SAVANNAH LAKE AND JOSEPH NICHOLSON UNC CHARLOTTE

Remediation by Degrees:
Enhancing ETD Metadata to Improve Discoverability

In 2022, staff at J. Murrey Atkins Library launched a project to remediate metadata for electronic theses and dissertations (ETDs) in the Niner Commons institutional repository, which hosts UNC Charlotte faculty, staff, and student scholarship on an open access model. Received several times a year in files encoded in ProQuest's own XML ETD metadata standard, which Atkins transforms into MODS, the ETD metadata in Niner Commons provided a basic level of access to student work but was marred by capitalization irregularities in title and note fields and, crucially, by the lack of controlled subject terms in the FAST (Faceted Application of Subject Terminology) vocabulary used in records for all other works in Niner Commons. The absence of controlled subject terms thwarted subject access to the ETD collection except through student-supplied keywords, which are generally poor in quality, and terms from ProQuest's own subject vocabulary. The remediation project addressed these metadata deficits by matching ProQuest subject terms in Niner Commons ETD metadata against FAST subject terms in an OpenRefine reconciliation procedure and inserting the terms into legacy records using XSLTs (Extensible Stylesheet Language Transformations), while making smaller adjustments to capitalization and style. The remediation project was, however, limited in scope, and did not address problems in other areas of the ETD records or attempt to rethink ETD metadata workflows at UNC Charlotte, which involve repository records in MODS and catalog records in MARC that are created through separate processes and staff and differ in quality.

This case study provides an account of the ETD metadata remediation project at Atkins Library, delineating the metadata problems it was designed to address, the remediation methods and tools used, the problems encountered during the course of the work, and the results of the project and findings. In describing Atkins' remediation process, it also reflects on some of the possibilities and contradictions of ETD metadata remediation work in the contemporary institutional repository environment, where staff shortages, legacy cataloging practices in other library units, and ProQuest's own distribution channels for ETD metadata can limit libraries' ability to ensure metadata quality and consistency across different systems and record formats. Atkins's experience suggests that a phased approach that does not tackle all remediation issues at once may be a viable strategy for remediating ETD metadata for institutions coping with staffing and technology constraints.

Literature Review

Doctoral dissertations, and to a lesser extent, master's theses, have been publicly distributed within the United States as far back as the 1930s, with microfilm copies facilitating relatively inexpensive and efficient distribution.[1] With the advent of digital publishing and online repositories, ETDs are even more readily available, with libraries playing an active role in this work. The Open Directory of Open Access Repositories (OpenDOAR) shows that of the 646 institutional repositories based in the United States and posting scholarly content like journal articles, 469 also report posting ETDs.[2] This means that approximately 73% of scholarly institutional repositories also host ETDs, suggesting that ingesting and managing ETD content is an ever-present responsibility of academic libraries.

In managing such workflows, libraries must consider whether digitally disseminating ETDs through ProQuest, an institutional repository, or both is the best fit. Such a decision involves careful consideration of staff bandwidth, discovery potential, and the costs of using a commercial publisher. ProQuest has administered

[1] Gail P. Clement and Fred Rascoe, "ETD Management & Publishing in the ProQuest System and the University Repository: A Comparative Analysis," *Journal of Librarianship and Scholarly Communication* 1, no. 4 (2013): 2-3. https://doi.org/10.7710/2162-3309.1074.

[2] "Advanced Search - Sherpa Services," OpenDOAR, accessed July 23, 2023, https://v2.sherpa.ac.uk/cgi/search/repository/advanced.

digital ETDs for over twenty-five years now, as far back as 1997, gaining widespread buy-in and momentum around 2006.[3] Given this legacy, the ProQuest Dissertations and Theses database (PQDT) holds appeal, as it is one of the largest databases of graduate works.[4] Additionally, working with ProQuest to distribute ETDs can be especially helpful to libraries with smaller cataloging and repository teams, as they may not have the staff to commit to collecting and cataloging several hundred ETDs on an annual basis. At the same time, housing ETDs in institutional repositories offers marked advantages, such as eliminating submission fees for students and collocating ETDs alongside faculty work as well as other graduate non-ETD work, such as capstone projects, articles, and conference proceedings. It is not surprising, then, that a 2017 survey of ETD policies and practices found that many institutions take advantage of both platforms; of 51 respondents, 40 load ETD metadata into their institutional repository and 24 load into PQDT, with the library catalog and OCLC WorldCat being popular destinations as well (34 and 29 respondents, respectively).[5]

