



*The Creative Commons-Attribution-Noncommercial License 4.0 International applies to all works published by IASSIST Quarterly. Authors will retain copyright of the work and full publishing rights.*

## **Evaluating data sharing practices: A case study for federally funded research using FAIR standards**

Jung Mi Scoulas<sup>1</sup>

### **Abstract**

While open science and access trends emphasize sharing data, the focus often prioritizes structured data, such as quantitative datasets. This emphasis leads to a lack of attention for sharing qualitative data and evaluating such practices. This paper examines both quantitative and qualitative data generated from a federally funded research project, led by the author, with an evaluation of its data management and sharing practices guided by FAIR principles. The research project aimed to develop assessment tools to understand undergraduate students' academic experiences, including library use, and how these experiences shaped their definitions of academic success. Conducted at public research universities, the two-year project (2022–2024) collected both quantitative and qualitative data including students' personal definitions of success. Following the project's completion, the author revisited the data management plan and data sharing policies to assess the data generated from the project, guided by FAIR standards. This process involved identifying the shared data and comparing it against the FAIR principles to determine the extent to which the data complied with those principles, and to identify areas for improvement, with the aim of increasing data reusability. This self-evaluation highlights key findings and critical takeaways in the management and sharing of various types of data, addressing a significant gap in current discussions that often prioritizes datasets. The findings offer actionable insights for researchers and information professionals, aiming to enhance data sharing, improve reusability, and refine their institutional practices.

### **Keywords**

Qualitative data, data sharing, reusability, survey

### **Introduction**

Open science and open access initiatives have increased the emphasis on sharing data and reuse among researchers. However, much of this focus has been on quantitative data. It is often assumed that quantitative research produces numerical data, while qualitative research generates non-numeric data, such as text, audio, or visual content. But is it accurate to draw a strict distinction between these two types of data? Even quantitative data are often accompanied by other types of data, such as survey instruments, open-ended text responses, and audio-visual recordings that document data analysis procedures or provide project background. Although these supplementary data or datasets contain both quantitative and qualitative data, they receive comparatively less attention in discussions

on the FAIR principles, which serve as a guiding framework to improve Findability, Accessibility, Interoperability, and Reusability of digital sources in machine-readable formats. Applying these principles in practice, however, is far from straightforward. Scholars note that the interpretation and implementation of FAIR principles can vary widely across research communities, as stakeholders retain considerable freedom of choice, which makes consistent adoption difficult (Jacobsen et al., 2020; Mons, Schultes, Liu, & Jacobsen, 2020). Yet, relatively few studies examine how FAIR principles can be applied to assess already shared data, particularly when those principles were not explicitly considered during data preparation. This gap raises questions about both the practicality of FAIR and its implications for meaningful user engagement with shared data.

This paper presents a case study from a federally funded research project led by the author, who is also the project director, examining data generated through quantitative methods that also incorporate qualitative components. The objectives of this study are to evaluate how these shared data align with the FAIR principles, to examine the outcomes of this evaluation, and to share critical takeaways for applying these principles to various types of data including qualitative data, and to address modifications. The outcomes and insights from this evaluation will help researchers understand key considerations for data sharing and provide useful guidelines to highlight important aspects, ultimately maximizing data reusability.

## Literature review:

### Federal funding agency's data sharing policies

U.S. federal agencies have increasingly required Data Management and Sharing Plans as part of grant applications to improve public access and preservation of research data. These requirements are reinforced by the 2022 Office of Science and Technology Policy (OSTP) memorandum on public access to research results and subsequent directives, which aim to make publications and supporting data immediately accessible, while promoting research integrity (White House Office of Science and Technology Policy, 2022). For example, the National Science Foundation (NSF) expects investigators to share outputs generated from the research project, such as primary data, samples, physical collections, and other supporting materials with other researchers within a reasonable timeframe (NSF, 2024). Award recipients are also encouraged to actively support and facilitate data sharing to enhance accessibility and collaboration, and to adopt FAIR principles to ensure that shared data is findable, accessible, interoperable, and reusable. The National Institutes of Health (NIH) provides more details in the types and amount of scientific data to be shared, the standards for data and metadata formats, and any necessary tools or software for access (NIH, n.d.). Additionally, NIH grants require investigators to describe data preservation such as repository selection, timelines for availability, and methods for ensuring data findability. Investigators are also encouraged to align their data management and sharing practices with FAIR principles. The federal funding agency well known in the library and information sciences is the Institute of Museum and Library Services (IMLS). According to the IMLS's Public Access Policy Guidance (IMLS, n.d.), researchers must deposit scientific data underlying peer-reviewed publications into a suitable repository immediately upon publication. When determining when data can be shared, researchers should account for various factors such as privacy, ethics, Indigenous rights, intellectual property, security, and legal restrictions. Awardees are also encouraged to align their data management and sharing practices with FAIR Principles, while also

considering CARE Principles for Indigenous data where relevant. Furthermore, “Recipients of IMLS research awards made on or after October 1, 2025 (i.e. FY-2026 awards and later), must comply with the key requirements outlined herein” (IMLS, n.d.). Across all agencies, the emphasis is on accessibility, reproducibility, and advancing scientific discovery while maintaining ethical considerations such as confidentiality and informed consent.

