

Electronic Theses and Dissertations: Differences in Behavior for Local and Non-Local Users

Midge Coates, Auburn University Libraries
coatemi@auburn.edu

Library Hi Tech, accepted (November 2013), preprint available via Emerald EarlyCite (June 2014)

Abstract:

Purpose:

This study examines two research questions: (1) How do users in different locations find Auburn University Electronic Theses and Dissertations? (2) Do users in different locations interact differently with the collection and, if so, how?

Design/methodology/approach:

Google Analytics data for user visits, landing pages, and page views were separated into groups based on user location. Visits data were also correlated with source (referring Web site), and landing pages and page views were grouped by type.

Findings:

Most local users came to the repository via Auburn University Web pages. This group usually landed on the collection home page and used internal navigation pages to find what they needed. Submission page views showed that most ETD depositors were local.

Most out-of-state users came to the repository via Web search engines. This group usually landed directly on bibliographic information pages for individual ETDs. They used internal navigation pages less frequently than local users. Users located within the state but outside of the local area interacted with the collection in a way that was intermediate between these two groups.

Practical implications:

Institutions interested in improving repository access for depositors will probably find it helpful to focus on in-state usage reports, while institutions seeking to improve access for end-users should exclude in-state users from their assessments.

Originality/value:

This is the first detailed examination of ETDs usage published since 2001 and shows how filtering tools available in Google Analytics allow comparisons of user behavior based on location and source (referring Web site).

Electronic Theses and Dissertations: Differences in Behavior for Local and Non-Local Users

Introduction

Many institutions offering graduate degrees are now requiring electronic theses and dissertations (ETDs) from their successful degree candidates, replacing the paper versions required in the past. Many of these documents are deposited into institutional, consortial, or national repositories, from which they are available without a fee (sometimes after an embargo period). Open access ETDs may be indexed by their local institutions, by consortial, state, and national indexes, and/or by international collective indexes such as Networked Digital Library of Theses and Dissertations (NDLTD) (NDTLD, 2012) and Open Access Theses and Dissertations (OATD) (OATD, 2013).

Usage patterns for digital collections such as ETDs are important because they allow managers to allocate resources more effectively towards making these collections easier to find and use. This study examines two research questions: (1) How do users in different locations find the Auburn University ETDs (AUETDs) collection? (2) Do users in different locations interact differently with the collection and, if so, how?

Literature Review

Despite the wide-spread adoption of institutional and ETD repositories, there is a dearth of research on usage patterns for them. McKay's article on institutional repository users (McKay, 2007) pointed out that this scarcity of end-user research resulted in a lack of knowledge about:

whether [repository] users are local or from outside the hosting institution;

whether they find the [repository] via the institutional homepage or via search engine referrals; ... what kind of information they look for and use; [and] how they use the functionality offered by the [repository].

Schmitz noted the lack of available information about repository end-users in a report from the Council on Library and Information Resources (Schmitz, 2008). Bailey's bibliography of resources on institutional and ETD repositories further documented this lack (Bailey, 2011). Both McKay and Schmitz suggested that it would be necessary to include studies of the use of other digital resources such as online encyclopedias, scholarly databases, and collections of digitized books if one wanted to get a fuller picture of repository end-users.

Tenopir and Rowlands showed that Web search engines such as Google were used by researchers of all ages to find books (Tenopir and Rowlands, 2007). They found that students looking for articles preferred searching, while faculty preferred browsing and reference-chaining (sometimes called "footnote-chasing"). Ismail and Kareem found that student researchers at both novice and advanced levels of expertise went first to Web search engines to find research information or answer questions (Ismail and Kareem, 2011). Connaway and Dickey observed that researchers at all levels spent little time actually reading the digital items they found online, focusing instead on finding documents first, then downloading them (or printing them out) for later reading off-line (Connaway and Dickey, 2010).

Institutional Repositories

Armbruster and Romary described four types of open access repositories for scholarly content: subject-based repositories, research repositories, national repository systems, and institutional repositories (Armbruster and Romary, 2010). They pointed out differences in who

administered these repositories, what types of materials they contained, and whether deposit was voluntary or mandatory. The institutional repositories they examined usually contained scholarly materials produced at a college or university, without restrictions based on the area of research or funding source. These collections were usually administered by the institution's library and contained article preprints and post-prints and conference presentations. Sometimes they also contained ETDs, datasets, teaching materials, and/or student papers.

Mukherjee and Nazim made a quantitative study of institutional repositories around the world (Mukherjee and Nazim, 2011). They found that most of the content in these repositories consisted of material from published literature—articles (36.0%) and books, including chapters (11.0%)—and unpublished or “grey” literature—conference and workshop presentations (19%), unpublished reports and working papers (17.0%), and ETDs (1.7%). Datasets accounted for 2.2% of institutional repository content, and learning objects, 2.8%.