For UNC Charlotte, the benefits of both platforms were clear, and we have similarly opted to have theses and dissertations featured in both. Dual online submission of ETDs into institutional repositories and PQDT is made possible through a variety of different workflows, including utilizing the ProQuest ETD Administrator, FTP, or harvesting.[6] With metadata records generated by ProQuest in its own XML–as opposed to an established schema like MODS, which our repository uses–a key part of our local workflow with ingesting ETDs into the institutional repository involves crosswalking ProQuest metadata to MODS. In doing this for several years, we have navigated several issues with repurposing ProQuest's metadata for our own repository. These issues have been documented in the literature as well; a case study from University of Iowa Libraries, for example, discussed limitations with ProQuest metadata, including a lack of departmental mapping, which prevents users from browsing ETDs alongside other works coming from the same department and hinders departments from getting a cohesive picture of their scholarly output.[7]

Even without the complicating factor of crosswalking metadata from ProQuest's schema, metadata can be a sticking point for ETD management. A review of thirteen conferences on ETDs and gray literature, for example, specifically recommended metadata improvements as a way to add value to ETDs housed in institutional repositories, to improve their discoverability.[8] ETDs, in particular, are subject to have "considerable variations" with metadata, such as differing descriptors to describe university programs, degree levels, and dates (which can range from the date the ETD was made available online, to the date it was submitted, to the date the student graduated).[9] The study reviewed repositories using ProQuest XML metadata records as well as other standards, suggesting that metadata remediation is a key component of any form of ETD management.

Like many other universities navigating the terrain of ETD management and dual online submission, historically we have addressed the differences in ProQuest's XML with our local standards through crosswalking with XSLTs. Adjusting for subject terminology has been a bit trickier, as ProQuest uses its own controlled vocabulary instead of Library of Congress Subject Headings (LCSH) or FAST.[10] Previously, we had simply carried over ProQuest's supplied terms, despite feeling that the

[3] Marielle Veve, "ETDs in ProQuest and the Institutional Repository: A Descriptive Study of the Current Workflows Available for Dual Online Submission," *The Journal of Academic Librarianship* 47, no. 5 (2013): 1-2. https://doi.org/10.1016/j.acalib.2021.102429.

[4] Clement and Rascoe, "ETD Management & Publishing in the ProQuest System and the University Repository," 17.

[5] Emily Alinder Flynn and Janet H. Ahrberg, "Electronic Theses and Dissertations (ETDs) Metadata Policies, Workflows, and Practices: A Survey of the ETD Metadata Lifecycle at United States Academic Institutions," *Journal of Library Metadata* 20, no. 2–3 (2020): 102-103. https://doi.org/10.1080/19386389.2020.1780689.

[6] Veve, "ETDs in ProQuest and the Institutional Repository," 3.

[7] Shawn Averkamp and Joanna Lee, "Repurposing ProQuest Metadata for Batch Ingesting ETDs into an Institutional Repository," *The Code4Lib Journal*, no. 7 (2009). https://journal.code4lib.org/articles/1647.

[8] Joachim Schöpfel, "Adding Value to Electronic Theses and Dissertations in Institutional Repositories," *D-Lib Magazine* 19, no. 3/4 (2013): 4. https://doi.org/10.1045/march2013-schopfe.

[9] Eun G. Park and Marc Richard, "Metadata Assessment in E-theses and Dissertations of Canadian Institutional Repositories," *The Electronic Library* 29, no. 3 (2011): 404. https://doi.org/10.1108/02640471111141124.

