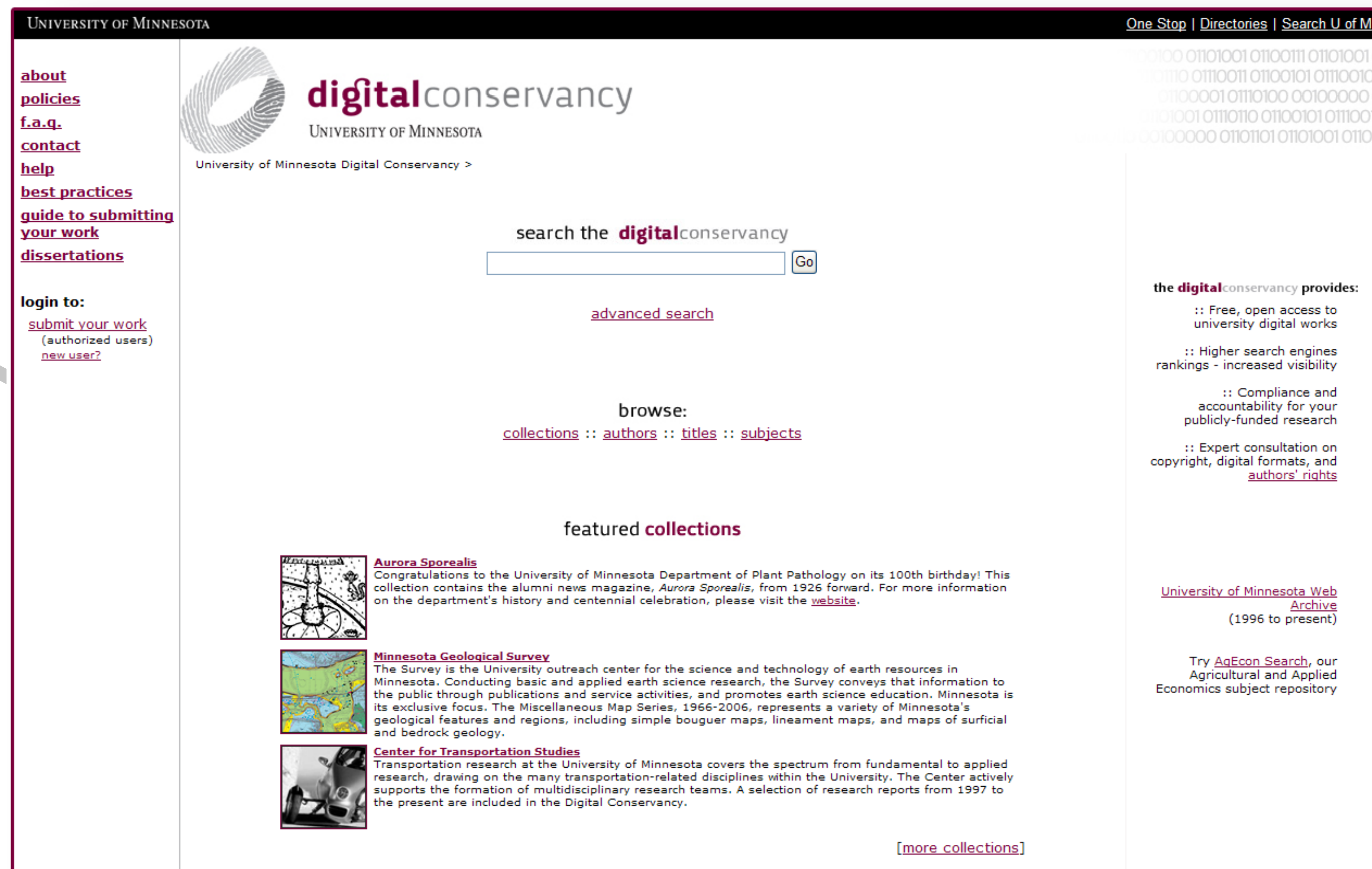
ASSUMPTIONS

The available data on users of typical IR content is scant at best, and detailed published information on users of archival content in IRs is non-existent (4). During the planning phase, we assumed that users of the administrative content in the Digital Conservancy would behave in much the same ways that users of the traditional University Archives behave: members of the campus community would continue to call or email the Archives with requests for information, the archivists staff would serve as intermediaries between those users and the collection—either print or digital—and access would continue to be as mediated and filtered as ever. In a way, this expectation was borne out: once the Digital Conservancy was up and running, archives reference statistics indicated no appreciable decrease in the number of traditional reference requests from the campus community.

