

COMMENTARY

Moving Towards FAIR Data Practices in Pharmacy Education

Jacqueline E. McLaughlin, PhD, MS,^a Alexander Tropsha, PhD, MS,^a Joseph A. Nicolazzo, PhD,^b Anita Crescenzi, PhD, MS,^a Kim LR Brouwer, PharmD, PhD^a

^a University of North Carolina, Eshelman School of Pharmacy, Chapel Hill, North Carolina

^b Monash University, Faculty of Pharmacy and Pharmaceutical Sciences, Melbourne, Australia

Corresponding Author: Jacqueline E. McLaughlin, University of North Carolina. Eshelman School of Pharmacy, Chapel, 322 Beard Hall, Chapel Hill, NC 27599. Tel: 919-966-4557. Email: jacqui_mclaughlin@unc.edu

Submitted March 30, 2021; accepted June 15, 2021; ePublished June 2021

Pharmacy schools are generating significant amounts of data across the training continuum, including data about student selection, performance, and job placement. However, current data practices limit our ability to effectively leverage the vast amounts of data available within and across our institutions. In an effort to improve data practices and promote the quality and reusability of data, the FAIR guiding principles for data management and stewardship were developed and published in 2016. These principles state that digital objects should be [F]indable (eg, data have a unique identifier and are registered in a searchable resource), [A]ccessible (eg, data are retrievable by their identifier using an open, free, standardized protocol), [I]nteroperable (eg, data use a formal, accessible, shared, and broadly applicable language and include qualified references to other data), and [R]eusable (eg, data are described with accurate and relevant attributes, released with a data usage license, and meet domain-relevant community standards). This commentary advocates for improved data practices and provides recommendations for advancing FAIR data principles in pharmacy education.

Keywords: research; rigor; reproducibility; FAIR data

INTRODUCTION

Pharmacy schools are generating significant amounts of data across the training continuum, including data about student selection (eg, recruitment, interviews), performance (eg, observations, examinations), and job placement (eg, career placement, residency matching). While some data are generated primarily for the purpose of operational decision-making, such as selecting students or evaluating student skills, other data might be generated primarily for education research, such as surveying underrepresented students about their career interests.¹ In some cases, data collected for operational purposes might also be used for research, such as examining residency attainment across pharmacy schools.²

Pharmacy educators and educational researchers increasingly rely on computational tools to act on a wide range of data types, formats, and protocols that support empirical research and decision-making. If you consider a student's entire educational record, for example, there might be numeric data related to admissions, course grades, class attendance, objective structured clinical examination (OSCE) ratings, experiential grades, Pharmacy Curriculum Outcomes Assessment (PCOA) scores, a North American Pharmacy Licensing Examination (NAPLEX) score, and demographic characteristics. There might also be non-numeric (or unstructured text/image) data such as admissions statements or recommendations, instructor and preceptor comments, photographs, and portfolios. Further, these data may reside in numerous systems or software that use varied database structures and processes for access, analysis, and reporting such as learning management systems, human resources systems, testing or course evaluation software, and admissions or experiential databases. While these types of data can provide insight into critical questions about learning and development, current data practices limit our ability to effectively access, federate, process, and analyze the vast amounts of data available within and across educational institutions. If one wanted to examine multiple years of admissions data and their relationship to student performance, for example, one would be faced with issues related to aggregating data from multiple databases, understanding how admissions interview scores and student grades were measured, addressing changes to admissions and grade methodologies over time, operationalizing demographic variables, and handling missing data. Attempting to aggregate data across multiple institutions might further amplify these types of challenges since institutions may use different approaches to define, collect, process, and analyze their data. These examples illustrate possible inconsistencies in data practices, such as (1) discrepancies in protocols for data acquisition, curation, and validation, (2) lack of clear data standards and definitions, (3) increasing volume and complexity of data and analytic tools, and (4) limited data sharing. Inconsistent data practices can further manifest themselves in publications as

insufficient descriptions of methods and results, selective data presentation, inconsistent ontologies, poor data management, and incorrect analytic techniques, all potentially leading to concerns regarding data reproducibility and trustworthiness.³