[10] "Institution Dissertations FAQ," accessed July 23, 2023, https://about.proquest.com/en/dissertations/proquest-dissertations-frequently-asked-questions/proquest-dissertations-institutions-frequently-asked-questions/.

terms were insufficient for meaningful discovery within our systems. Other libraries take a similar approach; in their case study of repurposing ProQuest metadata, Pennsylvania State University Libraries similarly uses ProQuest-provided subject terminology, relying on these user- and ProQuest-generated records to streamline procedures in a time in which "cataloging and metadata departments are being asked to provide new services while still keeping up with traditional workflows."[11] In fact, the university ceased manual LCSH subject cataloging several decades ago in 1975 for most of their dissertations in the interest of expediting workflows.[12]

Conversely, "through the efforts of the special format unit and many others involved in the process," University of Arkansas Libraries perform record-by-record subject analysis to ProQuest ETD metadata, to ensure LCSH subject terms are applied to ETDs. This has yielded meaningful impacts on discovery, as a subsequent survey found that library users and reference librarians credit the subject headings for improving access.[13] While perhaps best practice, record-by-record cataloging may be aspirational or out of reach for many. A case study of an ETD remediation effort at the University of Houston Libraries, for example, found assigning LCSH terms to ETD records required the additional help of a cataloging librarian and ultimately was too significant a commitment of time and labor to continue.[14] More broadly, a 2016 study of institutional repositories posting ETDs found that 61% of repositories relied on author-submitted keyword terms and 28% used another standardized thesaurus, while only 31% used LCSH.[15] Though the study did not specify, there is likely some overlap between the respondents who mentioned another standardized thesaurus and author-submitted keyword terms, as ProQuest provides both in its metadata records, with controlled subject terminology coming from its own vocabulary along with student-submitted subject keywords.

Limited staff hours are a recurrent issue in metadata creation, experienced by many libraries and cited as adversely impacting metadata quality.[16] As a library with staff constraints, we were interested in exploring automated or batch efforts for assigning controlled subject terminology to improve our metadata quality and discovery experience while also acknowledging our limited staff bandwidth. A promising presentation from the University of South Florida describes crosswalking ProQuest metadata via an XSLT, with a brief mention of utilizing the XSLT to append LCSH terms to ETD records.[17] This was of particular interest to our work, for its potential to partially automate what can be a time-consuming process. Balancing automated processes and ProQuest-provided metadata with our local standards for metadata quality, we aim to add to the growing literature on managing and remediating ETD metadata records from ProQuest, specifically in the space of subject metadata, to provide a robust analysis and case study that will help other universities replicate our process and facilitate better discovery of ETD records.

Problem Space

ProQuest metadata records include two types of subject metadata: "subjects," drawn from its in-house controlled vocabulary and applied by students during the ETD submission process, and "keywords," which are descriptors created and supplied by students.[18] Importantly,

[11] Ken Robinson, Jeff Edmunds, and Stephen C. Mattes, "Leveraging Author-Supplied Metadata, OAI-PMH, and XSLT to Catalog ETDs: A Case Study at a Large Research Library," *Library Resources & Technical Services* 60, no. 3 (2016): 200. https://doi.org/10.5860/lrts.60n3.191.

[12] Ibid., 192-195.

[13] Cedar C. Middleton, Jason W. Dean, and Mary A. Gilbertson, "A Process for the Original Cataloging of Theses and Dissertations," *Cataloging & Classification Quarterly* 53, no. 2 (2015): 240-245. https://doi.org/10.1080/01639374.2014.971997.

[14] Santi Thompson, Xiping Liu, Albert Duran, and Anne Washington, "A Case Study of ETD Metadata Remediation at the University of Houston Libraries," *Library Resources & Technical Services* 63, no. 1 (2019): 74. https://doi.org/10.5860/lrts.63n1.62.

[15] Tom Steele and Nicole Sump-Crethar, "Metadata for Electronic Theses and Dissertations: A Survey of Institutional Repositories," *Journal of Library Metadata* 16, no. 1 (2016): 53–68. https://doi.org/10.1080/19386389.2016.1161462.

[16] Heather Moulaison Sandy and Felicity Dykas, "High-Quality Metadata and Repository Staffing: Perceptions of United States–Based OpenDOAR Participants," *Cataloging & Classification Quarterly* 54, no. 2 (2016): 113. https://doi.org/10.1080/01639374.2015.1116480.