### **Benefits and ethical considerations of data sharing**

The push for data sharing is grounded in well-documented benefits and ethical considerations. Across various disciplines, the benefits of data sharing enhance the accuracy of results and reproducibility, accelerating scientific discovery, promoting collaborations, improving transparency, and supporting evidence-based decision-making (Institute of Medicine in US, 2013; U.S. Geological Survey, n.d.-a). For researchers, depositing data increases visibility and impact. For example, a study found that 48% of cancer microarray trials with publicly available data received 85% of citations, indicating a link between data sharing and higher citation rates (Piwowar, Day, & Fridsma, 2007).

While data sharing offers many benefits, it also raises ethical challenges. These include the inappropriate use of data or its misuse out of context (U.S. Geological Survey, n.d.-a), or misinterpretations or incorrect conclusions when others attempt to use the data, as well as the challenge of safeguarding sensitive information and ensuring confidentiality (Institute of Medicine in US, 2013). To address these issues, frameworks described by Dankar (2023) recommend mechanisms, such as opt-out models, and the use of trust rating systems to evaluate research institutions or data users before granting them access to sensitive data. Nonetheless, even with such measures, legal requirements for informed consent remain complex (Kemp, Nelson, Jenkins, 2023), and participants continue to express concerns about misuse, exploitation, and privacy (Mello, Lieou, & Goodman, 2018).

### **FAIR Principles**

Given the diverse benefits and concerns surrounding data sharing, addressing these issues effectively requires a framework that ensures data is both accessible and reusable. The FAIR principles were published in the paper by Wilkinson et al. (2016), which outlined their key components as a framework to enhance the discoverability and reusability of scientific data. As outlined in Box 2 (Wilkinson et al., 2016, p. 4), these principles provide guidance for implementing FAIR practices, with particular emphasis on enabling machines to automatically find and use data in addition to supporting human reuse. Findable emphasizes that data elements are assigned unique identifiers and indexed in searchable resources. Accessible ensures data can be retrieved through its unique identifier, even if the data itself is no longer available. Interoperable promotes the use of widely applicable languages for knowledge representation. Reusable requires data to include clear and accurate descriptions with well-defined attributes. These principles are applicable across various disciplines and types of data (Sadeh et al., 2023). FAIR principles are often associated with research data, but their intended scope is broader. Wilkinson et al. (2016) state, “it is our intent that the principles apply not only to ‘data’ in the conventional sense, but also to the algorithms, tools, and workflows that led to that data (p. 1).” This suggests that the FAIR principles can be applied broadly to research data as well as to diverse supplementary research outputs, such as instruments, instructional materials, and workflows, guiding their management and reuse. In this paper, the FAIR principles were used to evaluate the publicly

shared research data, including both quantitative and qualitative data, from the author's federally funded project to assess whether they meet the established criteria and identify areas for improvement to enhance future data reusability. To contextualize this study, it is also useful to consider broader practices and challenges in implementing FAIR principles, as identified in the next section.

### **Practices and challenges in implementing FAIR Principles**

Following the publication of the FAIR principles (Wilkinson et al., 2016), a growing body of literature has examined community practices and the challenges associated with implementing these principles. For example, Mons et al. (2020) note that FAIR is a set of guidelines rather than strict standards, promoting gradual progression towards greater data usability, and they discuss the early implementation of the FAIR principles across research communities. Early adopters implemented practices ranging from basic metadata schemas and persistent identifiers to advanced licensing and machine-readable protocols. The authors also discuss persistent challenges, including inconsistent or incomplete metadata, limited infrastructure and tools, unclear standards across disciplines, and the need for skills that are not traditionally part of standard research training. They emphasize that collaborative, community-driven approaches and shared best practices are crucial to overcoming these barriers and enabling the broader adoption of FAIR principles.

Other scholars have also highlighted challenges in implementing FAIR principles and ways to improve their use. For instance, Jacobsen et al. (2020) note that while communities can adopt or develop solutions to fill gaps, differing approaches can lead to inconsistent interpretations and incompatible outcomes. To address this, they provide systematic interpretation and implementation considerations for each principle, guiding stakeholders in applying them consistently. They also discuss implementation choices and challenges, encouraging solutions that can be reused across communities to provide a comprehensive view of FAIR in practice. Similarly, David et al. (2020) emphasize persistent issues, noting that FAIR does not explicitly address data quality, despite its importance for data reuse. They also highlight that the process of making data compliant with the FAIR principles requires considerable time, expertise, and resources, with unclear responsibility for who should contribute. These challenges underscore why FAIR implementation remains varied and inconsistent across disciplines and countries, even after years of investment. To support researchers, they developed an assessment tool to help researchers from diverse disciplines understand FAIR-related vocabularies, prepare their data, and measure progress towards compliance.

