Interactive Session: Exploring the Data Stewardship Toolbox Created by and for Multi-Disciplinary Research Teams

Dr. Carla B. Zoltowski, Purdue University – West Lafayette (College of Engineering)

Carla B. Zoltowski is an associate professor of engineering practice in the Elmore Family School of Electrical and Computer Engineering (ECE) and (by courtesy) the School of Engineering Education, Director of the Vertically Integrated Projects (VIP) Program, and Interim Director of the EPICS Program, all within the College of Engineering at Purdue. She holds a B.S. and M.S. in Electrical Engineering and a Ph.D. in Engineering Education, all from Purdue. Dr. Zoltowski's research includes the professional formation of engineers, human-centered design, and engineering ethics.

Dr. Wei Zakharov, Purdue University – West Lafayette

Dr. Wei Zakharov is an Associate Professor and Engineering Information Specialist in Libraries and School of Information Studies and (by courtesy) the School of Engineering Education at Purdue University. She is the faculty liaison to four Schools Aeronautical and Astronautical Engineering, Electrical and Computer Engineering, Semiconductors, Engineering Education, and Aviation and Transportation Technology from Polytechnic Institute at Purdue University. She works with students and faculty to address their needs regarding information resources, research data management, and scholarly communication. Dr. Zakharov teaches and conducts research in data and information literacy education, and AI and online learning. As a principal investigator (PI) or co-PI, she has multiple internal and external grants, including grants from National Science Foundation (DUE, CCRI, RAPID) and one Institute of Museum and Library Services (IMLS) National Leadership Grant. She serves as an editorial member and reviewer for several international research journals.

Dr. Senay Purzer, Purdue University – West Lafayette (College of Engineering)

Senay Purzer is a Professor in the School of Engineering Education at Purdue University. Her research is on engineering design reasoning.

Interactive Session: Exploring the Data Stewardship Toolbox Created by and for Multi-Disciplinary Research Teams

Abstract

This purpose of this interactive session is to support ethical data management education within undergraduate research programs, particularly in multi-disciplinary settings. We introduce the Data Stewardship Framework and Toolkit aimed at guiding mentors and students through the ethical, disciplinary, and practical aspects of managing research data responsibly. The participants will explore the Data Stewardship Toolkit and engage in small group discussion on ways to use the tools and the framework in support of multi-disciplinary projects.

Introduction

In today's world, where data are ubiquitous, practicing ethical data management is crucial especially in the era of AI and big data. Ensuring that data are findable, accessible, interoperable, and reusable is becoming increasingly important. We are a multi-disciplinary research team whose goal is to embed ethical data life cycle management education into undergraduate research programs in Engineering and Applied Sciences. In this paper, we present the "Data Stewardship Framework," which we have developed to serve as a guide for mentoring faculty and graduate students. We define Data Stewardship as the care and management of data throughout the project. It involves being able to identify underlying disciplinary, data management, and ethical principles and articulate trade-offs involved in making decisions throughout the project. We then provide an overview of the toolkit that we developed to prompt mentors and undergraduate researchers to reason through ethical considerations, future consequences, and trade-offs as they plan, collect, describe, manage, visualize, curate, and share the data. Finally, we provide evidence of the framework's educational impact.

At the conference, we will present a short overview of the Data Stewardship Framework and toolkits, then lead interactive activities which will allow the participants to engage with examples of how the tools have been applied in an undergraduate research project and explore how these can be used in their own research and design projects. Participants who will benefit from learning how to be better data stewards in their research and design projects by engaging with these tools that integrate frameworks from data management, ethics and design reasoning, and as a result, effectively guide the graduate and undergraduate researchers to reason through ethical considerations, future consequences, and trade-offs throughout the data management lifecycle.

Why does Research Need a Framework focused on Data?

Data are key to both enabling the research and to providing evidence of its effectiveness and impact. The Cambridge Dictionary defines the action (verb) of research as "to study a subject in detail, especially in order to discover new information or reach a new understanding" [1]. Papers written to disseminate the findings often include methods sections that describe specific data collection and data analysis processes. Although not explicitly stated in the definition, researchers have a responsibility for the ethical management of data throughout the project [2]. For students who participate in real-world undergraduate research projects with faculty and graduate mentors, "they cross the threshold from being knowledge consumers to becoming knowledge creators. Information and data literacy are integral to the research experience" [3, p1].

However, responsible management of data is often not addressed explicitly in many disciplines, including engineering and applied sciences. As a result, undergraduate research teams often face challenges related to ethical data management across all stages of the research data lifecycle [3-5]. The study by Arigye et al. [6] graduate mentors of engineering undergraduate research teams reported persistent challenges in areas such as documentation quality, metadata creation, ethical practices, and long-term data preservation.

