The Journal of Electronic Theses and Dissertations

Volume 2 Article 5

2022

Open Science ETDs and Institutional Repositories: Making Research Data FAIRer

Andrew Mckenna-Foster
Figshare, andrew@figshare.com

Maria Cotera
Figshare, maria@figshare.com

Mark Hahnel
Figshare, mark@figshare.com

Open Science ETDs and Institutional Repositories: Making Research Data FAIRer

Cover Page Footnote

Correspondence concerning this article should be addressed to Andrew Mckenna-Foster, email: andrew@figshare.com

Open Science ETDs and Institutional Repositories: Making Research Data FAIRer

Andrew Mckenna-Foster[1], Maria Cotera[2], and Mark Hahnel[3]

[1]Figshare, ORCiD: 0000-0001-7641-6588

[2]Figshare, ORCiD: 0000-0002-1375-0738

[3]Figshare, ORCiD: 0000-0003-4741-0309

Figshare, 4 Crinan St. London, N1 9XW, UK

Author Note

Correspondence concerning this article should be addressed to Andrew Mckenna-Foster, email: **andrew@figshare.com**

Abstract

Graduate students, as potential future full-time researchers, are a population that should show proficiency in data sharing. Though there are many resources that teach data sharing best practices for students, it is difficult to tell how well students do when sharing their data. We compared the FAIRness of non-traditional research output metadata associated with theses and dissertations for records shared in a generalist repository by individual students, and records shared through an institutional repository using the same repository platform. Those shared in an institutional repository were significantly FAIRer, as measured by metadata richness and interoperability, and had higher views per month. The only measure where records shared by students exceed institutional records is listing funding sources. We also examine how multiple related research outputs are grouped and offer suggestions to improve interoperability. We conclude that our sample population of graduate students sharing research outputs are not yet proficient in applying the FAIR principles. The review process and oversight that are often part of institutional repositories can offer a measurable benefit to non-traditional ETD outputs.

Keywords: electronic thesis and dissertation, dataset, FAIR data, graduate student, research outputs

Introduction

The publication and sharing of research data is beneficial to all because it provides credibility to research and increases reproducibility, and it is directly beneficial to authors because it increases the reuse and citation of a related paper (Colavizza et al., 2020). Graduate students are at the beginning of their careers and may particularly benefit from the practice of sharing data (Azevedo et al., 2022; Nelson & Kong 2020). They also may be the best population on which to focus training to maximally grow open data practices and enmesh them in standard practice. One could argue that if today's graduate students learn data sharing skills and philosophies as part of their course work, they would bring those skills into academia as they begin their careers, and FAIR data sharing would become standard practice faster. Unfortunately, Azevedo et al. (2022) write that while open principles are increasingly practiced in research, they are not widely taught in graduate programs as part of traditional course work. Training programs and resources do exist, for students and researchers - and often take the form of online resources, grant funded programs, trainings through university libraries, and online short courses (e.g., Adamick et al., 2012; Bailey 2021; Kathawalla et al., 2021; Roberts-Pierel et al., 2021). They generally address the benefits of data sharing as part of open science and reproducible research and provide the basic skills to share data responsibly. There is also growing support for those who train students (and researchers) in data sharing, notably the Framework for Open and Reproducible Research Training (Azevado et al., 2022) and reference books such as Carlson and Johnston's (2015) *Data Information Literacy: Librarians, Data, and the Education of a New Generation of Researchers* and Clare et al.'s (2019) *Engaging Researchers with Data Management: The Cookbook.*

Academic libraries and their services are positioned to play a key role in training students to share data appropriately. Libraries often provide support services and infrastructure in the form of scholarly communication and data librarians, online data sharing guides, and institutional repositories (Briney et al., 2015; Cox et al., 2019). Data-capable, institutional repositories can ease data sharing workflows for researchers because they offer a low barrier data sharing location that usually comes with more support than alternative generalist repositories. Librarians can help check the FAIRness of the repository records and help maintain the published records over time. However, not all libraries can offer these resources. Some institutions have repositories which cannot handle large data files or provide the functionality that a researcher may require to comply with funder or publisher requirements.

Furthermore, despite these resources, there is evidence that graduate students lack data management skills (Sharma & Qin 2014; Wiley & Kerby 2018), that they must still learn on their own rather than from faculty mentors (Pasek & Mayer 2019) and that they do not have the data management and sharing skills the students themselves think are important (Rantasaari 2021).