Studies of Institutional Repository Users

Based on the behavior of end-users of journal databases and research repositories, McKay predicted that institutional repository end-users would (a) make short visits, viewing only a few items per visit; (b) navigate the repository by browsing and performing simple searches; (c) use Web search engines in preference to library tools to locate materials (McKay, 2007). These predictions were supported by St. Jean and co-workers, who interviewed a group of institutional repository end-users (St. Jean, *et al.*, 2011). They found that users lacked general awareness of institutional repositories and usually went to Web search engines and to library databases when looking for information materials. When seeking a known item in the repository, they either went directly to the item (via a Web search engine or a bookmark) or they used the

repository's search function (searching first by author, then by title). When seeking unknown items, they preferred to use the repository's browse function.

Organ analyzed usage statistics for an institutional repository at the University of Wollongong, Australia (Organ, 2006). He found that 80.9% of the cover page downloads were from Google users and 95.8% of the full-text downloads for which the source was known (51.1% of the total) were also from Google users. For this collection, 75.5% of cover page downloads and 69% of full text downloads were from domains in the United States, while Australian users accounted for 6.2% of cover page downloads and 10.1% of full-text downloads.

ETD Repositories

Given the fact that some institutional repositories include ETDs as a significant portion of their content (Ware, 2004; Lynch and Lippincott, 2005), one might expect to see similarities in the usage patterns for institutional repository materials and ETDs. Yiotis postulated that ETD repositories constituted a subset of institutional repositories (Yiotis, 2008). She also provided a historical overview of the development of ETDs since their introduction in the 1990s.

Studies of ETDs Users

Prior to the adoption of ETDs, users wishing to consult a thesis or dissertation had to obtain a paper or microfilmed copy, either from the originating institution (in person or via inter-library loan) or from University of Michigan's Digital Dissertations (now ProQuest Dissertations & Theses). Lee-Smeltzer and Hackleman found that local graduate students were the major users of a collection of paper theses and dissertations located at Oregon State University (OSU) (Lee-Smeltzer and Hackleman, 1995). The primary reasons given for use were

general research, research towards students' own theses and dissertations, and templates for style or format. Students using the OSU collection for research found these documents while searching the library catalog for information on a topic and were not searching specifically for theses and dissertations. These same reasons for use should apply to ETDs, although online availability expands their potential audience. Online availability also makes it more likely that these documents will be found during a general search for information.

In interviews with researchers, Ismail and Kareem found that students preferred ETDs over other types of institutional repository content due to the breadth and depth of the research found in them (Ismail and Kareem, 2011). Students also considered these documents more trustworthy than other types of repository materials (preprints, conference presentations, etc.). Graduate students in the study also used ETDs as format guides for their own theses and dissertations.

Zhang, Lee, and You analyzed transaction logs to learn more about the users of an ETDs collection at the Korea Institute of Science and Technology Information (KISTI) (Zhang, *et al.*, 2001). This national collection was multi-lingual, but the user interface was only available in Korean. Not surprisingly, most users (80%) were located within South Korea, but the remaining users came from a wide variety of countries. The search function was the most frequently used system function for this collection. Most users made short visits (60% lasting a minute or less), and the duration of the visits was proportional to the number of individual pages viewed.

Alemneh and Phillips used Google Analytics to study the usage patterns for the University of North Texas (UNT) ETDs collection (Alemneh and Phillips, 2011). During the study period (2009-09-01 to 2011-03-22), local users from Denton, Texas, viewed 9.51 pages per visit (3.6% of total visits), as compared to 6.00 pages per visit (11.6% of total visits) for all

Texas users, 4.05 pages per visit (68.9% of total visits) for all USA users, and 3.99 pages per visit for all users world-wide. Over 62% of visitors were directed to the collection by Web search engines. Wikipedia referred 7.7% of collection visitors, and the social networking sites Facebook, Twitter, and Reddit, 1.9%. Referring sources with “unt” in their URLs accounted for 3.5% of user visits (Alemneh and Phillips, 2011).

Methodology

In 2005, the Graduate School of Auburn University (GSAU) began requiring ETDs instead of physical copies. To promote scholarly communication, these are accessible to the public, not just the University community. The AUETDs collection was created by Auburn University Libraries (AUL) as a DSpace repository (<http://etd.auburn.edu/etd/>) cooperatively administered by AUL and the GSAU. The collection may be browsed by date, author name, advisor name, title, and department. Brief metadata pages list title, author name, abstract, and date and provide a link to the full-text PDF. Full metadata pages also include advisor name(s), department name, and embargo details, if applicable.

Students upload their own ETDs in PDF format (after committee approval), and these are approved (or rejected) by GSAU staff. When students upload their ETDs to the collection, they may request an embargo period during which only the metadata will be viewable, but this period is limited. If no embargo period has been requested, the ETDs are available to the public immediately after GSAU approval. About 400-500 new ETDs are added to the collection each year. As of August 31, 2013, the collection contained 3,467 theses and dissertations in PDF format.

The AUETDs collection is indexed by Web search engines such as Google, Yahoo!, etc.,

and by Google Scholar. It is also indexed by databases such as WorldCat and by the international ETD directories NDLTD and OATD. There are also many local avenues of access to the repository. The AUL Web site links to the collection from a pull-down menu on the home page, an Open Access Collections page, and an Articles and Databases page. Records for individual ETDs are in AUL's Online Public Access Catalog (OPAC) and its VuFind discovery layer catalog. GSAU links to the repository from its own Web site, as do some individual colleges and departments. Some of the latter also link to ETDs uploaded by their own students. Finally, the Auburn University Web site has a search function that can be used to locate the repository and/or individual ETDs.