Data Stewardship Framework

The Data Stewardship Framework integrates frameworks from multiple disciplines to support ethical data life cycle management education and practice into undergraduate research programs. Data Stewardship is the care and management of data throughout the research project. It involves being able to identify underlying disciplinary, data management, and ethical principles and articulate trade-offs involved in making decisions throughout the project. It is informed by prior learning and experiences related to the project and consideration of future implications of the work and facilitates learning and new knowledge. There are three components of the Data Stewardship Framework:

1. Data Principles
2. Ethical Principles
3. Design Reasoning

The data principles include Findability, Accessibility, Interoperability, and Reusability (FAIR) [7]. According to Wilkinson et al [7], "All scholarly digital research objects—from data to analytical pipelines—benefit from application of these principles, since all components of the research process must be available to ensure transparency, reproducibility, and reusability." Any data generated including the algorithms that led to locating and generating the data must be findable with unique identifiers and metadata, be in formats that are accessible and retrievable, in formats that are usable with different software, well-described including data usage licenses and limitations to allow reuse.

The framework incorporates the ethical principles from Reflexive Principlism [8], which include beneficence (preventing harm and providing benefits), non-maleficence (avoiding causation of harm), respect for autonomy, and justice (fairly distributing benefits, risks, and costs). The ethical reasoning process of Reflexive Principlism is a "process of iterative specification, balancing, and justification that aligns well with the engineering design process" [8, p. 284].

The design reasoning connects data and ethics principles and includes four components: experiential, first-principles, trade-offs, and futures reasoning. Experiential reasoning values learning or past lived experiences. First-principles reasoning is the application of disciplinary, data, or ethical principles that inform data and research practices. Trade-offs reasoning engages researchers to negotiate risks and benefits and weigh conflicting requirements and the pros and cons of different data choices. Futures reasoning is the ability to imagine future scenarios in new contexts to assess implications and unintended impacts of data-related decisions.

Educational Use and Impact of the Data Stewardship Toolkit

Although data management practices have a long tradition for those scholars and educators from library and data science, many researchers with engineering backgrounds are not familiar with those concepts and practices. Thus, we have adapted toolkits and provided examples of how they can be used to better support undergraduate engineering researchers and their mentors. Our toolkit consists of the checklists and activities shown in Table 1.

Table 1. List of Toolkits and Their Descriptions

Toolkit	Brief Description
Data Management Planning Discussion Guide	Provides a summary or checklist to help research team reach consensus and implement good data stewardship practices.
Data Quality Checklist	Provides a checklist to assess data quality across six key dimensions—Data Credibility, Timeliness, Completeness, Accuracy, Consistency, and Deduplication.
File Level Metadata: Data Dictionary Creation Guide	Provides a data dictionary template to help you document sufficient information about your dataset at the file level.
Project Level Metadata: README Creation Guide	Provides a list of core components of a README template to help you document essential information about your dataset at the project level.
File Naming Convention Development Guide	Provides best practices for creating clear, consistent, and descriptive file names.

File and Folder Organization Guide	Provides tips and examples for organizing files and undergraduate research team project folders.
Data Storage and Backup	Provides guidance on selecting data storage solutions and the 3-2-1 backup data strategy.
Data Preservation Protocol	Provides guidance for preserving data in an organized and consistent manner.
Data Repository Selection Guide	Provides recommendations for choosing appropriate data repositories to publish data.

Evidence of Data Stewardship Framework's Educational Impact

Data stewardship principles and toolkits were embedded into the design and delivery of four undergraduate learning experiences across coursework and research-based instruction [9]. the course, *Understanding Your Research Data*, offered in Spring 2024 and 2025 to fifteen undergraduate research scholars, engaged students in applying ethical data management practices directly to their ongoing research through a hybrid instructional format. Reflection data indicated increased student confidence, efficiency, and ethical awareness in managing data-intensive projects, with students presenting their work at the 2024 and 2025 Purdue Spring Undergraduate Research Conference; the course now fulfills requirements for Purdue's upcoming Undergraduate Research Certificate launching in Fall 2026. Foundational instruction was further scaled through the course entitled *Foundations of Data Management*, offered in Spring and Fall 2025 to a total of 75 students, many pursuing a data science certificate, where students developed core competencies in data organization, preservation, publication, and, in Fall 2025, ethical considerations surrounding AI-enabled data tools.