Graduate students are not alone in this struggle. While the concept of research data sharing, and increasingly, FAIR data (Wilkinson et al., 2016), is now widespread within the scholarly communication milieu, putting the concepts into practice is a difficult task even for seasoned researchers. Data can be difficult to obtain even if there is a data sharing statement and if they are available and they are often not formatted in a way conducive to reuse (e.g. Cooper 2021; Couture et al., 2018; Gabelica et al., 2022; Valentin Danchev et al., 2021; Van Tuyl & Whitmire 2016).

The online training and help sites, grant funded programs, data librarians, and institutional repositories are a growing patchwork of data sharing resources. Taken as a whole,

we would expect more and more graduate students to be familiar with data sharing skills as the resources have expanded. We would also expect that students who receive help publishing data through their institution may benefit from librarian expertise and their outputs should be FAIRer than those shared without assistance.

In this paper, we examine how well graduate students are adopting best practices when sharing data and non-traditional research outputs (NTROs, e.g. figures, media, and other supporting materials) related to an electronic thesis or dissertation (ETD), and if there is any evidence that practices are improving. We assess their data sharing skills by checking for FAIRness of theses and dissertation related objects on a generalist repository platform and compare them directly to objects shared through institutional repositories using the same platform. We make the assumption that the latter are shared with the help of librarians and / or research data management professionals. Our analysis indicates the importance of institutional repositories and related services and also lays out a method for tracking student skills over time.

Methods

The Figshare repository platform is a free generalist repository platform, open to all with no curation where researchers can publish their ETD outputs. Figshare also provides infrastructure for research organizations to publish their ETDs and related outputs and provide curation and review checks (*figshare for institutions—A repository for research data of all types* (n.d.))**.** This uniquely provides the opportunity to study records that were most likely published directly by graduate students (hereafter referred to as figshare.com records) to those published through an institution (hereafter referred to as institutional records). An openly documented API

enables the upload and download of metadata across all Figshare repositories making it relatively easy to harvest information at scale.

We assessed metadata for two types of digital objects on the Figshare platform: records which hold zero to many files with accompanying metadata, and Collections which aggregate records under unifying metadata. Data collection involved two metadata harvesting runs using the Figshare API (htttps://docs.figshare.com) each searching all metadata fields for "thesis OR dissertation." One run collected up to 1,000 metadata records each for three record types: datasets, figures, and media (https://docs.figshare.com/#articles_search). We also collected the number of views for each record using a statistics API endpoint (https://docs.figshare.com/#stats_totals_endpoint_format). The second run collected metadata from as many Figshare Collections as possible. Collecting metadata for many Collections was necessary because we could not pre-emptively exclude results that mentioned theses or dissertations in the metadata, for example when listing sources for a review paper.

We programmatically and manually checked the sample to include only records published from academic repositories or from individual researchers. We manually removed records and Collections that were not directly related to an ETD. We assumed that records and Collections from figshare.com are very likely published directly by graduate students (rather than mediated by a library professional); and that those from an institutional repository went through some level of curation or metadata enhancement.

We evaluated records against components in three of the FAIR principles (Wilkinson et al., 2016) - Findable: Data are described with rich metadata; Interoperable: (Meta)data include qualified references to other (meta)data; Reusable: (Meta)data are richly described with a plurality of accurate and relevant attributes. We quantified the richness of metadata using proxy

measures: we counted the number of words in the title and description, counted keywords, and counted categories. We also assessed findability by dividing a record's view count by the number of months since the record was posted to obtain a views per month measure. To quantify interoperability, we counted the number of reference links either in the reference metadata field or in the description field and counted links to funder records. Because all measures are count data, we used a negative binomial model to examine the relationship between assessed values and the repository type (figshare.com versus an institutional repository). To account for multiple tests, Bonferroni corrected the significance level to 0.006 (0.05/8). Data and code are available in the Figshare repository (Mckenna-Foster, 2021).

We assessed Collections as extensions to records rather than standalone records themselves. Collection metadata describes a set of records and augments the record level metadata with more context. For example, a student can create separate, citable records for their ETD, their ETD data, slides and recordings of related presentations, and any other supplementary information, and then aggregate those records in a Collection under a unifying set of metadata and DOI. Ideally, each record would contain a DOI link referring to the Collection. In addition, the Collection can contain links to related resources (e.g. data published in a disciplinary repository) and, if a copy of the ETD is not present as a record, then a link to the ETD. For each Collection, we looked for the related ETD or published papers and documented what types of records were shared in the Collection. We also looked at each record in the Collection for a link pointing back to the Collection.