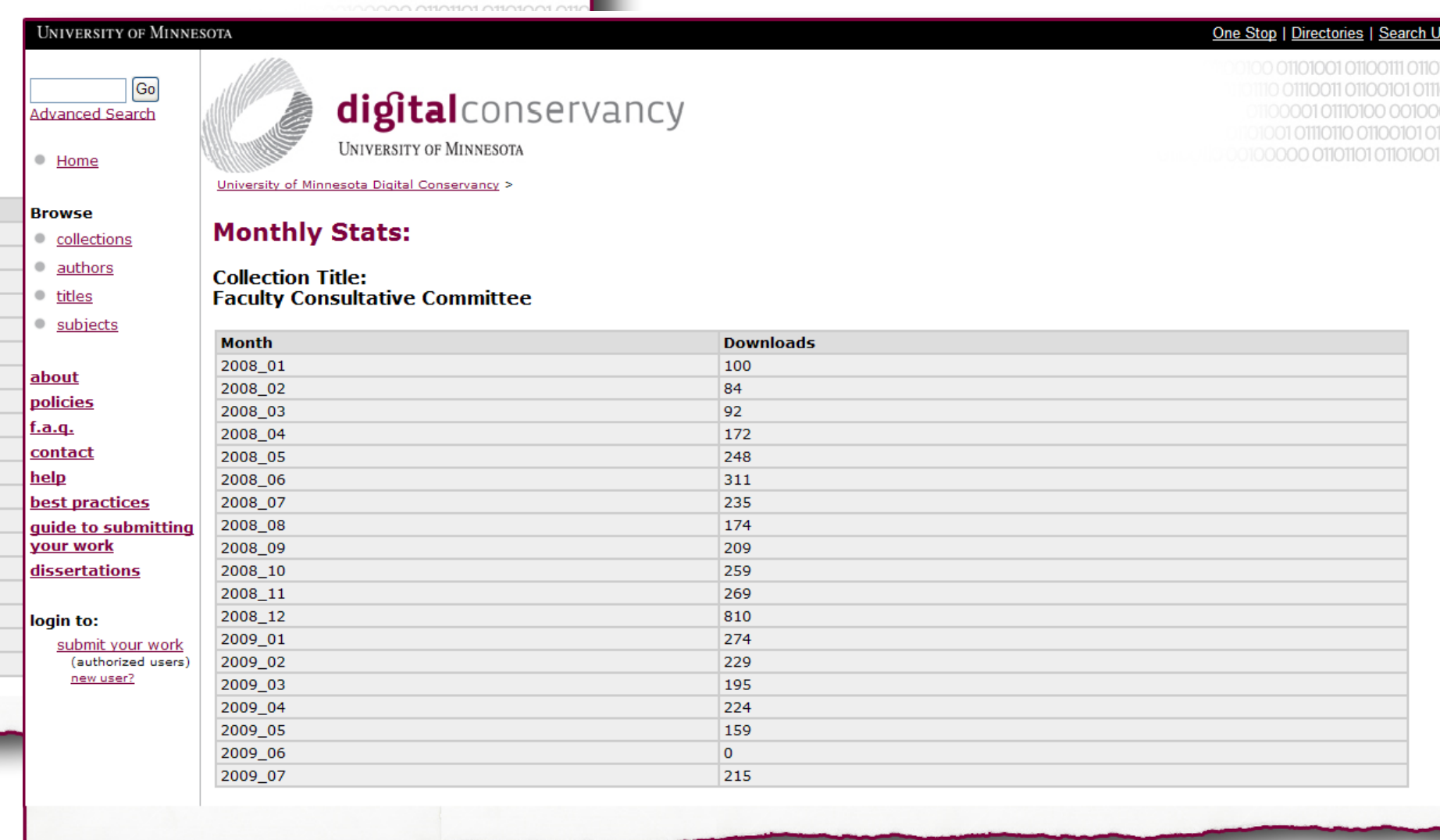
READING THE DATA

The implementation of a statistics package in spring 2009 revealed a whole new—and startlingly large—category of users: campus stakeholders who appeared to be going directly to the Digital Conservancy to find information on their own, bypassing the archives' traditional gate-keeper function entirely.