Beyond formal coursework, these principles were reinforced through mentorship of undergraduates working on faculty-led, multi-semester research projects as part of the Vertically Integrated Projects (VIP) Program at Purdue. Undergraduates and graduate students working on VIP projects involving Unmanned Aerial Vehicles (UAVs), computer vision, and AI tutors integrated data stewardship practices into their research projects. In addition to enhancing their ethical data practices, this mentorship led to the acceptance of a peer-reviewed data journal article [10]. Student reflections from another undergraduate research team in VIP highlighted how the Data Stewardship Framework and Toolkit helped them to identify new principles related to data management [11]. The students' initial reflections focused significantly on the technical approaches and methods used to implement the project, including how they partitioned the data for testing and training, and challenges due to the limited and unbalanced dataset. However, they did demonstrate some awareness of data management issues in their recognition of the challenges they faced related to inconsistent file naming and testing datasets [11].

As a result of the Data Stewardship Toolkit, the students made several improvements related to their data management practices, which included improved file organization, data control, and articulation of additional ethical considerations. They also identified a need for more transparency regarding data sourcing and acquisition, which provided insights to mentors on how to more effectively support ethical data management practices and data stewardship learning in technical undergraduate research projects [11].

Description of Session

The session will begin with a short overview of the Data Stewardship Framework and Toolkit we have developed to serve as a guide for mentoring faculty and graduate students and examples of how they have been applied. In small groups, session participants will engage with examples of how the tools have been applied in an undergraduate research project and explore how these can be used in their own research and design projects and strategies for highlighting and integrating the different disciplinary perspectives. Small groups will share key takeaways to facilitate a broader discussion among session participants. The session will finish with a discussion of next steps for our project and how participants can continue to learn engage in ongoing research and development of educational materials.

Agenda (90 minutes)

- Introductions – 10 minutes
- Overview of Data Stewardship Framework, Tools, and Application Examples – 20 minutes
- Small group discussion of application examples – 15 minutes
- Large group discussion – 15 minutes
- Small group discussion of strategies for highlighting and integrating multiple disciplinary perspectives – 10 minutes
- Large group discussion – 10 minutes
- Wrap-up: 10 minutes

Acknowledgment & Declaration of AI-assisted Technologies

This material is based upon work supported by a National Science Foundation under Grant No. 2236241. Any opinions, findings and conclusions, or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

During the preparation of this work, the authors used Microsoft Copilot to improve check grammar and the readability of paragraphs. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

[1] Cambridge University Press & Assessment (2026). Research. In Cambridge Dictionaries. Retrieved January 21, 2026, from https://dictionary.cambridge.org/dictionary/english/research

[2] Frugoli, J., Etgen, A. M., & Kuhar, M. (2010). Developing and communicating responsible data management policies to trainees and colleagues. *Science and engineering ethics*, *16*, 753-762.

[3] Burress, T. (2022). *Data literacy practices of students conducting undergraduate research*. College Research Libraries, 83(3), 434.

[4] Bossaller, J., & Million, A. J. (2023). The research data life cycle, legacy data, and dilemmas in research data management. *Journal of the Association for Information Science and Technology*, *74*(6), 701–706.

[5] Carlson, J., Fosmire, M., Miller, C., & Nelson, M. S. (2011). Determining data information literacy needs: A study of students and research faculty. *Portal: Libraries and the Academy*, *11*(2), 629–657.

[6] Arigye, J., Zakharov, W., Zoltowski, C., Sewell, S. & Purzer, S. (2026). Investigating graduate mentor perspectives on data information literacy and stewardship in undergraduate engineering research projects. *Journal of eScience Librarianship, 15*(1): e1171. https://doi.org/10.7191/jeslib.1171.

[7] Wilkinson, M.D. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship," Scientific data, vol. 3, no. 1, pp. 1–9. doi: 10.1038/sdata.2016.18 (2016).

[8] Beever, J., & Brightman, A. O. (2016). Reflexive principlism as an effective approach for developing ethical reasoning in engineering. *Science and Engineering Ethics*, 22(2), 275–291.

[9] Zakharov, W., Purzer, S., & Zoltowski, C. B. (2025). Cultivating ethical data stewardship for undergraduate research: A framework and toolkit duo. *Information Outlook,* 2025 Fall Issue: 17-20. https://docs.lib.purdue.edu/lib_fsdocs/329/

[10] Shi, L., Chang, Y., SommerKohrt-, K, Zakharov, W., Mou, S., Sundaram, S., & Goppert, J. (2025). UAS competition ROSbag datasets. *Data in Brief*, *62*, 111986. https://doi.org/10.1016/j.dib.2025.111986.

[11] Purzer, S., Zakharov, W., & Zoltowski, C. B. (n.d.). Stewards of Data: A Practical Handbook for Research in Engineering and Applied Sciences, In preparation.