Results

We collected a total of 2,606 records and 9,000 Collections through the Figshare API (Table 1). Cleaning the sample left 709 records (428 from figshare.com and 281 from 33 institutions) and 46 Collections (19 from figshare.com and 27 from 16 institutions).

Table 1.

Record and Collection Sample Summary.

Object	Date collected	Initial search results	Final sample set	Institutional repositories represented
Record	9 Sept 2021	2,606	709	33
Collection	3 June 2021	9,000	46	16

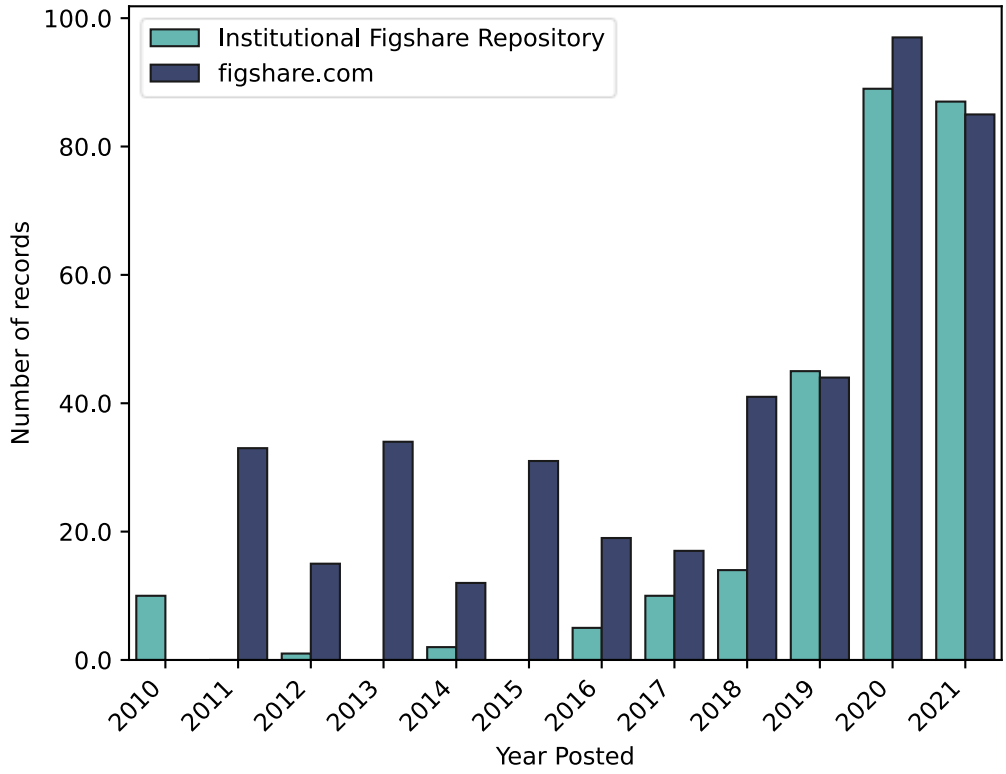
Record Results

The record dataset includes 458 datasets (65%), 128 figures (18%), and 123 media (17%) records. The distribution of types was significantly different with more datasets and figures from figshare.com and more media records from institutional repositories ((2, N=709) = 33.24, p-value<.0001). The 709 records represent 434 unique first authors with over 50% of authors represented by one record. One author is represented by 44 figshare.com records and all records have sparse metadata. The distribution in record posting date is relatively similar with most records for both figshare.com and institutions posted after 2018 (Fig. 1). Forty-six percent of records are related to doctoral work and another 46% could not be identified with a specific degree. Six percent are related to a master's degree and two percent are related to a bachelor's

degree. We did not remove bachelor's degree related records from the sample. Only four records contained non-English text in the title or description and all four also contained English text.

Figure 1.

The Number of Records in the Sample Dataset Published Each Year on figshare.com and in Institutional Repositories using the Figshare Platform.



Note. Institutions began using the Figshare platform in 2015. Records from prior dates exist because posted dates can be migrated with records from another platform.