The FAIR data principles for scientific data management and stewardship provide guidelines aimed at improving data practice and promoting the quality and reusability of data.⁴ Briefly, FAIR principles state that data should be [F]indable (eg, data have a unique identifier and are registered in a searchable resource), [A]ccessible (eg, data are retrievable by their identifier using an open, free, standardized protocol), [I]nteroperable (eg, data use a formal, accessible, shared, and broadly applicable language and include qualified references to other data), and [R]eusable (eg, data are described with accurate and relevant attributes, released with a data usage license, and meet domain-relevant community standards).⁴ These principles inform options for data generation, tool development, object (re)use, and long term stewardship, effectively enabling researchers to advance their digital ecosystems and promote discovery, innovation, and post-publication integration.^{4,5}

FAIR principles are particularly well suited for education given: 1) their emphasis on improving the rigor and reuse of data to advance knowledge and 2) the increasing ability of computers to enable access to and analysis of data. In health professions education, a small yet growing body of literature advocates for the improvement of our data practices and uptake of FAIR principles.⁶⁻¹⁰ Schwartz and colleagues, for example, argue for digital repositories in medical education to promote replicability of studies, enable secondary data analysis to answer new scientific questions, and overcome limitations associated with small sample single-site studies.⁹ To that end, utilizing FAIR principles might also empower experiential educators overcome statistical challenges associated with small sample sizes by enabling them to aggregate data across multiple practice sites or institutions.

Further, some health professions education journals now encourage or require data sharing, data citations, and/or data statements.^{11,12} *Medical Education*, for example, “encourages authors to share the data and other artefacts supporting the results in the paper by archiving it in an appropriate public repository. Authors should include a data accessibility statement, including a link to the repository they have used, in order that this statement can be published alongside their paper.”¹¹ *BMC Medical Education* requires data availability statements that “should include information on where data supporting the results reported in the article can be found including, where applicable, hyperlinks to publicly archived datasets analysed or generated during the study. By data we mean the minimal dataset that would be necessary to interpret, replicate and build upon the findings reported in the article.”¹²

Utilizing FAIR data principles can 1) enable reproducibility and replication of research, 2) facilitate secondary data analyses that answer new research questions (eg, additional analyses, meta-analysis), 3) address increasing requirements from funding sources and publishers, 4) advance science, and 5) promote scholar reputation through citations and visibility.^{6-8,10} It is both timely and critical that pharmacy education adopt data strategies aimed at promoting the quality and trustworthiness of our educational practice and research. To that end, we offer the following recommendations for moving toward the adoption of FAIR use in pharmacy education.

Recommendations

- **Conduct Needs Assessments.** Assessing data practices within pharmacy education will elucidate current practices, needs, infrastructure, and barriers and enable educators to identify gaps between current and FAIR data practices. This would provide a comprehensive understanding of 1) the current state of data practices, including the requirements and infrastructures provided by journals, funding agencies, data repositories, and other institutions; 2) the met and unmet data needs of key stakeholders of FAIR data practices; and 3) the real and perceived barriers to FAIR data practices for various stakeholder groups. As a starting point, this type of needs assessment could be conducted for an individual research team/laboratory, an entire pharmacy school, or across a specific pharmacy education sub-discipline (eg, OSCE data). Ultimately, this deep understanding of the current state of data practices and related gaps will provide the foundation for an evidence-informed roadmap to FAIR data practices for the scholarly community and help craft specific recommendations for educational and research endeavors tailored to individuals and institutions.
- **Train Educators and Researchers.** Educators and researchers must be equipped with the knowledge and skills needed to teach and apply FAIR data principles. As noted by Mons and colleagues, “FAIR compliant data stewardship will require many different skills that are not traditionally covered by the research curricula of contemporary students and researchers. Therefore, extensive training capacity and training materials are needed, and in need of development.”⁵ Furthermore, FAIR practices rely heavily on humans to format and communicate data, tools, and findings in a way that enables understanding and mitigates ambiguity associated with natural language barriers (eg, jargon). Although some researchers believe access to shared data will advance research,

training will need to address strategies for overcoming identified FAIR barriers, such as: unfamiliarity with data sharing; strategies for handling the effort, time, and cost of preparing data to share; uncertainty regarding benefits to oneself or others; and ethical and legal issues associated with FAIR practices.^{9,13}

