FAIRising Pedagogical Documentation for the Research Lifecycle

Deborah A. Garwood (✉) and Alex H. Poole

Metadata Research Center, Drexel University, Philadelphia, PA 19104, USA
dgarwood@drexel.edu

Abstract. How can pedagogical research complement academic research and vice versa? This case study revisits two research projects on curricular resources: the first, in 2019, analyzes partially structured syllabi data in digital humanities, and the second, in 2021, focuses on unstructured course titles and descriptions in LIS course catalogs. Findings reexamine data collection and analysis processes in which the lack of linked semantic metadata and persistent digital objects in curricular resources impedes fruitful research on how (inter)disciplinary topics are taught and future researchers trained. Consequently, the case study locates a gap: the role of pedagogical documentation in the research lifecycle has not been considered. As suggested by the emergence of FAIR principles, metadata expertise is a foundation for establishing the findability, accessibility, interoperability, and reuse of persistent digital objects in research outputs. FAIRising pedagogical documentation for the research lifecycle holds potential to link curricular resources with other research outputs. Information professionals have a leadership role in assisting faculty to create FAIRised pedagogical documentation, and curricular resources so prepared address the gap for integrating pedagogical documentation with the research lifecycle. Benefits include recognition of curricular resources as vital research outputs and facilitating longitudinal research on (inter)disciplinary pedagogical practices in the FAIR ecosystem.

Keywords: Metadata · Linked semantic metadata · Pedagogical research · (Inter)disciplinary research · Persistent digital objects · FAIR digital objects · FAIR ecosystem

1 Introduction

How can pedagogical research complement academic research and vice versa? To explore the research question, this case study first considers literature on the research lifecycle and the role of metadata expertise. Second, background on metadata’s increasing utility for measures of integrity in scholarly publications adumbrates the FAIR principles’ inclusive approach to research outputs well beyond scholarly publications [1, 2]. The literature leaves a gap for exploring how curricular resources contribute to the research lifecycle, and how “FAIRised” [2, p. 8] pedagogical documentation aids scholarly analysis of (inter)disciplinary pedagogy practices.
The third section presents the case study’s reexamination of semi-automatic and manual data collection and analysis techniques for two pedagogical research projects, one investigating open source digital humanities syllabi, the other analyzing unstructured course titles and descriptions in Library & Information Science (LIS) course catalogs at American Library Association accredited US LIS programs. Fourth, a discussion section outlines the implications of findings. Fifth, the conclusion summarizes key points and recommends future research directions.

2 The Research Lifecycle, Metadata’s Role, and the FAIR Ecosystem

Information professionals whose metadata expertise assists faculty to prepare scholarly research for publication and data deposit perform a crucial step in the research lifecycle [3–5]. Teams of faculty and students launch academic research projects and often consult with information professionals to deploy metadata and build skills [3, 4, 6]. Information professionals – librarians, archivists, and data curators who often have overlapping technical and soft skills – are well-positioned to participate in research initiatives and guide the process of metadata creation throughout projects [7–9].

In short, pedagogy, research outputs, and metadata expertise are intertwined, yet the role of curricular resources in the research lifecycle has not been considered. Rather, metadata’s capacity to structure research outputs historically is associated with scholarly publications. The FAIR principles (Findability, Accessibility, Interoperability, and Reusability) reach far beyond scholarly publications to include such research outputs as software, workflows, algorithms – all components of the research lifecycle [1, 2]. Curricular resources are not yet recognized as research outputs, however.

Ensuring that all research outputs are reliable and persistent, machine actionable digital objects is the primary objective of the FAIR principles [2]. The holistic research culture FAIR envisions relies on networks of policies, data management plans, persistent identifiers, standards, and repositories, and the FAIR ecosystem is representative of stakeholders’ investment in the integrity of global open science [2].

Metadata has a key role in implementing major changes to entrenched tenets of the predominant research culture, notably that of scholarly publications as the sine qua non for research integrity and professional recognition [1, 2, 10, 11].