Records from institutions are FAIRer with significantly longer titles, significantly longer descriptions, significantly more references, and significantly more keywords (Fig. 2). Compared to figshare.com records, institutional records had 1.81 times the number of words in the title

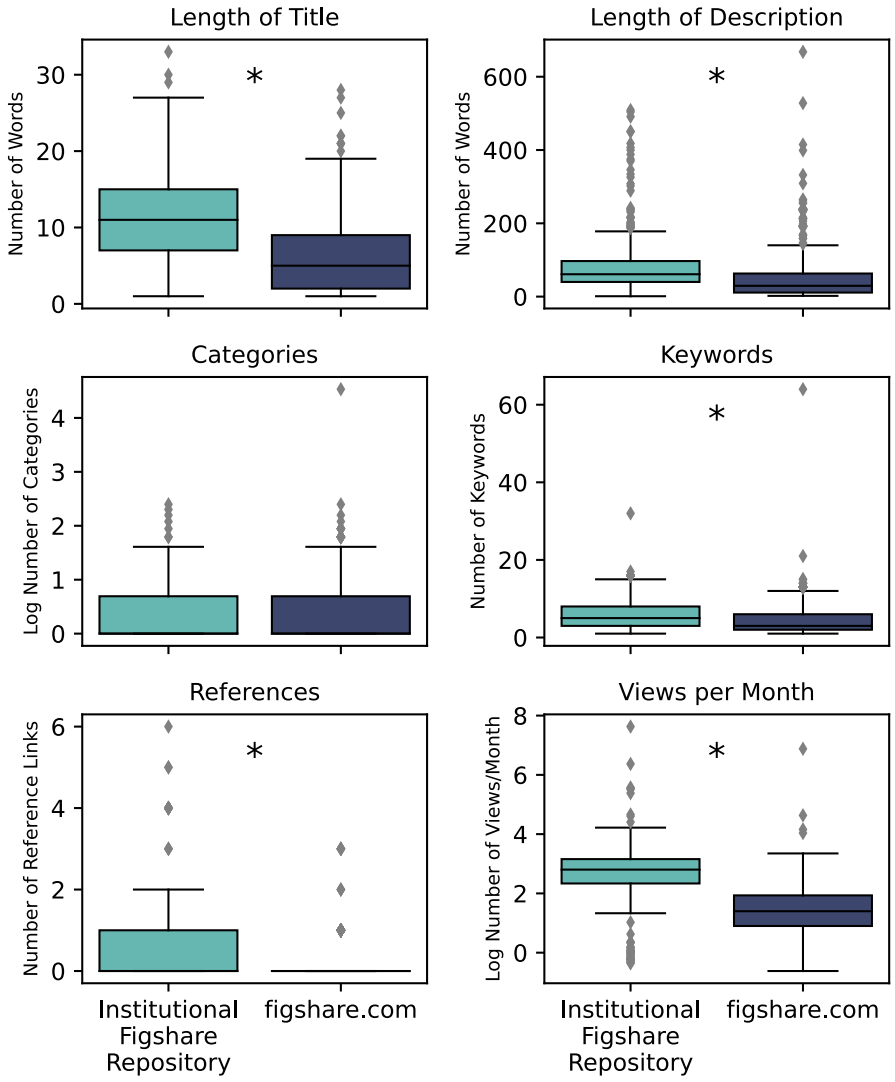
(95% confidence interval [CI], 1.64 to 2.00; P <.001), 1.56 times the number of words in the description (95% CI, 1.33 to 1.82; P < .001), 3.60 times the number of references (95% CI, 2.68 to 4.87; P < .001), and 1.31 times the number of keywords (95% CI, 1.17 to 1.47; P < .001). The number of categories was not significantly different. Institutional records had 3.85 times the number of views per month, and this was significant (95% CI, 3.25 to 4.57; P < .001). Removing the author with 44 records from the analysis did not affect the significance of any differences we observed. Seventy-two percent of figshare.com records had no links in the description or references field, while only 54% of institutional records lacked a link in either field.

The one area where figshare.com records compare to and even surpass institutional records is listing and linking to funding sources. Figshare metadata includes a field that accepts free text grant information or enables the user to select a grant from a list sourced from the dimensions.ai database. In the latter case, the field automatically inserts a link to the grant record. Eighty-two figshare.com records (19%) list at least one grant compared to 78 institutional records (28%). The 82 figshare.com records list a total of 202 grants and 14 (7%) of those link to a grant record in dimensions.ai. Institutional records list 148 grants and only one (0.7%) is linked.

The difference in quality of record metadata between institutional repositories and individual users is visible overtime and shows signs of widening for some measures in recent years. Records in institutional repositories show longer titles and more references than those shared by individuals on figshare.com since 2017 (Fig. 3). Institutional records more consistently have longer descriptions and more keywords over time. No consistent trends are visible for figshare.com records.

Figure 2.

Comparisons of Metadata Quality for Records from Figshare.com and Institutional Repositories.



