

The ETD lens on the IR and the University

Gail McMillan

Scholarly Communication, Virginia Tech Libraries, USA

gailmac@vt.edu

Abstract

Electronic Theses and Dissertation (ETDs) form the foundation of many institutional repositories (IR), and ETDs can provide an important lens on how the IR represents the work that is being done university-wide. To test this theory, I studied a microcosm of VTechWorks, the IR at Virginia Tech, anticipating that it would encapsulate the characteristics of the repository as a whole. I analyzed the Lesbian, Gay, Bisexual, Transgender, Queer/Questioning (LGBTQ) microcosm through the ETD lens.

I compiled a controlled vocabulary of 161 LGBTQ terms and phrases that I searched in the IR's ETD collection, recording number of hits for each term. To give context to this data and to understand more broadly who is doing the scholarship and research in this microcosm, I gathered the same data from the IR's faculty publications and academic units' web publications which are often aimed at alumni or the public. I searched the same vocabulary in the university website to establish the measure of the LGBTQ domain and to contextualize the data from the IR.

Using a microcosm to look at graduate student works in the IR and university web publications provides manageable data and the opportunity to reflect on the contribution of ETDs to the development of institutional repositories.

Keywords: IR content, unconscious bias, LGBTQ, institutional repositories, assessment

Introduction

ETDs form the foundation of many institutional repositories (IRs) that manage and preserve community assets and gray literature as well as peer-reviewed publications. An IR gives the university both a digital library and a showcase, but **do IRs accurately reflect their home institutions?** Assessing IRs from the perspective of the institution's scholarly output is a little used frame of reference, and ETDs can provide an important lens to help assess whether

the IR is representative of the scholarship and activities of the home institution. This study was a foray into one aspect of how we might assess our universities' digital repositories.

IR content is not developed like library collections with subject specialists who have budgets assigned to areas that represent the foci of the university's research and teaching. Instead, IRs depend on people's time to locate and deposit works. VTechWorks, Virginia Tech's seven-year-old IR, has been populated using a variety of methods. There is self-deposit by members of the university community—by choice or by edict. For example, our faculty may deposit into the IR through integrated systems like Symplectic [Elements](#), which Virginia Tech uses as an electronic faculty activity data system to collect and manage information about research and scholarship, creative works, teaching, extension, outreach, and service activities. Graduate students deposit [ETDs approved by the Graduate School](#) through a local online system, and some students deposit directly into VTechWorks as a course requirement.

VTechWorks staff select and deposit both unsystematically, for example, based on something they read about in the daily online VT News. They also make project-based deposits, including articles supported by the library's [Open Access Subvention Fund](#). VTechWorks also receives automated harvests from publishers such as BioMed Central, Hindawi, and MDPI who use the SWORD protocol.

VTechWorks hosted more than 71,000 works in April 2019 when I collected the data for this study. About 99% of the content was publicly available and about 85% were textual works.

Literature Review

Among the wealth of articles written about IRs, there is a dearth of articles about content development and content scope analysis. Assessing IRs from the perspective of the institution's scholarly output is a little-used frame of reference. Instead, articles about assessing IRs tend to cover topics such as usability, preservation, marketing, workflow, intellectual property, etc. But these articles do not attempt to assess the relationship of the content to the research and teaching foci of the home institution as an indicator of the value of the IR.

Articles and presentations that addressed diversity within the academy influenced this study. At the 2017 Coalition for Networked Information's fall meeting, Amanda Rust, a librarian at Northeastern University (Boston MA, USA) presented «Design for Diversity». This grant funded project focused on ways in which information systems embody and reinforce cultural norms (e.g., data models that enforce strict gender binaries), and she addressed designing systems that take into account diverse cultural materials.

Sam Winn, a Special Collections librarian at Virginia Tech (Blacksburg VA, USA), made several salient points in her article, «The Hubris of Neutrality in Archives», that could also be applied to IRs. She pointed out [p.2] that ...archivists contribute to the omission or erasure of

historically marginalized groups in the archives. And, that ...a «radically inclusive historical record» [like an IR] will not happen by accident.