[17] Annie Glerum and Dominique Bortmas, "Migrating ETDs from Dublin Core to MODS: Automated Processes for Metadata Enhancement," presented at the ALCTS Metadata Interest Group Virtual Pre-Conference (2016): 54.

[18] "ProQuest Electronic Thesis and Dissertation (ETD) Administrator: Student Submission Libguide" (2023): 20. https://proquest.libguides.com/ld.php?content_id=64364001.

neither of these subject terms align with established controlled vocabularies like LCSH or FAST. This negatively impacts discovery within our institutional repository, which uses FAST, since ETDs on the same subjects as other works within the repository are not assigned the same subject term, preventing collocation and browse.

In addition to these issues with collocation, we also found some of the subject metadata within these ProQuest records to be of such low quality as to be effectively useless in helping users find works. This is especially the case with the "keywords," which are wholly uncontrolled and generated by students. While the ProQuest team reviews and edits this self-submitted metadata, there remain significant issues. We currently have twenty ETDs on "Applied Physics,"[19] for example. Instead of meaningful, descriptive subject terms, these papers often have keywords that are so broad as to essentially be meaningless (such as "Flipped," "Design," and "Inverted"); or conversely, so specific that they are unlikely to be used by many to browse the repository (such as "Bovine Serum Albumin," "Choline Dihydrogen Phosphate," and "Centrifugal Radial Inflow Bubble Heating"). Even worse are the terms that are essentially synonyms but show up in different variants, since these keywords are uncontrolled (such as "Bohm" and "Bohmian" as well as "Sleeping Beauty Transposase" and "Sleeping Beauty Transposon"). These various issues add noise, creating a long tail of keywords that have only one work associated; within our own repository of 3,090 ETDs, there are over 6,000 keywords with just one associated work.[20]

In thinking through our repository holdings as a whole, we identified this disconnect with subject metadata between ETDs and the repository at large as a meaningful area for improvement. In addition to providing for better discovery and collocation within other materials within the repository, we saw making this metadata improvement as an investment in our repository, which is relatively young. As the repository becomes more established, we hope to create additional research

support services for our campus community. Such services could include generating metrics and reports for departmental administrators, for example, so that administrators would have a better understanding of the scholarly output of their faculty and students. Creating a more cohesive metadata ecosystem within the repository will be instrumental in developing such services and demonstrating the value of the repository, which we hope will increase engagement and use.

Process

To assign FAST subject terminology to ETDs without performing record-by-record analysis, we first began with the subject metadata provided by ProQuest; specifically, the "subject" terms from their in-house controlled vocabulary, given the great irregularities present in the student-supplied "keywords." After loading these terms into OpenRefine, we then used a FAST reconciliation service[21] to reconcile the ProQuest subject terms against FAST. While most terms had fairly high confidence matches, there were a few that required manual review. In this review, we determined that some ProQuest terms required two FAST terms; "Canadian History," for example, has no direct FAST equivalent, so we assigned the FAST terms "Canada" and "History" to that term.

Once we had the list of reconciled FAST terminology, we incorporated these terms into our existing workflows. Prior to the ETD remediation project, Atkins Library used an XSLT to transform incoming batches of ProQuest XML ETD records into MODS and remediate some of the metadata problems that are a noted characteristic of records received through ProQuest ETD Administrator workflows.[22] The XSLT mapped ProQuest XML elements such as title, thesis author, and advisor to equivalent title and name elements in MODS. Student-supplied keywords, meanwhile, were crosswalked to a MODS note element rather than to MODS subject elements, a step taken in order to provide some form of subject access and yet prevent Niner Commons

[19] "Search Results," Niner Commons, accessed July 23, 2023, https://ninercommons.charlotte.edu/islandora/search?type=dismax&islandora_solr_search_navigation=0&f%5B0%5D=mods_relatedItem_host_titleInfo_title_ms%3A%22UNC%5C%20Charlotte%5C%20electronic%5C%20theses%5C%20and%5C%20dissertations%22&f%5B1%5D=mods_name_personal_author_affiliation_ms%3A%22Applied%5C%20Physics%22.