Note. Data for number of categories and views per month is displayed as the natural log of the value to show the data distribution more clearly. The horizontal line within each box is the median, the upper and lower box ends represent 75% and 25% of the data respectively, and the vertical lines indicate the highest and lowest values up to 1.5 times the height of the box. Points outside that range are outliers and are represented as diamonds. *p<.006

The difference in quality of record metadata between institutional repositories and individual users is visible overtime and shows signs of widening for some measures in recent years. Records in institutional repositories show longer titles and more references than those shared by individuals on figshare.com since 2017 (Fig. 3). Institutional records more consistently have longer descriptions and more keywords over time. No consistent trends are visible for figshare.com records.

It is difficult to assess the use of Collections from record metadata because record metadata does not automatically indicate membership in a Collection. However, by comparing the list of ETD related records with the list of records sourced from our sample of Collections, we were able to identify fifty-seven records (8%, 40 from figshare.com and 17 from institutions) that are members in a Collection. Other records in our sample may be included in Collections that were not part of our Collections sample, thus 8% is a conservative estimate.

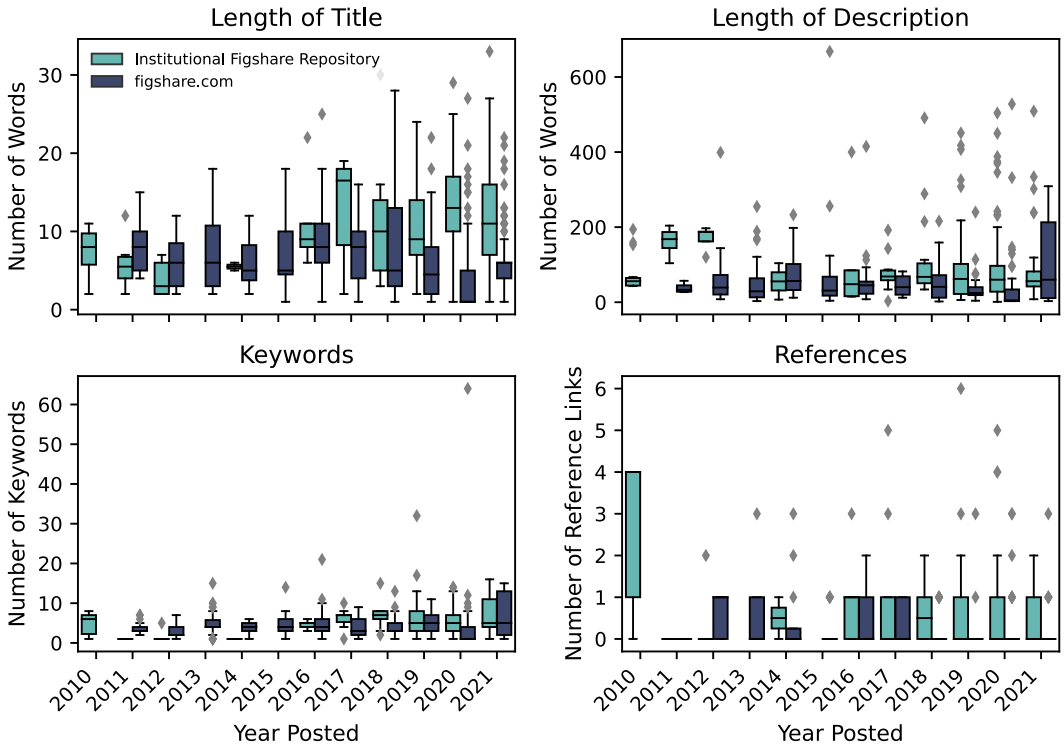
Collection results

The 46 Collections contained 407 records. Datasets, code, media, and figures are found in 83% of Collections with presentations, geospatial files, links to other data, and supplementary files making up the rest. A slight majority of Collections (59%) do not contain the ETD and do not reference the ETD or a published paper. Only 13% contained the ETD itself, but 41% contained or linked to a document with more information about the research. Nine percent link to a peer reviewed paper. Only nine percent of Collections contain records that link back to the Collection. Collections shared in an institutional repository were not more or less likely to contain an ETD than those shared on figshare.com ($\chi^2(1, N=46) =0.199$, $p=.655$). Linking to published papers and linking to funders were equal for institutional Collections and Collections

on figshare.com. Two of the figshare.com Collections and two of the institutional Collections linked to a published paper related to the ETD. Three fighsare.com Collections listed funding sources and three institutional Collections listed funding sources.

Figure 3.

Quality of Metadata Over Time for Figshare.com Records and Institutional Repository Records.