Since March 2009, AUL has been collecting user statistics for the AUETDs collection using the Standard (free) version of the Google Analytics Web tracking service (Google, 2013; Wikipedia, 2013). JavaScript tracking code has been added to the AUETDs Web page template. When a browser accesses any collection page, Google records the URL for that page, the access time and date, referring Web site, and user location, provided that the browser has enabled JavaScript, caching, and cookies. (If any of these are not enabled, data cannot be collected.)

This study examined Google Analytics data for AUETDs page views and user visits for a one-year period (2012-01-13 through 2013-01-12). The data were filtered according to user location (City, Region, Country). In some cases, the data were also sorted by page type (based on URL) and/or by source (the Web sites which referred users to AUETDs).

Results and Discussion

For this study, users have been divided into four groups based on location information provided by Google Analytics. The Local group consisted of users whose Region location was

Alabama and whose City location was Auburn or Opelika (a city adjacent to Auburn). The Alabama–not Local group consisted of users whose Region location was Alabama but omitted Auburn and Opelika users. The USA–not Alabama group consisted of users whose Country location was given as United States but omitted Alabama users, while the World–not USA group consisted of all users except those whose Country location was United States.

It was expected that groups in different locations would consist of different types of users and would therefore interact differently with the collection. For example, many in the Local group were likely to be AU students in the process of submitting their theses and dissertations, while the remaining members of this group might consist of faculty and students doing research and AUL and GSAU staff members performing administrative tasks. The other three groups might contain some students submitting ETDs, but these groups were expected to be predominantly non-AU researchers looking for information.

It was also expected that different types of users might find their way to the collection via different types of sources (referring links). In the tables below, similar types of sources have been grouped together. Direct indicates users that came to AUETDs either via a browser bookmark or by entering a URL into their browser address bars (by typing or by copy-and-pasting). The University Sources group combines users who came to the site via AUL sources (library home page, catalogs, digital library page, etc.) with users from the GSAU Web site, Web pages for individual AU colleges and departments, and the AU Web site search function. The Scholarly Databases group combines users from Google Scholar, the NDLTD and OATD databases, the OhioLink ETD directory, WorldCat/FirstSearch, ScienceDirect, Summon, and similar sources.

The Search Engines group combines users of the Web search engines most recognizable

to US audiences (e.g., Google, Yahoo!, Bing) with users of lesser-known search engines (e.g., Duckduckgo, PDFQueen) and non-English-based search engines (e.g., Babylon, Sogou, Yandex). The Email and Social Media group combines users of any recognizable email service (i.e., with “mail” as part of the URL). This group may be under-represented: if a user copied-and-pasted a link from an email, rather than clicking on it, the visit would be designated as Direct rather than having an email source. This group also includes users of the social media sites Facebook, Reddit, StackOverflow, LinkedIn, and Wikipedia. The Other group contains all users who came to the site via sources not fitting into any of these categories.

Most of the tables below present Google Analytics usage data that have been sorted by page type in order to see how user groups interact with the AUETDs collection. In these tables, browse, search, and advanced search pages have been combined as Navigation Pages. Simple (abbreviated) and full (complete) metadata pages have been combined as Bibliographic Information Pages. Views of PDFs in the browser window have been listed in the tables as PDF Views. (The counting of PDF right-click downloads was not enabled during the study period.) Views for all pages associated with submitting an ETD to the collection have been combined as Submission Pages. Data for all other page types (including repository administration) have been combined as Other Pages.

Usage Data

Overall user visit data for the four location groups are given in Table 1. Most of the visits to AUETDs during the study period were from the two out-of-state groups—USA—not Alabama (38.9%) and World—not USA (43.2%)—and these two groups viewed just over 2 pages per visit. The number of visits from users outside of the USA was substantially higher than the number of

Table 1. AUETDs user visits and page views, filtered by user location (2012-01-13 through 2013-01-12).

	Local Users	Alabama-not Local Users	USA-not Alabama Users	World-not USA Users	Collection Total
Visits (% total visits)	10,998 (14.2%)	2,824 (3.6%)	30,178 (38.9%)	33,502 (43.2%)	77,502 (100%)
Pages viewed (% total views)	99,321 (39.4%)	15,719 (6.2%)	65,673 (26.1%)	71,089 (28.2%)	251,802 (100%)
Avg. pages/visit	9.03	5.57	2.18	2.12	

non-domestic users (20%) reported for the KISTI collection (Zhang, *et al.*, 2001). This may be due to the fact that the AUETD user interface is in English, a more common second language for international researchers than Korean.

Table 1 shows that the two in-state groups made fewer visits than the out-of state groups—Local (14.2%) and Alabama—not Local (3.6%)—but viewed many more pages per visit—9.03 pages per visit for Local users and 5.57 pages per visit for Alabama—not Local users. The differences in the number of pages viewed per visit show that groups in different locations interacted differently with the collection.