[20] "Search Results," Niner Commons, accessed July 29, 2023, https://ninercommons.charlotte.edu/islandora/search/?type=dismax&islandora_solr_search_navigation=0&f[0]=mods_relatedItem_host_titleInfo_title_ms:%22UNC\%20Charlotte\%20electronic\%20theses\%20and\%20dissertations%22.

[21] Ryan Johnson, "Remerjohnson/Fast-Reconcile, " Python (2021). https://github.com/remerjohnson/fast-reconcile.

[22] Veve, "ETDs in ProQuest and the Institutional Repository," 8.

subject facet displays from combining controlled FAST subject terms from non-ETD repository records with uncontrolled keywords of wildly varying quality in ETD records. Additionally, terms from ProQuest's own subject vocabulary for ETDs were mapped to another MODS note element and displayed in a separate field in Niner Commons' public interface.

In an effort to minimize capitalization irregularities in ProQuest XML records, where ETD titles and student-supplied keywords are erratically capitalized, the XSLT for incoming ProQuest records capitalized all titles and all keywords in Niner Commons ETD records. The belief at the time was that capitalization of title and keyword fields in all ETD records was preferable to inconsistent capitalization in such fields from record to record. Subsequently, however, Atkins staff came to see camelCase displays of keyword and title data as more intelligible to users, partly as a result of a review of literature on best practices for metadata displays.[23] The absence of controlled subject vocabulary was, of course, an even more serious liability.

In order to address the subject heading and capitalization issues, a suite of two remediation XSLTs was developed for the remediation project.[24] The first XSLT inserted one or more FAST subject terms into the legacy ETD records based on the ProQuest subject terms already present in the metadata, addressed the capitalization issues, and inserted administrative metadata that documented the remediation actions taken and the remediation date. To create it, staff used the templating function in OpenRefine to map the spreadsheet data containing FAST subject terms matched against the ProQuest terms in the reconciliation procedure to blocks of XSL "variable" elements. The ProQuest vocabulary subject terms in the legacy records were similarly mapped to clusters of XSL "if" elements using the same OpenRefine functionality. The "transpose columns" function in OpenRefine was crucial to this procedure.

Next, the clusters of XSL "variable" and "if" XSL elements were exported from OpenRefine in XML format and dropped into an XSLT document that contained ad-ditional templates for adjusting capitalization and creating administrative metadata. The XSLT was constructed in such a way that when it encountered a specific ProQuest subject term in an ETD record, it applied one or more matched ProQuest subject terms and their uniform resource identifiers in new MODS subject elements, as well as smoothed out capitalization and other style issues. The original ProQuest subject terms were retained in the legacy records. During tests, staff discovered that the XSLT was applying duplicate FAST subject terms to some ETD records. Rather than attempt to address this issue in the first XSLT, staff built a second stylesheet that stripped out any duplicate headings applied during the first transformation.

To apply the XSLTs, staff downloaded the legacy ETD records from Niner Commons using a CRUD (Create, Read, Update, Delete) app in the Islandora repository platform and moved them into Oxygen XML Editor project folders on a local computer. An Oxygen transformation scenario was created that applied the two XSLTs sequentially to 2,640 legacy ETD records in a single batch process. Requiring some 12 hours to complete, the transformation would doubtless have finished sooner if a more powerful computer had been used. Following spot checks of the transformed records, some manual edits were made with find and replace to address lingering capitalization irregularities, a process also described in an account of an ETD remediation process at the University of Iowa Libraries.[25] Like the authors of that study, Atkins staff hope to craft a more automated solution for normalizing capitalization in future XSLTs.

Due to a problem with the CRUD app that interfered with replacing the Niner Commons legacy ETD records with the transformed versions, staff enlisted the help of an Atkins developer to reingest the files. The ETD collection in Niner Commons was then reindexed so that the new FAST subject terms would display properly. After the remediation procedure, all that remained to be done was an extensive revision of the XSLTs for incoming ProQuest ETD records so that the same group of FAST subject terms would be applied to all future ETD

[23] See, for instance, Pragya Srivastava and Ms. Vinita Sharma, "Best Practices of UI Elements Design," *International Research Journal of Engineering and Technology* 6, issue 6 (2019) and Quovantis, "Why Letter Casing Is Important To Consider During Design Decisions," UX Planet, June 25, 2018, https://uxplanet.org/why-letter-casing-is-important-to-consider-during-design-decisions-50402acd0a4e.