Discussion

ETD-related NTROs shared in repositories with institutional oversight are more findable, interoperable, and reusable than those shared without institutional oversight. They have richer metadata, and except for listing and linking to funders, are better linked to related records. These

results indicate that graduate student mastery of FAIR sharing is still limited and that they likely benefit from the services attached to institutional repositories. Many repositories are managed by librarians; we assume their oversight may be the main reason for the increased FAIRness of ETD research outputs associated with an institution. Librarians can ensure metadata completion before and after publication and apply appropriate discovery terms, ensuring a higher chance for reuse of a student's work.

The higher use of the linked funding option in figshare.com records was unexpected. It may be due to the migration of legacy records from a former institutional repository platform into the Figshare based platform. Linking grants requires manual work for every record and is possibly too time consuming for a repository manager to attempt.

Our analysis of metadata quality over time (Fig. 3) shows that there is a lot of variability for both figshare.com and institutional records in each of the four measures. The description field in particular shows variation, possibly because authors may paste in the abstract or a figure caption from their thesis or dissertation while others only add a few words of explanation. Importantly, there is no indication that ETD related metadata on figshare.com is improving. It is difficult to tell how generalizable this is. Our sample may be too small to capture changes over time, metadata practices may be improving more visibly in other generalist repository platforms and not on figshare.com, or there may in fact be no overall improvement. One indication that the latter may be true is that institutional records do show increases in title length and more consistent use of the references metadata field. This is not to say that data sharing training and resources are not effective, it may be that it is taking time for the data sharing ethic, and the skills associated with it, to be accepted across research disciplines and across career stages. Graduate students are certainly not alone in having trouble completing metadata. Studies regularly find

that shared research data often has incomplete or low-quality metadata (e.g. Gonçalves & Musen 2019; Van Tuyl 2019; Van Tuyl & Whitmire 2016) Harvesting metadata to specifically examine changes over time, perhaps including data from other generalist repository platforms, would be a useful next step.

The higher number of views per month for institutional repository records indicate a measurable benefit to students. We cannot determine if the increased number of views is due to the higher quality metadata or if the views result from the association with an institution. Any record shared in a Figshare based repository is searchable within the figshare.com search interface and is one of more than six million objects as of August 2022. An institution's repository landing page and search interface provides access to a subset of those records and increases the chances a record will be viewed by a user. Users most often arrive at repository records directly from search engine results (Coates, 2014; St. Jean et al., 2011) but data from Coates (2014) suggest that about one third of users may be finding records from repository landing pages. Ultimately, our results suggest that there is a significant visibility benefit for students who share ETD outputs in an institutional repository.

Figshare Collections offer a natural way to group related records of many different file types, a difficulty surfaced by both Collie and Witt (2011) and Van Tuyl (2019). This could be especially important for graduate students who may want to showcase in one place datasets, code, presentations, and posters related to their work. Those outputs may be the first significant contributions they make to their field and being able to reference them as a set with a DOI could be useful. However, ideal use of Collections requires an author or curator to manually add the Collection DOI to each member record, just as they would need to manually add links to other related outputs. Collections do not do this automatically because they can contain records that an

author does not own, thus providing a way for authors to use datasets published by others to cite all the works together and individually. Manually adding a link to the Collection would not make sense for all Collection use cases but does make sense when an author is specifically creating a Collection to group related records. In our sample, linking between records and the Collection was rare for both institutional and figshare.com records. It is therefore difficult to know if a record is part of a Collection, difficult to navigate from a record to its parent Collection, and difficult to find related records. As of August 2022, Collections make up about 20% of the records available in the Figshare search interface (https://figshare.com/search, 1,242,554/6,214,118). Although our sample of Collections is very small and specific to graduate work, it may indicate that authors in general do not think to link their own records to their own Collections when it might make sense to do so.

Strengths and Weaknesses

This study's strengths lie in its quantitative nature, its reproducibility, and replicability. However, the record data are analyzed in aggregate and may lack important context because, aside from confirming that a record was related to an ETD, we were not able to look in detail at every record or at any of the associated research objects. Our assessment of metadata quality was not comprehensive and only addressed parts of the FAIR principles. There are two variables that we could not control and may be important. We could not tell if figshare.com records were published with the assistance of an institution curator. Many institutions list generalist repository options on their data sharing resource pages or LibGuides and suggest authors get in touch with the library for general assistance. If many figshare.com records were prepared in this way with the assistance of a curator, our assumption that curatorial assistance in institutional repositories