Table 2 contains data for all pages viewed by users in the four location groups, while Table 3 contains data on the landing pages for these groups. Landing pages are the pages at which visitors arrive when they first enter a collection. Data in both tables have been sorted by page type. Examining both sets of data gives an indication of how users interacted with the collection and points out some differences between the location groups.

Table 2 shows that 43.1% of the pages viewed by users in the Local group were Navigation Pages. Landing page data in Table 3 show that 50.7% of this group landed on the Home Page, while only 25.7% landed on Bibliographic Information Pages. These findings are consistent with Table 1 data which show that Local users viewed 9.03 pages per visit, the most for any location group. Visitors who land on a collection's home page must use internal navigation tools (e.g., browse, search, advanced search pages) to find the documents they need, which results in additional pages viewed per visit.

In contrast with the Local group, most of the users in the two out-of-state groups landed on Bibliographic Information Pages, 82.4% of the USA—not Alabama group and 88.4% of the World—not USA group. Given that Table 1 shows that these users viewed just over 2 pages per

Table 2. Page views, sorted by page type and filtered by user location (2012-01-13 through 2013-01-12).

	Home Page	Navigation Pages	Bibliographic Information Pages	PDF Views	Submission Pages	Other Pages	Location Totals
Local Users	visits (% page type)	42,833 (43.1%)	27,954 (28.1%)	4 (<0.1%)	11,622 (11.7%)	8,298 (8.4%)	99,321 (100%)
Alabama-not Local Users	visits (% page type)	8,391 (53.4%)	3,934 (25.0%)	1 (<0.1%)	1,346 (8.6%)	604 (3.8%)	15,719 (100%)
USA-not Alabama Users	visits (% page type)	18,764 (28.6%)	38,337 (58.4%)	48 (<0.1%)	2,767 (4.2%)	2,188 (3.3%)	65,673 (100%)
World-not USA Users	visits (% page type)	19,540 (27.5)	46,488 (65.4%)	96 (0.1%)	135 (0.2%)	2,537 (3.6%)	71,089 (100%)
Page Type Totals	visits (% total)	89,528 (35.6%)	116,713 (46.4%)	149 (<0.1%)	15,870 (6.3%)	13,627 (5.4%)	251,802 (100%)

Table 3. Landing pages, sorted by page type and filtered by user location (2012-01-13 through 2013-01-12).

		Home Page	Navigation Pages	Bibliographic Information Pages	PDF Views	Submission Pages	Other Pages	Location Totals
Local Users	visits (% page type)	5,575 (50.7%)	2,048 (18.6%)	2,825 (25.7%)	2 (<0.1%)	116 (1.1%)	432 (3.9%)	10,998 (100%)
Alabama-not Local Users	visits (% page type)	961 (34.0%)	610 (21.6%)	1,200 (42.5%)	1 (26.9%)	8 (0.2%)	44 (1.6%)	2,824 (100%)
USA-not Alabama Users	visits (% page type)	2,266 (7.5%)	2,735 (9.1%)	24,860 (82.4%)	32 (0.1%)	10 (<0.1%)	275 (0.9%)	30,178 (100%)
World-not USA Users	visits (% page type)	1,274 (3.8%)	1,983 (5.9%)	29,623 (88.4%)	85 (0.2%)	2 (<0.1%)	535 (1.6%)	33,502 (100%)
Page Type Totals	visits (% total)	10,076 (13.0%)	7,376 (9.5%)	58,508 (75.5%)	120 (0.2%)	136 (0.2%)	1,286 (1.6%)	77,502 (100%)

visit, it might be supposed that they were using external navigation tools (e.g., search engines, databases) to locate the documents they needed rather than the collection's own navigation tools. Table 2, however, shows that 28.6% of pages viewed by the USA-not Alabama group and 27.5% of the pages viewed by the World-not USA group were Navigation Pages. Apparently, these groups did use the collection's internal navigation tools, just not as frequently as users in the Local group. The behavior of the Alabama-not Local group appeared to be transitional between the Local and out-of-state groups.

To look for location-related differences in how users found the AUETDs collection, the user visit data in Table 4 has been sorted by source (referring Web link). These data show that 53.2% of users in the Local group came to the collection via University Sources, while 29.0% came via Search Engines. In sharp contrast, more than 70% of users in both out-of-state groups came via Search Engines and less than 10% from University Sources. For both of these source categories, Local users viewed significantly more pages per visit than did the out-of-state users. As observed previously, the behavior of the Alabama-not Local group was transitional between that of the Local group and the two out-of-state groups.

Direct Visitors

Direct visitors came to the collection via bookmarks or by manually entering a URL into the browser address box. The majority of Direct users were in the two out-of-state groups: USA-not Alabama (35.8%) and World-not USA (37.3%). In-state Direct visitors viewed 6-10 pages per visit, while those in the out-of-state groups viewed 2-5 pages per visit. If one assumes that many Direct users came to the collection via bookmarks, then the difference in pages viewed suggests that in-state and out-of-state visitors were using bookmarks differently—that in-state

Table 4. User visits, sorted by source and filtered by user location (2012-01-13 through 2013-01-02).