2.1 Metadata’s Role in Research Integrity as a Precursor to the FAIR Principles

Prior to FAIR’s emergence, measures of research integrity and professional recognition centered on scholarly publications. In the 2000s, scholars in the scientific community proposed to structure abstracts with metadata as part of the peer review process [12, 13]. The tool at hand was text mining software, and the structuring technique involved fitting abstracts with “electronically annotated information,” or EAI [12, p. 1178].

Proponents viewed publishers as arbiters of research integrity due to the connection between publications and their deposit in electronic databases. Proponents held that if EAI necessarily began with researchers, peer review during the publication process mitigated bias and ensured research integrity [12, 13]. Conversely, proponents assigned
a secondary role to a curator “versed in the particular content descriptors for a given species or subject of research” [13, p. 1]. In sum, the scientific scholars considered metadata expertise ancillary to peer review.

By the 2010s, publishers based accountability for research integrity on reproducible data sets [14, 15]. The unreliability of researcher data sets, however, posed obstacles to—even a crisis in—reproducibility [5, 7, 14, 16, 17]. Greater integration of journals and data repositories, along with the involvement of researchers in data management, now supports research integrity whether data is discrete or aggregated [6, 16, 18, 19]. Persistent digital objects are key to this integrated process but do not, in themselves, ensure the success of (re)using resource content. Persistent digital objects must be reliable to sustain the integrity of resource content [1, 2, 11]. By the middle of 2016, FAIR crystallized these concerns and provided a way forward [1].

Scholars increasingly view FAIR as an assist for (inter)disciplinary research and knowledge creation. The FAIR principles highlight far-reaching opportunities for data (re)use to enrich scholarship, substantiate incentives for scholars, and promote a broader FAIR research ecosystem[1, 2, 20, 21]. Yet even as scholars embrace the FAIR principles’ importance for research outputs, they have not considered these principles’ application in pedagogy. This case study explores how pedagogical research can complement academic research and contribute to the research lifecycle via “FAIRised” [2, p. 8] pedagogical documentation.

3 Pedagogical Research on Digital Humanities Syllabi and LIS Curricular Resources

Given the absence of digital objects in syllabi and LIS curricula more generally, two studies relied on semi-automated and manual data collection techniques to analyze curricular resources in two areas: digital humanities syllabi and health-related LIS courses [22, 23].

3.1 Digital Humanities Syllabi Research

The first study, completed in 2019, represents combined semi-automatic and manual content analysis techniques as a means for exploring the relationship between formal digital humanities training and skills needed to conduct public-funded (inter)disciplinary research [22]. Digital humanities syllabi links, uploaded to an open access Zotero group, provided a means to analyze digital humanities training [24]. Zotero is a software platform for citing bibliographic data and permits subscribers to create “groups” for sharing resources. Spiro’s (2012) Zotero groups upload digital humanities syllabi links for this purpose.1

Revisiting the “DHSyllabi” subfolder on August 20, 2020 for the present case study indicates that syllabi link uploads stopped in 2014 with the exception of one entry added

---

1 https://www.zotero.org/groups/25016/digital_humanities_education/items/collectionKey/MXXEMX7P.
in 2015 and one in 2017.\footnote{https://www.zotero.org/groups/25016/digital_humanities_education/collections/MXXEMX7P/items/R5WVPCUP/collection.} Use of syllabi links in the 2019 study found that most are inoperable, but exporting syllabi links from Zotero reuses semi-structured data in the links. All Zotero metadata fields were selected for export into a spreadsheet. Analysis of 236 syllabi links spanning a 17-year period from 1998–2014 appears in Table 1.