explains the metadata quality differences may be incorrect. One or more other factors may explain the differences. For example, there is the possibility that institutional records may be dominated by records migrated from another repository platform so the metadata was not entered through the Figshare interface and may not be an ideal comparison. Another variable we could not control is the language of the record's author. Figshare.com is a free service to researchers around the globe with an interface and help pages in English; while we only found four records in our sample where another language appeared in the title or description in addition to English, we assume that non-native English students may also be using the platform to share research outputs, and therefore there may be many other records written by authors with lower English proficiency. Authors sharing through an institutional repository, even in a non-English speaking country, may have more resources to help translate material, assist with using the platform, and understand data sharing best practices. In a similar vein, we acknowledge that this paper has a North American viewpoint, especially related to the review of data sharing resources for graduate students. It should be viewed as a starting point for further investigation and evaluation of graduate student data sharing.

Conclusion

We determined that there are significant differences in the quality of some metadata fields between ETD non-traditional research outputs (NTROs) shared on figshare.com - presumably directly by graduate students - and ETD NTROs shared through an institutional repository - presumably mediated by a library or repository professional. Records from institutions also had higher views per month. The results indicate that institutional repositories help students increase both the FAIRness and the visibility of their research outputs. The

Collection object type in the Figshare platform is a good way for students to group and showcase their outputs; however, many Collections and their member records, both from figshare.com and from institutions, lacked appropriate linking to the related ETD and related objects, making it difficult for end users to navigate between objects.

This is a relatively small study and indicates that an analysis of more records using a more granular metadata quality assessment method, such as that used by Quarati and Raffaghelli (2020), could offer deeper insights in graduate student NTRO sharing. At the very least, our methods can be replicated over time to take snapshot assessments. As data sharing continues to be accepted as a standard component of scholarly communication, it will remain important to assess student and early researcher skills.

This study was conducted by Figshare researchers, and the results led the authors to actively explore ways we can help all repository users make research data FAIRer. We updated help articles to include specific guidance for students and curators sharing outputs related to ETDs to include information on completing metadata, using Collections, and linking objects. We also created a specific help page for graduate students.

References

Adamick, J., Reznik-Zellen, R., & Sheridan, M. (2012). Data management training for graduate students at a large research university. *Journal of EScience Librarianship*, *1*(3), 180–188. https://doi.org/10.7191/jeslib.2012.1022

Azevedo, F., Liu, M., Pennington, C. R., Pownall, M., Evans, T. R., Parsons, S., Elsherif, M. M., Micheli, L., Westwood, S. J., & Framework for Open, Reproducible Research Training (FORRT). (2022). Towards a culture of open scholarship: The role of pedagogical

communities. *BMC Research Notes*, *15*(1), 75. https://doi.org/10.1186/s13104-022-05944-1

Bailey, Charles W., Jr. (2021). *Research Data Curation and Management Bibliography*. Houston: Digital Scholarship. http://digital-scholarship.org/rdcmb/rdcmb.htm

Briney, K., Goben, A., & Zilinski, L. (2015). Do you have an institutional data policy? A review of the current landscape of library data services and institutional data policies. *Journal of Librarianship and Scholarly Communication*, *3*(2), eP1232. https://doi.org/10.7710/2162-3309.1232

Carlson, J., & Johnston, L. R. (Eds.). (2015). *Data Information Literacy: Librarians, Data, and the Education of a New Generation of Researchers*. Purdue University Press. https://www.jstor.org/stable/j.ctt6wq2vh

Clare, C., Cruz, M., Papadopoulou, E., Savage, J., Teperek, M., Wang, Y., Witkowska, I., & Yeomans, and J. (2019). *Engaging Researchers with Data Management: The Cookbook*. Open Book Publishers. https://doi.org/10.11647/OBP.0185

Coates, M. (2014). Search engine queries used to locate electronic theses and dissertations: Differences for local and non-local users. *Library Hi Tech*, *32*(4), 667–686. https://doi.org/10.1108/LHT-02-2014-0022

Colavizza, G., Hrynaszkiewicz, I., Staden, I., Whitaker, K., & McGillivray, B. (2020). The citation advantage of linking publications to research data. *PLOS ONE*, *15*(4), e0230416. https://doi.org/10.1371/journal.pone.0230416

Collie, W. A., & Witt, M. (2011). A Practice and Value Proposal for Doctoral Dissertation Data Curation. *International Journal of Digital Curation*, *6*(2), 165–175. https://doi.org/10.2218/ijdc.v6i2.194