	Direct	University Sources	Scholarly Databases	Search Engines	Email and Social Media	Other Sources	Location Totals
Local Users	visits (% location)	1,843 (16.8%)	60 (0.5%)	3,186 (29.0%)	35 (0.3%)	21 (0.2%)	10,998 (100%)
	pages/visit	9.56	1.32	4.87	3.40	3.67	
Alabama-not Local Users	visits (% location)	299 (10.6%)	37 (1.3%)	1,402 (49.6%)	28 (1.0%)	33 (1.2%)	2,824 (100%)
	pages/visit	6.55	1.27	3.59	3.86	4.73	
USA-not Alabama Users	visits (% location)	2,850 (9.4%)	1,827 (6.1%)	22,423 (74.3%)	356 (1.2%)	589 (2.0%)	30,178 (100%)
	pages/visit	2.54	1.53	1.83	1.55	2.03	
World-not USA Users	visits (% location)	2,975 (8.9%)	2,804 (8.4%)	25,961 (77.5%)	217 (0.6%)	812 (2.4%)	33,502 (100%)
	pages/visit	4.33	1.72	1.78	4.40	2.66	
Source Totals	visits (% total)	7,967 (10.3%)	4,728 (6.1%)	52,972 (68.3%)	636 (0.8%)	1,455 (1.9%)	77,502 (100%)
	pages/visit	4.98	1.64	2.03	2.73	2.47	

Direct users were preferentially bookmarking the collection's home page, while the out-of-state Direct users were bookmarking pages for individual ETDs.

University Sources Visitors

University Sources visitors came to the collection via Web sites for AUL, GSAU, the University itself, and individual colleges and departments. The majority of University Sources users (59.8%) were in the Local group. University Sources visitors in the Local group viewed 11.26 pages per visit, while those in the other three groups viewed 5-9 pages per visit. The comparatively high number of pages viewed by all four groups suggests that these visitors used University Web pages to find the overall collection, then used the collection's internal navigation tools to locate individual ETDs.

Scholarly Databases Visitors

Scholarly Databases visitors came to the collection via Google Scholar, international ETD directories such as NDLTD, OATD, and OhioLink, and vendor databases such as WorldCat/FirstSearch. Table 4 shows that 97.8% of Scholarly Databases visitors were in the two out-of-state location groups. This group viewed 1-2 pages per visit, no matter what their location. It is likely that these sources were used to find individual ETDs, not the overall collection.

Search Engines Visitors

Search Engines visitors came to the collection via Web search engines such as Google, specialized search engines such as PDFQueen, and non-English-based search engines such as Babylon and Yandex. Table 4 shows that 91.3% of Search Engines visitors were in the two

out-of-state location groups. Out-of-state Search Engines users viewed 1-2 pages per visit, while their in-state counterparts viewed 3-5 pages per visit. This suggests that some in-state visitors used these sources to find the AUETDs collection itself, while most out-of-state visitors used them to find individual ETDs.

Email and Social Media Visitors

Email and Social Media visitors came to the collection by clicking on a URL contained in an email from a Web-based email service (e.g., Gmail, Yahoo! Mail, Outlook) or by clicking on a link from one of the common social media sites Facebook, Reddit, StackOverflow, LinkedIn, and Wikipedia. Email and Social Media sources accounted for only 0.8% of all visits to AUETDs. Of these, 90.2% were in the two out-of-state location groups. However, the overall numbers for this group were so low that it is difficult to draw conclusions about user behavior based on them.

Location Groups' Interaction with AUETDs

Local Group

This group consisted of users living or working in Auburn, Ala., and Opelika, Ala. It may be assumed that most, if not all, of this group were AU students, faculty, or staff. How users in the Local group will interact with the collection will depend on who they are. Graduate students at the end of their programs must upload their approved ETDs to the repository. Other students and faculty may seek out ETDs for research purposes or to check specific format requirements (Lee-Smeltzer and Hackleman, 1995; Ismail and Kareem, 2011). AUL and GSAU staff may perform administrative and cataloging tasks (McKay, 2007).

Table 4 shows that visitors in the Local group used three types of sources to find the collection: Direct, 16.8%, University Sources, 53.2%; and Search Engines, 29.0%. Very few used Scholarly Databases or Email and Social Media sources. As this group should have direct knowledge of the AUETDs collection, they might have been expected to use University Web pages almost exclusively. Although they did use these sources more than any other location group, they still made significant use of Web search engines, as predicted by earlier studies (Tenopir and Rowlands, 2007; McKay, 2007; St. Jean, *et al.*, 2011).

About half (50.7%) of users in the Local group landed on the home page when they entered the collection (Table 3). Once they were in the collection, Local users found the information they needed using internal navigation tools (browse, search, advanced search pages): 43.1% of the pages viewed by users in this group were Navigation Pages, more than any other page type (Table 2). This fact and the fact that some of these users were viewing Submission Pages (presumably to upload their own ETDs) were probably the reasons why users in the Local group viewed 9.03 pages per visit (Table 1).