[24] Joseph Nicholson, "ProQuest2FAST1," Github, 2021, accessed January 30, 2023, https://github.com/SedizioseVoci/XSLTs/tree/master/ProQuest2FAST_XSLT_1.

[25] Averkamp and Lee, "Repurposing ProQuest Metadata for Batch Ingesting ETDs into an Institutional Repository."

records from ProQuest as they were transformed into MODS and ingested.[26] Since the remediation project, all newly arriving ETD records have received one or more FAST subject terms upon ingest.

For all new receipts of ProQuest ETD records, staff coordinate closely with Atkins developers, who now apply the ingest XSLTs within the Islandora system. Once the ingest XSLTs have been run, staff spot check the records and run additional diagnostic XSLTs devised since the completion of the remediation project to identify records that were not assigned a FAST subject heading and those that have been assigned inappropriate headings during the transformation. After the ETD records are loaded into a test collection and additional quality control spot checks are performed, they are ingested in the ETD collection in Niner Commons.

Results

In the year since its implementation in April 2022, this process for normalizing capitalization and appending FAST terms to ProQuest ETD metadata has worked well, integrating seamlessly with existing workflows and reliably producing accurate, quality metadata. We have run the process several times as part of batch ETD ingests without issue. Conceivably, as more ETDs come in on novel topics, there may be new ProQuest subject terms to reconcile against FAST, which will require us to update the corresponding XSLTs. Relatively speaking, however, maintaining this process has not been especially time consuming or a burden in our ETD workflow.

One limitation of Atkins Library's remediation project was its narrow focus on a small handful of metadata problems that staff had identified as particularly crucial for retrieval and use of the ETD collection in Niner Commons. Unlike a more ambitious remediation effort at the University of Houston Libraries,[27] which was launched in order to bring ETD metadata into harmony with revised metadata guidelines for records contributed to a statewide ETD repository in Texas, staff at Atkins

Library did not attempt to standardize or control names of authors, advisors, or thesis committee members. Authority control measures like these have been identified as important for digital collections by both Waugh et al.[28] and McCutcheon.[29] Nor did the remediation project address diacritics problems in abstracts or title fields, which have been mostly handled on a record-by-record basis in Niner Commons, or seek to remediate or entirely remove the most flawed student-supplied keywords. Yet the relatively small-scale remediation actions performed in Atkins Library's project certainly do not preclude more extensive remediation work later. One benefit of the project's modest dimensions is that they allowed staff to test out remediation techniques on a smaller scale that can later be applied much more broadly in the repository. A second, more ambitious remediation effort that will address such issues as authority control is currently in the planning stages.

Another limitation of the project was that it did not attempt to apply the improvements made to Niner Commons ETD records in MODS to the corresponding MARC records for ETDs in Atkins's catalog (also received from ProQuest and then locally enhanced) or resolve the discrepancies in metadata quality that have resulted from Atkins' habit of creating and managing two sets of ETD records in different systems, one derived from student-supplied metadata and the other created by catalogers. Described by Rebecca Lubas as "double deposit,"[30] this commonplace practice in academic libraries can involve not only duplicative metadata management work for the same resources by staff in different units, but also records that do not share the same controlled access points or level of detail. At Atkins, double deposit in two linked but separate systems with different functionalities has made it difficult to ensure that changes to one group of records are mirrored in those in the other system. Though harmonizing separate ETD metadata management practices in MARC and MODS at Atkins could over the long term help reduce

[26] Joseph Nicholson, "FINAL_ProQuest_XML_to_MODS_XSLT_Troika," Github, 2022, accessed January 30, 2023, https://github.com/SedizioseVoci/XSLTs/tree/master/FINAL_ProQuest_XML_to_MODS_XSLT_Troika.