- **Build Communities of Practice.** Since the FAIR principles were first described in 2016,⁴ an estimated 1,000 communities of practice have implemented the FAIR principles across various disciplines.⁵ These communities have explored ontology mapping, machine learning, automation, and annotation and curation, among other FAIR issues.¹¹ Within pharmacy education, several strategies could be utilized to identify and connect with those interested in building communities aimed at advancing these efforts. Professional organizations, such as the American Association of Colleges of Pharmacy (AACP) and Research Data Alliance (RDA), often include special interest groups (SIG) with shared interests related to assessment, data, and FAIR use (eg, AACP Assessment SIG, RDA Raising FAIRness in Health Data and Health Research Performing Organisations Group, RDA Health Data Interest Group). As an increasing number of funding agencies adopt and advocate for improved data practices, such as the National Science Foundation's "Harnessing the Data Revolution",¹⁴ individuals may also find potential collaborators among lists of funded projects.

Pharmacy educators might also find individuals and/or committees within their departments, schools or universities focused on relevant data issues, such as data governance, data science, and information science. Because student pharmacists are taught by experts from various disciplines, such as biomedical science, education, library science, and healthcare practice, it is natural and necessary to create interprofessional FAIR communities that embody relevant expertise. Together, pharmacy educators and researchers must leverage this diversity of expertise to identify and engage collaborators that can identify and implement FAIR strategies.

- **Integrate into Educational Practice and Research.** First generation implementation of FAIR data principles indicates disciplinary differences when it comes to FAIR use, including legal and workflow issues.¹⁵ The use of FAIR data principles in education could include navigating legal restrictions associated with accessing and sharing data (eg, Family Educational Rights and Privacy Act), adhering to ethical principles (eg, deidentifying data), and consistently operationalizing commonly-used constructs. For example, a single construct, such as wellness, may be measured differently by different institutions, challenging the interoperable nature of the data.

The strategies and related effort required to implement FAIR principles can vary widely depending on the goals of the individual or institution, and the complexity of the data. Some strategies may be achieved with minimal additional effort, such as making a full survey instrument and related de-identified data available. Others might require more effort, such as publishing qualitative transcripts, given the unique processing required to ensure confidentiality of participants.¹⁶ Even more complex would be data subject to FERPA restrictions, in which case data instruments may be easily shareable yet the data itself could not be shared. As noted by BMC Medical Education, "it is not always possible to share research data publicly, for instance when individual privacy could be compromised."¹²

As a discipline, we must elucidate how FAIR principles intersect with education and how to best position pharmacy educators to integrate them into their research and practice. A survey of Society for Directors of Research in Medicine Education members, for example, indicated that data sharing was strongly supported, yet few reported actually using data repositories.⁹ Clearly, the challenges facing implementation of FAIR use in education are not trivial and will likely require sustained effort and strategy across a wide range of expertise.

- **Leverage Machine Learning.** Data-intensive research and decision-making rely on the ability of humans to discover, access, integrate and analyze task-appropriate scientific data.⁴ As the amount and diversity of data related to pharmacy education (such as aforementioned data on student selection, performance, and job placement) grows and becomes available in FAIR-compliant datasets, their manual, "eye-balling" analysis of such data becomes increasingly unattainable. Thus, computational tools such as machine learning are growing in use for education data analysis and trend elucidation.¹⁷

Machine learning relates to a large group of computational and statistical algorithms and respective software tools that can explore the relationships and correlations between various characteristics of digital objects; for instance, machine learning models can help predict student performance in pharmacy curricula based on student admission data.¹⁸ FAIR-compliant data alone are a "lame substrate" without workflows that empower the formatting, communication, and consumption of that data.⁵ Thus, we should continue to leverage machine learning tools to make the most efficient use of education data to improve student selection, advance their education, and help them find the most satisfying employment. Achieving these goals is, arguably, a common objective of all educational institutions and the use of bigger datasets typically improves the accuracy and reliability of machine learning models. Thus, promotion of FAIR data principles in pharmacy education will

enable data standardization and integration across multiple schools, and, consequently, growth of respective educational databases. We surmise that these advances will catalyze data-driven decision support and improve the efficiency of pharmacy education.