Table 1. Analysis of data in syllabi links exported from Zotero group “DHSyllabi”.

| Zotero field with syllabi data | Syllabi data description | Total cells with data | Percent of sample |
|-------------------------------|--------------------------|-----------------------|------------------|
| Item Type                     | Webpage                  | 236                   | 100%             |
| Title                         | Course title             | 236                   | 100%             |
| URL                           | Syllabi hyperlink        | 236                   | 100%             |
| Date Added                    | Month-day-year           | 236                   | 100%             |
| Date Modified                 | Month-day-year           | 236                   | 100%             |
| Access Date                   | Month-day-year           | 236                   | 100%             |
| Date                          | Syllabi dated by URL     | 95                    | 40.25%           |
| Author                        | Instructor name          | 56                    | 23.73%           |
| Publication Year              | Year (i.e. 2014)         | 34                    | 14.41%           |
| Publication Title             | Course title             | 30                    | 12.71%           |
| Abstract Note                 | Course description       | 22                    | 9.32%            |
| Short Title                   | Course title             | 13                    | 5.51%            |
| Type                          | Syllabus                 | 10                    | 4.24%            |
| Manual tags                   | Instructor keywords      | 9                     | 3.81%            |

Zotero automatically parses syllabi data into the 14 Zotero-labeled fields in column 1, “Zotero field with syllabi data”.\footnote{A few fields inapplicable to the syllabi, such as “Interviewer” and “Cosponsor”, populated sparsely with random data and are not included in Table 1. The “Cosponsor” field sparsely filled with “English”, suggesting that “Language” could be a useful field in structured syllabi.} Column 2, “Syllabi data description” corresponds to fields in syllabi. For example, Zotero’s “Item Type” identifies each syllabus as a webpage, “Title” fills with the course title if provided, random data or a blank cell if not; and “URL” fills with the syllabi’s hyperlink. These three fields apply to all 236 records (100%). Three other fields at 100%, “Date Added”, “Date Modified”, and “Access Date” refer to the link’s Zotero upload date and activities.

Although more than half of the Zotero fields are sparsely populated, six (“Date”, “Author”, “Publication Title”, “Abstract note”, “Type”, and “Manual Tags”) suggest crosswalks to basic syllabi elements. Unfortunately, essential syllabi elements, namely student assessments and course readings, drop out completely on export. More robust options for syllabi metadata are imperative for digital object properties.
3.2 LIS Course Catalog Research

The second pedagogical research project, completed in 2021, utilizes American Library Association (ALA) links to accredited LIS programs in the US to gather health-related course titles and descriptions from institutions’ course catalogs [23]. Course descriptions’ unstructured format necessitates laborious preprocessing to generate a coding system. Of 118 course titles and descriptions collected manually from 40 LIS programs, subsequent manual analysis techniques winnowed this data set to 35 (29.66%) having relevance for the project [23].

Conversely, curricular resources as digital objects could permit seamless extraction from course catalogs on institutions’ websites. Health and other (inter)disciplinary topics represent especially promising areas in which FAIR principles undergird topic genealogies and how they change over time. Data may be accessed and directed to problems devised by humans, or even by machines [11]. The tremendous potential for “FAIR-ised” [2, p. 8] curricular resources to enhance the research lifecycle and augment the global FAIR ecosystem lies as yet untapped.

4 Discussion

FAIR principles are a key step toward machine actionability, defined as a “continuum of possible states wherein a digital object provides increasingly more detailed information to an autonomously-acting, computational data explorer” [1, p. 3]. In order to apply FAIR principles and reap their benefits, digital objects must first have unique and persistent identifiers [1, 2]. The persistent identifier, part of a metadata application profile, wraps a digital object core, includes a license, and provides for long-term access on the Semantic Web [2, 25].

Structuring curricular resources with metadata proper to their status as persistent digital objects is a first step toward making them FAIR for the classroom as well as for research. As suggested in the 2019 and 2021 projects, “FAIRised” [2, p. 8] pedagogical documentation is needed to facilitate efficient and reliable pedagogical research on how (inter)disciplinary topics are taught and students trained. An added benefit is the capability to link curricular resources to publications and data sets in academic libraries, archives, and data repositories [7, 18, 26]. The (re)use of research outputs, inclusive of curricular resources, allows for more complete comprehension of interdependencies within the research lifecycle [27]. Consequently, “FAIRised” [2, p. 8] pedagogical documentation can function as an access point to entire networks of related (inter)disciplinary research outputs in the (inter)disciplinary research lifecycle [8, 28].