Cooper, K. A. (2021). Data sharing attitudes and practices in the plant sciences: Results from a mixed method study. *Journal of Agricultural & Food Information*, *22*(1–2), 37–58. https://doi.org/10.1080/10496505.2021.1891923

Couture, J. L., Blake, R. E., McDonald, G., & Ward, C. L. (2018). A funder-imposed data publication requirement seldom inspired data sharing. *PLOS ONE*, *13*(7), e0199789. https://doi.org/10.1371/journal.pone.0199789

Cox, A. M., Kennan, M. A., Lyon, L., Pinfield, S., & Sbaffi, L. (2019). Maturing research data services and the transformation of academic libraries. *Journal of Documentation*, *75*(6), 1432–1462. https://doi.org/10.1108/JD-12-2018-0211

figshare for institutions—A repository for research data of all types. (n.d.). Retrieved March 30, 2022, from https://knowledge.figshare.com/institutions

Gabelica, M., Bojčić, R., & Puljak, L. (2022). Many researchers were not compliant with their published data sharing statement: Mixed-methods study. *Journal of Clinical Epidemiology*, *150*, 33–41. https://doi.org/10.1016/j.jclinepi.2022.05.019

Gonçalves, R. S., & Musen, M. A. (2019). The variable quality of metadata about biological samples used in biomedical experiments. *Scientific Data*, *6*, 190021. https://doi.org/10.1038/sdata.2019.21

Kathawalla, U.-K., Silverstein, P., & Syed, M. (2021). Easing into open science: A guide for graduate students and their advisors. *Collabra: Psychology*, *7*(1). 18684. https://doi.org/10.1525/collabra.18684

Mckenna-Foster, A. (2021). Data and code to compare metadata quality of Figshare records related to electronic theses and dissertations. figshare. https://doi.org/10.6084/m9.figshare.16766932

Nelson, M. S., & Kong, N. N. (2020). Capturing their "first" dataset: A graduate course to walk PhD students through the curation of their dissertation data. *IASSIST Quarterly*, *44*(3). https://doi.org/10.29173/iq971

Pasek, J. E., & Mayer, J. (2019). Education Needs in Research Data Management for Science-Based Disciplines: Self-Assessment Surveys of Graduate Students and Faculty at Two Public Universities. *Issues in Science and Technology Librarianship*, *92*. https://doi.org/10.29173/istl12

Quarati, A., & Raffaghelli, J. E. (2020). Do researchers use open research data? Exploring the relationships between usage trends and metadata quality across scientific disciplines from the Figshare case. *Journal of Information Science*, *48*(4), 423–448. https://doi.org/10.1177/0165551520961048

Rantasaari, J. (2021). Doctoral students' educational needs in research data management: Perceived importance and current competencies. *International Journal of Digital Curation*, *16*(1), 36. https://doi.org/10.2218/ijdc.v16i1.684

Roberts-Pierel, B., Davis, E., & Rao, Y. (2021). *A Graduate Student's Road Map to Data Management Training* (Version1). ESIP. https://doi.org/10.6084/m9.figshare.14384456.v1

Sharma, S., & Qin, J. (2014). Data management: Graduate student's awareness of practices and policies. *Proceedings of the American Society for Information Science and Technology*, *51*(1), 1–3. https://doi.org/10.1002/meet.2014.14505101130

St. Jean, B., Rieh, S. Y., Yakel, E., & Markey, K. (2011). Unheard voices: Institutional repository end-users. *College & Research Libraries*, *72*(1), 21-42. https://doi.org/10.5860/crl-71r1

Valentin Danchev, Dp., Min, Y., Borghi, J., Baiocchi, M., & Ioannidis, J. P. A. (2021). Evaluation of data sharing after implementation of the International Committee of Medical Journal Editors data sharing statement. *Health Policy*, *4*(1), e2033972. https://doi.org/10.1001/jamanetworkopen.2020.33972

Van Tuyl, S. (2019). What's in the Box? Assessing the potential usability of four decades of thesis and dissertation supplementary files. *Journal of EScience Librarianship*, *8*(1), e1142. https://doi.org/10.7191/jeslib.2019.1142

— & Whitmire, A. L. (2016). Water, water, everywhere: Defining and assessing data sharing in academia. *PLOS ONE*, *11*(2), e0147942. https://doi.org/10.1371/journal.pone.0147942

Wiley, C. A., & Kerby, E. (2018). Managing research data: Graduate student and postdoctoral researcher perspectives. *Issues in Science and Technology Librarianship*, *89*. https://doi.org/10.5062/F4FN14FJ

Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., … Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*(1), 160018. https://doi.org/10.1038/sdata.2016.18