Alabama—not Local Group

Conceptually, this group can be sub-divided into PseudoLocal and NonLocal categories. The PseudoLocal category consists of AU students, faculty, and staff living outside the city limits and accessing the collection from their residences and AU researchers working at remote locations around the state. PseudoLocal users might be expected to interact with the collection in a way similar to that of the Local group. The NonLocal category consists of researchers not affiliated with AU. NonLocal users would not be expected to interact with the collection similarly to the Local group but to have more in common with the two out-of-state groups.

One might assume that any members of the Alabama–not Local group viewing Submission Pages would fall into the PseudoLocal category, as only AU students would need to upload ETDs. This suggests a way to obtain a rough estimate of the percentage of PseudoLocal users present in this group and in the two out-of-state groups. Table 5 contains page views data for Submission Pages for all four location groups. Based on the ratios of Submission Page views to total page views, one can obtain an estimate of 73% PseudoLocal users in the Alabama–not Local group.

This group exhibited behavior intermediate between the Local group and the USA–not Alabama group. Table 4 shows that users in the Alabama–not Local group used the same types of sources to find the collection as the Local group but in different proportions: Direct, 10.6%; University Sources, 36.3%; and Search Engines, 49.6%. Like the Local group, few found the collection via Scholarly Databases or Email and Social Media sources.

Table 3 shows that only 34.0% of this group landed on the Home Page, while 42.5% landed on Bibliographic Information Pages. Interestingly, Table 2 shows that Alabama–not Local users viewed Navigation Pages to an even higher degree (53.4%) than the Local group (43.1%). This, coupled with the smaller number of Submission Pages viewed by this group (8.6%), may mean that users in this group were more often interested in finding information than in uploading ETDs, even when using University Web pages to find the collection.

USA–not Alabama Group

Table 5 gives an estimate of 36% PseudoLocal users in the USA–not Alabama group. The remainder of this group may be assumed to be NonLocal researchers. Table 4 shows that Search Engines users constituted the majority of this group, 74.3%. The rest used: Direct, 9.4%;

Table 5. Submission Page Views Used to Estimate the Percentage of PseudoLocals in User Location Groups (2012-01-13 through 2013-01-12).

	Submission Page Views	Total Page Views	Submission Page Views as % of Total	Estimate of % PseudoLocals in Location Group
Local Users	11,622	99,321	11.70%	[100% Local]
Alabama-not Local Users	1,346	15,719	8.56%	73% PseudoLocals
USA-not Alabama Users	2,767	65,673	4.21%	36% PseudoLocals
World-not USA Users	135	71,089	0.19%	2% PseudoLocals

University Sources, 7.1%; and Scholarly Databases, 6.1%. Few used Email and Social Media sources.

Given the high proportion of Search Engines users, it is not surprising that 82.4% of this location group landed on Bibliographic Information Pages (Table 3). Table 2 shows that Navigation Pages accounted for only 28.6% of the pages viewed by this group. These last two observations suggest that this group frequently used external tools (Web search engines and databases) to go directly to the pages for individual ETDs (as opposed to going first to the collection home page and then using internal navigation tools). This is supported by the low number of pages per visit (2.18) for this group shown in Table 1. These data are consistent with the predictions of McKay that institutional repository end-users would make short visits, viewing only a few items, and would use Web search engines to locate repository materials (McKay, 2007). They are also consistent with the findings of St. Jean, *et al.*, that institutional repository end-users used Web search engines to go directly to items of interest (St. Jean, *et al.*, 2011).

World-not USA Group

Table 5 gives an estimate of only 2% PseudoLocal users in the World-not USA group, with the rest assumed to be NonLocal researchers. Search Engines users constituted the majority of this group, 77.5% (Table 4). The remainder used: Direct, 8.9%; University Sources, 2.2%; and Scholarly Databases, 8.4%. Again, Email and Social Media sources were seldom used.

Table 3 shows that 88.4% of this group landed on Bibliographic Information Pages. This group viewed Navigation Pages (browse, search, and advanced search) in similar amounts to the USA-not Alabama group, 27.5% of total pages viewed (Table 2). These last two observations suggest that this group, like the USA-not Alabama group, frequently used external tools (Web

search engines and databases) to go directly to pages for specific ETDs. This is supported by the low number of pages per visit (2.12) shown for this group in Table 1. As shown for the USA–not Alabama group, these data are consistent with the observations of McKay and St. Jean, *et al.*, about the behavior of institutional repository end-users (McKay, 2007; St. Jean, *et al.*, 2011).

Comparing AUETDs with Other Repositories

It is difficult to make direct comparisons of AUETDs to other ETD repositories because of the scarcity of data reported in the literature. Annual use data have been posted for the Virginia Tech ETDs repository for 1997/98 through 2012 (Virginia Tech, 2013). These data have presumably been obtained from server transaction logs, as most pre-date Google Analytics. Although overall collection use numbers are given, no locations or referring sources have been provided, and the only information about page type is limited to “html files ... mostly tables of contents and abstracts” and “PDF files ... mostly full ETDs”.