The statistics package that we incorporated into our DSpace installation was intended primarily as a content-recruitment tool for faculty works. That it revealed significant trends in the use of the archival content was serendipitous. We employed Apache Tomcat, a Java-based utility that captures web server log files, and AWStats, an open source Web analytics reporting tool that parses and analyzes log files and produces HTML reports (5). Download numbers sometimes are misleading since they can be easily manipulated even by those with little programming experience. To compensate for numbers inflated by robots, gaming and other quirks, we discounted unusually high numbers of downloads of individual files by particular IP addresses, noting that these were not reflective of normal patterns of use behavior. As the monthly statistics accumulate we began to see fairly consistent levels of downloads of archival content in the IR that could not possibly be generated by our own staff.



Month-by-month breakdown of numbers of downloads in the University Archives sub-community.



Monthly breakdown of downloads in one collection within University Archives, the minutes of the Faculty Consultative Committee.

Detail of the AWStats web statistics data for the entire Digital Conservancy, including archival and faculty works, April 2009.

Links from an external page (other web sites except search engines)	Pages	Percent	Hits	Percent
http://special.lib.umn.edu/uarchv/	152	11.3 %	152	8.8 %
http://www.lib.umn.edu	149	11.1 %	149	8.6 %
http://www.lib.umn.edu/inf	46	3.4 %	46	2.6 %
http://www1.umn.edu/usenate/committees/acfp.html	37	2.7 %	37	2.1 %
http://www1.umn.edu/usenate/committees/fcd.html	32	2.3 %	32	1.8 %
http://blog.lib.umn.edu/moore144/archives/	32	2.3 %	32	1.8 %
http://www.pdf-searcher.com/pdf-abstract/200agebra.html	26	1.9 %	26	1.5 %
http://en.wikipedia.org/wiki/Quadrilateral	21	1.5 %	21	1.2 %
http://www.vermontfdk.org/bc348.htm	17	1.2 %	17	0.9 %
http://ipil.com/search/	16	1.1 %	16	0.9 %
http://consensus.usg.edu/viewdoc/summary	15	1.1 %	15	0.8 %
http://translate.google.com/translate	14	1 %	14	0.8 %
http://www1.umn.edu/usenate/committees/acfd.html	14	1 %	14	0.8 %
http://special.lib.umn.edu	14	1 %	14	0.8 %
http://translate.google.co.jp/translate_s	13	0.9 %	13	0.7 %
http://www1.umn.edu/usenate/committees/acfp.html	13	0.9 %	13	0.7 %
http://74.125.47.132/search	13	0.9 %	13	0.7 %
http://www1.umn.edu/usenate/committees/acfd.html	12	0.8 %	12	0.7 %
http://74.125.47.132/search	11	0.8 %	11	0.6 %
http://lib.umn.edu	11	0.8 %	11	0.6 %
http://www.pdf-search-engine.com/quir-story-grp-pdf.html	11	0.8 %	11	0.6 %