Rebekah Scoggins, a librarian at Leander University (Greenwood SC, USA) reported in «Broadening Your Library's Collection: Implementing a LGBTQIA Collection Development Project» that her library was not meeting user needs because its collection was out-of-date and incomplete. Her article made me think about the limitations of studies that only consider meeting community needs through the traditional library collection (i.e., purchased books, serials, and other media) and not considering the content in the IR.

These works made me ask, «How would we know if our repository, VTechWorks, lacked diverse cultural materials and was contributing to the omission of historically marginalized groups?» These three works and the dearth of articles about IR scope analysis, influenced me to conduct a study to try to determine if VTechWorks was providing an inclusive record as well as being a digital library and showcase. Therefore, I chose to analyze the microcosm of LGBTQ works in VTechWorks, and in Virginia Tech's ETDs in particular but also VTechWorks collections of faculty scholarship and academic units web publications, and to compare these findings to the LGBTQ microcosm of university web publications.

Methodology

From several academic and community resources I compiled a controlled vocabulary of 161 LGBTQ terms and phrases. (See Appendix A) This list of terms (See Appendix B) may not be bias free or comprehensive, but it is a starting place for inquiry into the LGBTQ domain in the IR. (To reduce wordiness throughout this report, when I refer to terms, I mean both terms and phrases.)

Some of the terms were not practicable because they have historically different meanings, such as «gay» and «queer», I eliminated other terms because they were too broad, such as discrimination, but I did not exclude biological terms. The IR software, DSpace, uses the Solr search platform, which allows «fuzzy» (i.e., close) matches. Solr also «stems», that is, it expands words with common endings to include plurals, past tenses, etc. I used the common search practice of putting quotes around phrases.

To refine my investigation, to help understand who is doing the scholarship and research in the LGBTQ microcosm, and to give context to the data, I searched the 161 terms in five collections: (1) the university at large, (2) VTechWorks at large as well as within its collections of (3) ETDs, (4) peer-reviewed faculty publications, and (5) academic units' web publications. (At Virginia Tech we call academic units' colleges, so in this paper I refer to them as the Colleges Collection.) Table 1 shows the number of items in each collection, how many of the 161 terms got hits, and how many hits were found in each collection.

# items	collection	terms matched	hits
71,734	VTechWorks	1 29	21,455
32,557	ETDs	1 15	9,017
3,870	Faculty Research	4 0	858
14,590	Colleges	8 9	4,067
	Virginia Tech vt.edu	1 09	84,793

TABLE 1

I did not search vt.edu from the university's homepage because the indexing of Virginia Tech's websites includes VTechWorks. Instead, I searched (<https://www.google.com/>) using this search strategy: [term] site:vt.edu -site:vtechworks.lib.vt.edu -site:theses.lib.vt.edu. Searching VT in this way resulted in 109 of the terms getting 84,793 hits.

Because the five collections have radically different sizes and ages, I converted the number of hits for each term in each collection into a percentage of hits. I then compared the percentage of hits within and across collections which provided the lens of the graduate students through their ETDs, the faculty's use of the terms in their peer-reviewed publications, and the academic units' and university-wide use in works often aimed at the general public and alumni.

Results and Discussion

Creating spreadsheets for the search results of each collection allowed me to sort the data in various ways. Table 2 shows the search results for 15 terms sorted by the largest number of hits in the ETD collection. The term «gender» got nearly 50% of the hits, leaving the remaining 114 terms with hits between 3.5% and 0.01%.

ALL TERMS	% of hits in ETDs	VT ETDs in VTW (32,557) 9017 hits	ETDs oldest	ETDs newest
gender	49.2%	4433	1962	2019
heterosexual	3.5%	317	1955	2019
homosexual	3.3%	295	1968	2019
hegemonic	3.0%	269	1982	2019
sexual orientation	2.9%	266	1981	2019
lesbian	2.9%	259	1975	2019
asexual	2.5%	222	1910	2019
sexism	2.4%	218	1972	2019
gender identity	2.0%	182	1981	2019
gender bias	1.9%	174	1984	2019
androgynous	1.9%	169	1975	2018
bisexual	1.8%	159	1955	2019
gender equality	1.5%	136	1987	2018
gender inequality	1.3%	120	1989	2019
gender inequity	1.3%	120	1989	2019

TABLE 2

Table 3 shows search results for four collections (excluding VTechWorks at large) sorted by largest number of hits in the ETD Collection. Highlighted are the terms that got 2% or more of the hits.