CONCLUSION

Current data practices limit our ability to leverage the vast amount of data across pharmacy education. It is timely and critical to adopt data strategies that promote quality and trustworthiness in our education research. Given the complexity of data, data systems, and computations tools across pharmacy education, efforts to improve data practices must be intentional, strategic, and pursued with sustained commitment and multidisciplinary expertise. As first steps, pharmacy educators and researchers must conduct needs assessments, invest in FAIR training, build communities of practice, integrate FAIR principles, and pursue advanced data-analytical machine learning approaches that can inform and improve the training of healthcare practitioners. These efforts will help ensure that data are utilized with accuracy and efficiency across the continuum of education.

REFERENCES

1. Hahn FT, Bush AA, Zhang K, Patel A, Lewis K, Jackson A, McLaughlin JE. Exploring the career engagement, interests, and goals of students identifying as underrepresented racial minorities in pharmacy. *Am J Pharm Educ.* 2021; 84(4): Article 8364.
2. Lyons K, Taylor DA, Minshew LM, McLaughlin JE. Student and school-level predictors of pharmacy residency attainment. *Am J Pharm Educ.* 2018; 82(2): Article 6220.
3. Collins FS, Tabak LA. Policy: NIH plans to enhance reproducibility. *Nature.* 2014;505:612–613.
4. Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, et al. The FAIR guiding principles for scientific data management and stewardship. *Sci Data.* 2016; 3:160018.
5. Mons B, Schultes E, Liu FH, Jacobsen A. The FAIR principles: First generation implementation choices and challenges. *Data Intel.* 2020;2: 1–9.
6. Borgman CL. The conundrum of sharing research data. *J Am Soc Inf Sci Tec.* 2012; 63: 1059-1078. doi:10.1002/asi.22634
7. Fecher B, Friesike S, Hebing M, Linek S. A reputation economy: how individual reward considerations trump systemic arguments for open access to data. *Palgrave Comm.* 2017; 3(1): Article 1.
8. Feger SS, Wozniak PW, Lischke L, Schmidt A. ‘Yes, I comply!’: Motivations and practices around research data management and reuse across scientific fields. *Proceedings of the ACM on Human-Computer Interaction.* 2020; 4(CSCW2): 1–26. <https://doi.org/10.1145/3415212>
9. Schwartz A, Pappas C, Sandlow LJ. Data repositories for medical education research: Issues and recommendations. *Acad Med.* 2010; 85(5): 837–843.
10. Ju B, Kim Y. The formation of research ethics for data sharing by biological scientists: an empirical analysis. *Aslib J Info Manag.* 2019; 71(5).
11. Medical Education Author Guidelines. <https://onlinelibrary.wiley.com/page/journal/13652923/homepage/forauthors.html>. Accessed March 30, 2021.
12. BMC Medical Education Submission Guidelines. <https://bmcmededuc.biomedcentral.com/submission-guidelines/preparing-your-manuscript/research-article>. Accessed March 30, 2021.
13. Tenopir C, Allard S, Douglass K, Aydinoglu AU, Wu L, Read E, et al. Data sharing by scientists: Practices and perceptions. *PloS ONE.* 2011; 6(6): e21101.
14. National Science Foundation. Harnessing the Data Revolution. <https://www.nsf.gov/cise/harnessingdata/>. Accessed June 10, 2021
15. Jones S, Pergl R, Hooft R, Miksa T, Samors R, Ungvari J, Davis, RI, Lee T. Data management planning: How requirements and solutions are beginning to converge. *Data Intel.* 2020;2:208–219.
16. McGrath C, Nilsonne G. Data sharing in qualitative research: opportunities and concerns. *MedEdPublish.* 2018;7(4), Article 4. <https://doi.org/10.15694/mep.2018.0000255.1>
17. Kučak D, Juričić V, Đambić G. Machine learning in education – A survey of current research trends. *Annals of DAAAM & Proceedings.* 2018;29.
18. Muratov E, Lewis M, Fourches D, Tropsha A, Cox WC. Computer-assisted decision support for student admissions based on their predicted academic performance. *Am J Pharm Educ.* 2017; 81(3): Article 46.