Use data for the KISTI ETDs collection also pre-date Google Analytics and were obtained from server transaction logs (Zhang, *et al.*, 2001). As this collection was a national repository, no single group of users could be considered “Local”, users being identified only as from South Korea or “Other Countries”. No referring source data have been provided. Overall page views cannot be compared to those for AUETDs, as the KISTI collection was structured such that page views were recorded separately for individual document pages, rather than for the full ETDs.

Data for the UNT ETDs repository were obtained from Google Analytics rather than transaction logs (Alemneh and Phillips, 2011). However, the UNT repository (like KISTI) was structured such that page views were recorded separately for individual ETD pages. Even though

users were probably reading only a few pages online before downloading the PDF (Connaway and Dickey, 2010), this repository arrangement made it impossible to directly compare UNT page views to those of a repository like AUETDS which contained only PDFs.

Even though page views are not comparable, however, user visits to the two repositories may be compared. User location data provided by Alemneh and Phillips show that Denton, TX, (i.e., Local) users made 3.6% of the total visits to the UNT repository. Extrapolating from other location data provided, users in the rest of the state (i.e., Texas–not Local) made 7.9% of the total visits, users in the rest of the country (i.e., USA–not Texas) made 57.4% of the total visits, and users in the rest of the world (World–not USA) made 31.1% of the total visits.

A comparison of these numbers with those given in Table 1 shows a significant difference in visit numbers for Local user groups between UNT (3.6%) and AU (14.2%). This probably arises, at least in part, from differences in ETD upload procedures for the respective repositories. It appears that UNT library personnel uploaded ETDs to the repository (Phillips and Alemneh, 2011), while AU students uploaded their own. Since AU students had to visit the AUETDS Web site to submit their ETDs, this probably resulted in increased Local traffic to the site. It is difficult to predict how the absence of this traffic would have affected the overall distribution of visits from the various location groups.

Extrapolating from UNT data on sources, Web search engines provided 62.4% of the overall traffic to the UNT repository, while University Web pages provided 3.5%. Table 4 shows that 68.3% of AUETDS traffic came from Web search engines and 12.6% from University Web pages. Although the search engines numbers could be considered roughly comparable, the difference in traffic from University Web pages seems significant. It was probably related to the difference in ETD uploading procedures mentioned earlier. AU students uploading ETDs had to

access the repository and frequently used University Web pages to do so.

Conclusions

The first research question this study addressed was: How do users in different locations find the AUETDs collection? Users in Auburn and Opelika came to the collection via a combination of University Web pages, Web search engines, and direct access. The two out-of-state groups were much more likely to use Web search engines and much less likely to use University Web pages. Alabama users not located in Auburn and Opelika exhibited behavior in between that of the local users and the out-of-state users.

The second question addressed by this study was: Do users in different locations interact differently with the collection and, if so, how? Users in different locations did interact differently with the collection. Alabama users landed on the collection home page more often, viewed more pages per visit, and viewed internal navigation pages (browse, search, advanced search) and ETD submission pages more frequently than users in locations outside of Alabama.

One probable reason for the observed differences in behavior is that many Alabama users had prior knowledge regarding the existence of AUETDs and went to it intentionally, while most out-of-state users found items in the collection while searching generally for information. Another probable reason is that many Alabama users were depositors, instead of, or in addition to, being end-users, while the two out-of-state groups consisted largely of end-users. Depositors and end-users need different things from a repository, so it is to be expected that they will interact differently with it.

An institution interested in improving a repository's access, navigation, etc., from the point of view of depositors, therefore, will probably find it helpful to focus solely on usage

reports for in-state users. This group probably already knows about the repository and will use institutional resources to find it. Creating more local pathways will make these collections easier for this group to find. In-state users will probably also use the collection's internal navigation tools to locate the materials they need. Anything that can be done to expand these internal tools (e.g., full-text searching, indexing by department) or to make them easier to use will help this group.

However, when seeking to improve the repository for end-users, it is probably more useful to exclude in-state users from usage reports. For most end-users, finding a particular ETDs repository or even a particular ETD is not their goal; it is a by-product of their information-seeking process. Research shows that information seekers are more and more likely to begin their quest with a Web search engine (Tenopir and Rowlands, 2007). Therefore, exposing a repository collection to search engine indexing is the most important step towards increasing its use. Moreover, anything that can improve the quality of search engine indexing (making documents' full text available for indexing, for example) will help end-users find relevant materials in the collection.

McKay said that, without end-user research, one couldn't know (a) whether a repository's users were local or were located outside the institution; (b) whether end-users found the repository via institutional sources or external search engines; (c) what kind of information they sought and used; and (d) how they used the functionality offered by the repository (McKay, 2007). This study provides answers for the AUETDs collection for part (a)—only 14.2% of the visits to the collection were from users in Auburn and Opelika—and part (b)—12.6% of all visitors came to the collection via University Web pages, while 68.3% came via Web search engines (with the former used more by local visitors and the latter used more by visitors outside

the University community). This study also provides a partial answer for part (d)—35.6% of all pages viewed were internal navigation pages (with local visitors viewing these more often than external visitors). Answers to part (c)—the search terms used to find the collection and the types of items viewed and downloaded—will be the subject of future research.