[27] Thompson, Liu, Duran, and Washington, "A Case Study of ETD Metadata Remediation at the University of Houston Libraries," 62.

[28] Laura Waugh, Hannah Tarver, and Mark Edward Phillips, "Introducing Name Authority into an ETD Collection," *Library Management* 35, no. 4/5 (2014): 273.

[29] Sevim McCutheon, "Basic, Fuller, Fullest: Treatment Options for Electronic Theses and Dissertations," *Library Collections, Acquisitions, & Technical Services* 35 (2011): 65.

[30] Rebecca L. Lubas, "Defining Best Practices in Electronic Thesis and Dissertation Metadata," *Journal of Library Metadata* 9, issue 3-4 (2009): 253.

some of the ETD metadata flaws that Atkins' remediation project in Niner Commons was designed to address, the ambitious effort of restructuring ETD workflows at Atkins would require more staff and resources than the library currently possesses and was therefore beyond the scope of this effort.

As to the quality of the reconciled FAST metadata that we now append to the ETDs, as we drew from the existing ProQuest subject metadata, the words we reconciled are very general, covering disciplines or areas of study like "environmental science" and "adult education." As we do not have the staffing bandwidth for record-by-record analysis, this approach was a matter of necessity. In addition to being more general, this mode of subject description is more diffuse. Essentially, the terms are a translation of existing terminology instead of a result of direct analysis, which could potentially cause the description to be more blurred or imprecise. While we review the reconciled FAST terms against their ProQuest originals in a spreadsheet, we do not look at each ETD to ensure their reconciled FAST terms are perfect fits (aside from select spot checking with each batch ingest of ETDs into the repository). While this reliance on batch processes and more general subject terminology may be more lax, we have found the resultant ETD metadata to be more or less in line with the descriptive records for other works within the Niner Commons repository. Currently there is only one staff member responsible for ingesting works into Niner Commons and creating the corresponding metadata records, so as a matter of staff capacity each record receives two or three FAST terms. Accordingly, while this approach works for our cataloging needs, it may be too broad for institutions looking for more granular subject coverage.

An unexpected yet important consequence of this remediation project was that it highlighted DEIA (diversity, equity, inclusion, and accessibility) issues within our subject metadata. In particular, in running the FAST reconciliation service in OpenRefine and reviewing the results, we saw that the FAST equivalents of several of the ProQuest terms were offensive, problematic, and outdated. The reconciliation service had recommended "Oriental literature" for "Asian literature," for example, and "Sexual minorities" for "LGBTQ studies." Accordingly, this ETD remediation project was in part the impetus for a subsequent metadata initiative, in which we audited FAST metadata within the repository at large to identify and replace offensive terms. This initiative is in progress, as we continue to evaluate terms and develop cataloging guidelines that will help us be more inclusive and respectful of our users.

Finally, the remediation project was hampered by deteriorating functionalities of the Islandora platform that supports Niner Commons, which is currently running on an older, unsupported version. Unable to make use of the CRUD app to reingest the remediated metadata files through the Islandora interface, staff had to ask Atkins developers to replace the records through a command line procedure on the backend, a step that will be necessary for any future remediation actions. Staff have since received training in replacing files through the command line themselves, but the procedure remains a cumbersome workaround. These difficulties are a reminder of how repository system weaknesses, just as much as staffing and skill constraints, can negatively impact the scope and ease of a metadata remediation project. Atkins staff are presently exploring new repository platform options, with a migration tentatively scheduled to take place within the next year.

Conclusion

Atkins's ETD subject metadata remediation project has improved discovery within the repository, facilitating better collocation, browse, and cross-repository searching. Limited to capitalization and subject metadata, this remediation effort acknowledges staff constraints both by being targeted in scope and by utilizing batch tools and methods. Though ETD metadata workflows can vary by university and can be especially tricky, with metadata often coming from different sources and relying on user-submitted information, Atkins Library has found success with small-scale, sustainable remediation projects. For libraries lacking extensive repository or cataloging staff, project-based remediation efforts that yield integrated changes in cataloging workflows could be a useful strategy for continually improving the metadata quality of ETDs and other works.