This case represents a starting point for “FAIRising” [2, p. 8] syllabi digital object properties. Much as syllabi are sources for investigation, course titles and descriptions undergird LIS pedagogical research in the 2021 project [23]. As exemplified in Table 2, Zotero metadata and syllabi metadata may help prototype digital object metadata for curricular resources.

“FAIRised” [2, p. 8] curricular resources including syllabi, course titles and descriptions, and course content such as student assessments and readings could be reliably implemented with future reuse in mind. Future research on “FAIRising” [2, p. 8] pedagogical documentation may advance pedagogical research along three lines. First, digital
Table 2. Zotero metadata, syllabi metadata, and digital object metadata for curricular resources

| Zotero metadata | Syllabi metadata | Digital object metadata for curricular resources |
|-----------------|------------------|-------------------------------------------------|
| Item type       | Webpage          | Webpage(s)                                      |
| Title           | Course title     | Course title                                    |
| URL             | Syllabus hyperlink| Syllabus hyperlink                              |
| Publication year| Year (syllabus date) | Syllabus date, semester                        |
| Author          | Instructor name  | Institution, department                        |
| Abstract note   | Course description| Course description                             |
| Type            | Syllabus         | Syllabus                                        |
| Manual tags     | Instructor keywords | (inter)disciplinary topics                   |
| Manual tags     | Student assessments| Projects, tools, course outputs               |
| Manual tags     | Course readings  | Links to external research outputs             |

Object properties and metadata are to be determined, as suggested by case study research. Second, metadata for curricular resources should be linked to journal publications and data sets, fostering data (re)use. Third, because metadata both constitutes and contains documentation for curricular resources over time, “FAIRising” [2, p. 8] concerns both longitudinal and cross-institutional investigations. Based on these three recommendations, pedagogical research on “FAIRised” [2, p. 8] pedagogical documentation can complement academic research, enhance the research lifecycle, and advance the FAIR ecosystem long term.

5 Conclusion

Pedagogical research has potential to complement academic research, but the role of curricular resources in the research lifecycle has not been considered. This potential is circumscribed while pedagogical documentation remains unstructured text. “FAIRised” [2, p. 8] curricular resources are imperative for conducting pedagogical research on how (inter)disciplinary topics are taught and students trained. As part of the research lifecycle, such curricular resources augment research outputs. “FAIRising” [2, p. 8] pedagogical documentation for the research lifecycle is as much an opportunity for individual researchers and institutions as it is a necessary step in advancing scholarship holistically throughout the global FAIR ecosystem.