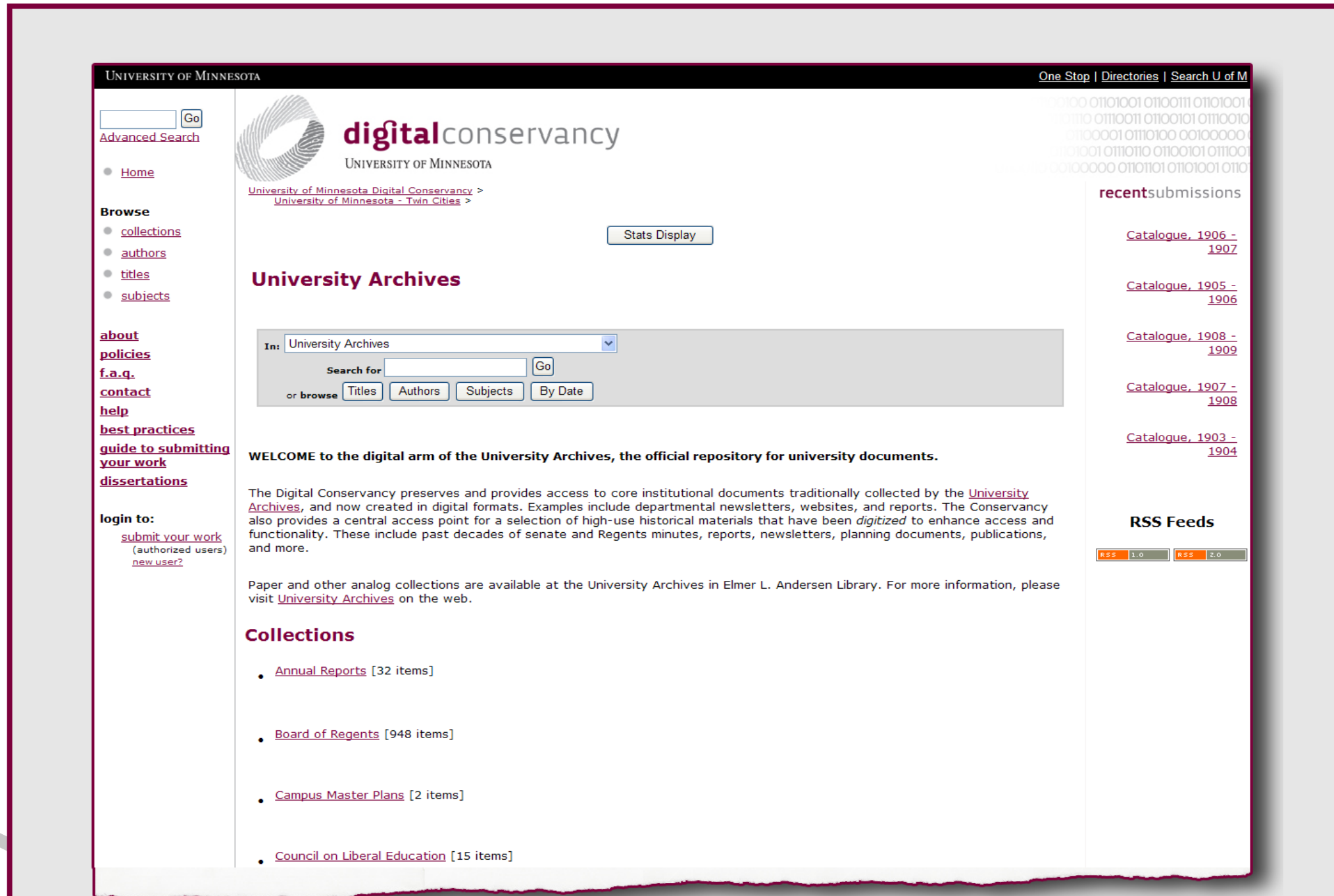
Top referring sites for April 2009. Six of the top 10 referring sites are internal to the University of Minnesota (umn.edu) indicating a high proportion of campus use.

FURTHER EXPLORATION

We recognize that caution is imperative when interpreting statistics and we share these findings as imprecise, but certainly suggestive. As the data roll in each month, the numbers and the analysis seem to suggest new and unexpected uses and user behaviors. Monthly download numbers, identification of particularly highly used documents, and information about referring sites, along with an accumulating body of anecdotal information, demonstrate a new area of potential for institutional digital repositories as important components of institutional archives. Managing and populating institutional repositories is a resource-intensive activity. But if they do provide a place to park digital content, as we had expected, and if users are beginning to link directly to them for their own information needs, the investment seems worth pursuing.

REFERENCES

1. The MIRACLE (Making Institutional Repositories A Collaborative Learning Environment) Project.
2. "Institutional Repositories and the Institution's Repository: What is the Role of University Archives with an Institution's On-line Digital Repository?" Douglas Bicknese, Archival Issues: Journal of the Midwest Archives Conference volume 28 (2), 2003-2004.
3. University Digital Conservancy <http://conservancy.umn.edu>.
4. "Institutional Repositories and their 'Other' Users: Usability Beyond Authors." Dana McKay. <http://www.ariadne.ac.uk/issue52/mckay>.
5. AWStats <http://en.wikipedia.org/wiki/Awstats> and Apache Tomcat <http://en.wikipedia.org/wiki/Apache_Tomcat>.



FACTS & FIGURES

- As of August 7, 2009, 10,083 items available in the Digital Conservancy.
- Archival content currently accounts for 5,697 (56.5%) of the total items available.

UPCOMING CONTENT

- 3,000 items from the University Relations office. (September 2009)
- 1,000 additional documents from the University Senate. (September 2009)