References

Alemneh, D. G., and Phillips, M. E. (2011), "Assessing the usage of electronic theses and dissertations: an overview of ETD statistics and metrics in the UNT libraries", paper presented at Texas ETD Association Annual Conference, 31 March 2011, Arlington, Texas, available at: <http://digital.library.unt.edu/ark:/67531/metadc32969/> (accessed 9 May 2013).

Armbruster, C., and Romary, L., (2010), "Comparing repository types: challenges and barriers for subject-based repositories, research repositories, national repository systems and institutional repositories in serving scholarly communication", *International Journal of Digital Library Systems*, Vol. 1, No. 4, pp. 61-73.

Bailey, C. W., Jr., (2011), "Institutional repository and ETD bibliography 2011", available at <http://www.digital-scholarship.org/iretd/iretd.pdf> (accessed 29 August 2013).

Connaway, L. S., and Dickey, T. J., (2010), "The digital information seeker: report of the findings from selected OCLC, RIN, and JISC user behaviour projects", available at: <http://www.jisc.ac.uk/media/documents/publications/reports/2010/digitalinformationseekerreport.pdf> (accessed 10 May 2013).

Google, (2013), "Get the power of Google Analytics", available at: <http://www.google.com/analytics/premium/features.html> (accessed 12 January 2013).

Ismail, M. A., and Kareem, S. A., (2011), "Identifying how novice researchers search, locate,

choose and use web resources at the early stage of research”, *Malaysian Journal of Library & Information Science*, Vol. 16, No. 3, pp. 67-85.

Lee-Smeltzer, J., and Hackleman, D. (1995), “Access to OSU theses and dissertations in Kerr Library: how are they used ... or are they?”, *Technical Services Quarterly*, Vol. 12, No. 4, pp. 25-37.

Lynch, C. A., and Lippincott, J. K., (2005), “Institutional repository deployment in the United States as of early 2005”, *D-Lib Magazine*, Vol. 11, No. 9, available at:

<http://www.dlib.org/dlib/september05/lynch/09lynch.html> (accessed 17 July 2013).

McKay, D., (2007), “Institutional repositories and their ‘other’ users: usability beyond authors”, *Ariadne*, No. 52, available at: <http://www.ariadne.ac.uk/issue52/mckay> (accessed 6 April 2013).

Mukherjee, B., and Nazim, M., (2011), “Open access institutional archives: a quantitative study (2006-2010)”, *DESIDOC Journal of Library & Information Technology*, Vol. 31, No. 4, pp. 317-324.

NDLTD: Networked Digital Library of Theses and Dissertations, (2012), available at:

<http://www.ndltd.org/> (accessed 25 May 2013).

Open Access Theses and Dissertations: OATD, (2013), available at: <http://oatd.org/> (accessed 25 May 2013).

Organ, M. K., (2006), “Download statistics—what do they tell us? The example of Research Online, the open access institutional repository at the University of Wollongong, Australia”, *D-Lib Magazine*, Vol. 12, No. 11, available at:
<http://www.dlib.org/dlib/november06/organ/11organ.html> (accessed 6 April 2013).

Phillips, M. E., and Alemneh, D. G., (2011), “Data desiccation: facilitating long-term access, use, and reuse of ETDs”, in *Proceedings of the 14th International Symposium on Electronic Theses and Dissertations, Cape Town, South Africa, 13-17 September 2011*, National Research Foundation, available at: http://dl.cs.uct.ac.za/conferences/etd2011/papers/etd2011_phillips.pdf (accessed 15 July 2013).

Schmitz, D., (2008), “The seamless cyberinfrastructure: the challenges of studying users of mass digitization and institutional repositories”, Council on Library and Information Resources, available at: <http://www.clir.org/pubs/archives/schmitz.pdf> (accessed 29 August 2013)

St. Jean, B., Rieh, S. Y., Yakel, E., and Markey, K., (2011), “Unheard voices: institutional repository end-users”, *College & Research Libraries*, Vol. 72, No. 1, pp. 21-42.

Tenopir, C., and Rowlands, I., (2007), “Information behaviour of the researcher of the future: age-related information behaviour”, available at:
<http://www.jisc.ac.uk/media/documents/programmes/reppres/ggworkpackageiii.pdf> (accessed on 11 May 2013).

Virginia Tech University Libraries, (2013), “Electronic theses and dissertations at VT: electronic theses and dissertations online access data”, available at:

<http://scholar.lib.vt.edu/theses/data/somefacts.html> (accessed 18 July 2013).

Ware, M., (2004), “Institutional repositories and scholarly publishing”, *Learned Publishing*, Vol. 17, No. 2, pp. 115-124.

Wikipedia, (2013), “Google Analytics”, available at

http://en.wikipedia.org/wiki/Google_Analytics (accessed 12 January 2013).

Yiotis, K., (2008), “Electronic theses and dissertation (ETD) repositories: what are they? where do they come from? how do they work?”, *OCLC Systems & Services: International Digital Library Perspectives*, Vol. 24, No. 2, pp. 101-115.

Zhang, Y., Lee, K., and You, B.-J., (2001), “Usage patterns of an electronic thesis and dissertations system”, *Online Information Review*, Vol. 25, No. 6, pp. 370-377.