% of hits in VT	ALL TERMS	% of hits in ETDs	% of hits in Fac Res	% hits in Colleges
39.3%	gender	49.2%	52.8%	48.9%
1.8%	heterosexual	3.5%	0.3%	0.8%
2.6%	homosexual	3.3%	0.9%	0.6%
1.5%	hegemonic	3.0%	0.8%	1.0%
13.8%	sexual orientation	2.9%	5.9%	13.8%
3.1%	lesbian	2.9%	0.9%	1.3%
0.8%	asexual	2.5%	3.1%	1.7%
2.2%	sexism	2.4%	1.5%	0.8%
7.8%	gender identity	2.0%	3.8%	3.8%
0.6%	gender bias	1.9%	9.3%	1.4%
0.3%	androgynous	1.9%		0.2%
1.6%	bisexual	1.8%	0.6%	1.2%
0.7%	gender equality	1.5%	0.9%	1.1%
0.5%	gender inequality	1.3%	1.2%	1.6%
0.1%	gender inequity	1.3%	1.2%	1.5%
1.7%	transgender	1.2%	0.1%	1.1%
1.4%	gender equity	1.1%	3.5%	2.6%
0.6%	gender neutral	1.0%	0.8%	0.8%
0.1%	lifestyle choice	1.0%	0.7%	0.2%
1.5%	underrepresented groups	0.9%	1.5%	2.7%
0.8%	homophobia	0.9%	0.1%	0.3%
0.5%	gay men	0.8%	0.5%	0.3%
0.0%	binarism	0.7%	0.1%	0.3%
0.2%	hermaphrodite	0.6%	0.8%	0.3%
0.4%	gay rights	0.6%		0.2%
0.1%	biological sex	0.6%	0.3%	0.1%
0.8%	homophobic	0.5%		0.2%
0.3%	sexual preference	0.5%		0.1%
0.2%	heteronormative	0.4%	0.5%	0.3%
2.3%	LGBTQ	0.4%	0.3%	1.1%
0.2%	heterosexism	0.4%		0.2%
0.1%	gender disparity	0.4%	0.6%	0.6%
0.2%	transsexual	0.4%		
0.3%	misogyny	0.3%		0.1%
0.1%	secondary sex characteristics	0.3%		
0.2%	intersex	0.3%	0.2%	0.5%
0.3%	gay marriage	0.3%		0.0%
0.2%	faggot	0.3%		0.1%
0.1%	gender binary	0.3%	0.1%	0.2%
0.3%	unconscious bias	0.3%	0.5%	0.3%
7.0%	gender expression	0.2%	3.6%	2.9%

TABLE 3

In ETDs «heterosexual» got the next highest percentage of hits after «gender». While only 3.5%, however, it was nearly twice what the same term got in vt.edu, 1.8%. Graduate students used it ten times more often than faculty in the VTechWorks collections studied. Other terms

that graduate students used three-to-nine times more than faculty, included «bisexual», «lesbian», «binarism», «hegemonic», «homophobia», and «homosexual». Faculty used only two terms significantly more often than graduate students. Faculty used «gender expression» 18 times more, and «gender» bias five times more.

Using vt.edu as the measure of LGBTQ scholarship and activities at the university, and comparing the percentage of hits in ETDs with the percentage of hits at vt.edu provides some evidence as to whether studying a microcosm is indicative of the university community as a whole. Table 4 shows search results for the top 20 hits in VT collections and compares the percentage of hits in each.