References

1. Wilkinson, M.D., et al.: The FAIR Guiding Principles for scientific data management and stewardship. Sci. Data 3, 1–9 (2016)
2. Collins, S., et al.: Turning FAIR into reality: final report and action plan from the European Commission expert group on FAIR data. Publications Office of the European Union, LU (2018)
3. Tenopir, C., et al.: Research data services in academic libraries: data intensive roles for the future? J. eSci. Libr. 4(2), e1085 (2015)
4. Vitale, C.R.H., Marshall, B., Nurnberger, A.: You’re in good company: unifying campus research data services. Bull. Am. Soc. Inf. Sci. Technol. (Online) 41(6), 26–28 (2015)
5. Stodden, V., et al.: Setting the default to reproducible: reproducibility in computational and experimental mathematics. In: Stodden, et al. (eds.): ICERM Workshop on Reproducibility In Computational and Experimental Mathematics, ICERM, pp. 1–19. Brown University, Providence, RI (2013)
6. MacMillan, D.: Data sharing and discovery: what librarians need to know. J. Acad. Libr. 40(5), 541–549 (2014)
7. Akmon, D., et al.: Building tools to support active curation: lessons learned from SEAD. Int. J. Digit. Curation 12(2), 76–85 (2018)
8. Mayernik, M.S., et al.: Enriching education with exemplars in practice: iterative development of data curation internships. Int. J. Digit. Curation 10(1), 123–134 (2015)
9. Hirsch, B.D.: Digital Humanities Pedagogy: Practices, Principles and Politics. Open Book Publishers, Cambridge, MA (2014)
10. Harrower, N., et al.: Sustainable and FAIR Data Sharing in the Humanities: Recommendations of the ALLEA Working Group E-Humanities. In: Harrower, N. et al. (eds.) ALLEA Report February 2020, pp. 1–42. All European Academies, Berlin (2020)
11. Thompson, M., et al.: Making FAIR easy with FAIR tools: from creolization to convergence. Data Intell. 2(1–2), 87–95 (2020)
12. Leitner, F., Valencia, A.: A text-mining perspective on the requirements for electronically annotated abstracts. FEBS Lett. 582(8), 1178–1181 (2008)
13. Seringhaus, M., Gerstein, M.: Manually structured digital abstracts: a scaffold for automatic text mining. FEBS Lett. 582(8), 1170 (2008)
14. Christian, T.-M., et al.: Operationalizing the replication standard. Int. J. Digit. Curation 13(1), 114–124 (2018)
15. Dafoe, A.: Science deserves better: the imperative to share complete replication files. PS: Polit. Sci. Polit. 47(01), 60–66 (2014)
16. International Society for Biocuration: Biocuration: distilling data into knowledge. PLOS Biol. 16(4), e2002846 (2018)
17. Marwick, B.: Computational reproducibility in archaeological research: basic principles and a case study of their implementation. J. Archaeol. Method Theor. 24(2), 424–450 (2017)
18. Niu, J.: Archival intellectual control in the digital age. J. Arch. Organ. 12(3–4), 186–197 (2015)
19. Marshall, B., et al.: Organizing, contextualizing, and storing legacy research data: a case study of data management for librarians. Issues Sci. Technol. Libr. https://explore.openaire.eu/search/publication?articleId=datacite____::5c55fccc1fbbcb531a1f7a378ae4d9845. Accessed on 02 Mar 2017
20. Dunning, A., De Smaele, M., Böhmer, J.: Are the FAIR principles fair? Int. J. Digit. Curation 12(2), 177–195 (2017)
21. Hank, C., Bishop, B.W.: Measuring FAIR principles to inform fitness for use. Int. J. Digit. Curation 13(1), 35–46 (2018)
22. Garwood, D.A., Poole, A.H.: Pedagogy and public-funded research: an exploratory study of skills in digital humanities projects. J. Doc. 75(3), 550–576 (2019)
23. Garwood, D.A., Poole, A.H.: Vital signs: health literacy and library and information science pedagogy in the United States. J. Educ. Libr. Inf. Sci. 62(1), n–n (2021)
24. Spiro, L.: Opening up digital humanities education. In: Hirsch, B.D. (ed.) Digital Humanities Pedagogy: Practices, Principles and Politics, pp. 331–363. Open Book Publishers, Cambridge, England (2012)
25. Greenberg, J., et al.: A metadata best practice for a scientific data repository. J. Libr. Metadata 9(3–4), 194–212 (2009)

26. Ramírez, M.L.: Whose role is it anyway? A library practitioner’s appraisal of the digital data Deluge. Bull. Am. Soc. Inf. Sci. Technol. (Online). Silver Spring, 37(5), 21–23 (2011). Accessed 02 Mar 2017

27. Borgman, C.L., Wallis, J.C., Mayernik, M.S.: Who’s got the data? Interdependencies in science and technology collaborations. Comput. Support. Coop. Work (CSCW) 21(6), 485–523 (2012)

28. Giles, C.L.: Scholarly big data: information extraction and data mining. In: Proceedings of the 22nd ACM International Conference on Conference on Information & Knowledge Management 2013, pp. 1–2. CIKM, San Francisco, California, USA (2013)