% VT	VT ranking	109 terms / phrases	ETDs ranking	% ETDs
39.3%	1	gender	1	49.2%
13.8%	2	sexual orientation	5	2.9%
7.8%	3	gender identity	9	2.0%
7.0%	4	gender expression	41	0.2%
3.1%	5	lesbian	6	2.9%
2.6%	6	homosexual	3	3.3%
2.3%	7	LGBTQ	30	0.4%
2.2%	8	sexism	8	2.4%
1.8%	9	heterosexual	2	3.5%
1.7%	10	transgender	16	1.2%
1.6%	11	bisexual	12	1.8%
1.5%	12	underrepresented groups	20	0.9%
1.5%	13	hegemonic	4	3.0%
1.4%	14	gender equity	17	1.1%
0.8%	15	asexual	7	2.5%
0.8%	16	homophobia	21	0.9%
0.8%	17	homophobic	27	0.5%
0.7%	18	gender equality	13	1.5%
0.6%	19	gender bias	10	1.9%
0.3%	20	androgynous	11	1.9%

TABLE 4

There were four notable disparities in the terms used more frequently at vt.edu than in the ETD Collection: «sexual orientation» (13.8% vs 2.9%), «gender identity» (7.8% vs 2.0%), and «gender expression» (7.0% vs 0.2%). These terms appear more in public-facing web documents than in graduate student's academic research output. Only one term appeared with considerably more frequency in ETDs: «gender» (49% vs 39%), that is «gender» was used more in graduate students' research than in the academic units' public facing university web sites. After analyzing terms that appeared in ETDs but not in VT at large and vice versa, the disparities were too small to be noteworthy for this study.

I initially thought it would be interesting to compare faculty and graduate students' non-use of the terms also. But, you may recall that of the 161 terms I searched, 115 appeared in ETDs but only 40 appeared in Faculty Research. There were 75 unused terms, which was too many to be meaningful for this study. [See my November 2019 article in *Against the Grain* for a fuller discussion of the faculty publications and college collections.]

Conclusion

The goal of this study was to investigate whether the Institutional Repository represents the scholarship and activities of its home institution, and whether this can be determined by studying a microcosm of the Repository, and particularly its ETD Collection, and comparing it with the same microcosm in the university through its website, vt.edu. This study used a controlled vocabulary to analyze the LGBTQ microcosm in the university website, vt.edu, and the VTechWorks, particularly its ETD collection. Using a microcosm to look at graduate students' ETDs and university website provided the opportunity to reflect on the contribution of ETDs to the development of Institutional Repositories through a manageable data set. The percentage of search hits on the selected terms at the university's website was the measure of the LGBTQ domain and contextualized the data from ETDs as well as the whole Repository.

Since 94% of the LGBTQ terms appearing in the university website through a Google search of vt.edu also appeared in the VTechWorks, and 94% of the terms had a less than two percent (<2%) difference in the percentage of hits, the institutional repository as a whole reflects quite well the scholarship and activities of the university in this microcosm.

The discrepancy in the percentage of hits between VT and the VTechWorks collection of ETDs for the 99 terms they had in common is less than one percent (<1%). 91% of the terms have a nearly equal percentage of hits in each collection, therefore, the ETD lens quite accurately reflects the scholarship and activities at the university, at least in the LGBTQ microcosm. It will be necessary to study additional microcosms in order to better understand whether this is a high enough correlation to determine if the IR and its ETD Collection are truly representative of the scholarship and activities at the university.

[I have begun a similar study of Indigenous Peoples. If anyone is interested in commenting on the study methodology and criteria to assess IRs, and/or would be interested in conducting similar studies to compare across institutions, I'd like to hear from them: gailmac@vt.edu.]

References

- Rust, A. (2017). Design for diversity: Towards inclusive information systems for cultural heritage. Coalition for Networked Information [presentation] Dec. 9, 2017. <https://www.cni.org/topics/assessment/design-for-diversity-towards-inclusive-information-systems-for-cultural-heritage> Accessed May 15, 2019.
- Scoggins, R. (2018). Broadening your library's collection: Implementing a LGBTQIA collection development project. *C&RL News*, 79(3), 114–116, 126. <https://doi.org/10.5860/crln.79.3.114> Accessed May 15, 2019.
- Winn, S. (2017). The hubris of neutrality in archives. *On Archivy*. <https://medium.com/on-archivy/the-hubris-of-neutrality-in-archives-8df6b523fe9f> Accessed May 15, 2019.
- McMillan, G. (2019). Does the Repository Reflect the Institution? *Against the Grain*. [forthcoming November 2019]