Although existing literature focuses on technical and policy aspects of FAIR and the challenges of preparing researchers for its implementation, less attention has been given to the practical evaluation of datasets already shared before FAIR principles became part of funding agency requirements. As this project was completed prior to the implementation of current federal data-sharing requirements, this study contributes to addressing this gap by assessing an already shared dataset through the lens of the original FAIR principles, updating metadata, and identifying lessons for improving FAIR compliance and long-term data reusability.

## Methods

This paper examines how data shared from a completed federally funded research project, directed by the author and carried out by the research team, aligns with the FAIR principles in terms of data management and sharing practices. Specifically, this study addresses the following research questions:

1. How well does the shared data generated from the federally funded research project align with the FAIR principles?
2. What outcomes result from applying a self-evaluation of FAIR alignment?
3. What lessons and challenges emerge that can guide future data sharing practices?

### About the project:

The research project introduced in this paper is about developing assessment tools to explore the relationship between students' psychological factors and academic engagement, how students defined academic success and perceived the library's contribution to it, and the factors associated with students' GPA and their own definitions of success. The research team developed tools and studies aimed at helping academic libraries assess their impact on students' success and better supporting students' needs. The assessment tools, datasets, and instructional videos created through this project were scheduled for public release in summer 2024 via the Project Website (University of Illinois Chicago Library, n.d.-a)

### Data and data sharing:

Table 1 presents the data used for analysis in this paper, including the types of data and their DOIs/URLs for data sharing. The data include survey instruments, datasets, codebooks, and recordings. The survey instruments, referred to as assessment tools, consist of two components: a one-time survey and online weekly journals. These tools are designed to capture students' academic engagement and their personal definitions of academic success.

In the first phase of the project (Year 1: August 2022-July 2023), the assessment tools were developed by the research team and validated (Scoulas, Shotick, De Groot & Osorio, 2024) and then piloted at the University of Illinois Chicago (UIC), with data collection occurring in Spring 2023. The tools included both multiple-choice and open-ended questions. In the second phase (Year 2: August 2023-July 2024), the tools were implemented at both UIC and Northern Illinois University (NIU), with data collection in Spring 2024. The tools were revised after the implementation stage.

The datasets for this research project were collected in two phases: the pilot phase at UIC (2023), and the implementation phase at UIC and NIU (2024). Data were collected from the *Student Academic Engagement and Success (SAES) Survey* and the *Online Weekly Journal*, both of which were administered through Qualtrics. All data were de-identified and anonymized before being deposited in UIC's Institutional Repository, INDIGO (UIC Library, n.d.-b) in 2024. Deposited materials include datasets, codebooks, and survey instruments. Additionally, recorded instructional files (e.g., project overview and guidance on using the survey instruments) were created and made available via Youtube, with links those resources shared on the project website:

<https://libscholar.digital.uic.edu/assessment-tools/>.

### Data analysis guided by FAIR Principles:

The summary of the shared research materials is outlined in Table 1, and they were included in the analysis using FAIR principles (Wilkinson et al., 2016). The following steps were undertaken by the author:

1. Review of FAIR principles and familiarization with shared research materials
  - a. The four foundational FAIR Principles and their 15 guiding principles were reviewed.
  - b. Each of the shared research materials, including datasets, codebooks, instruments, and recorded instructional files, was examined to become familiar with its contents and to identify potential gaps or areas for improvement in FAIR alignment.
2. Structured evaluation
  - a. A structured spreadsheet was created with criteria for marking each element as “Met”, “Partially met”, or “No.”
  - b. All shared research materials were examined to evaluate the accompanying meta(data) against the FAIR principles, and the results were documented in the spreadsheet.
  - c. During this process, the author recorded observations, areas for revision, challenges, and lessons learned related to implementing FAIR principles.
  - d. These reflections were documented alongside the evaluation to capture practical insights and inform recommendations for future data sharing practices.
3. Updating metadata based on findings
  - a. Based on the evaluation and observations from the FAIR assessment, the author collaborated with the project team member who originally created the metadata in the institutional repository to update a detailed list of recommended revisions.
  - b. These revisions were implemented after completion of the research project to ensure that all metadata in the repository was accurate, complete, and aligned with FAIR principles.

Table 1. Data, types of data, and depository

Data	Data Type	Depository
Survey instruments (Assessment tools): SAES survey and OWJ-generic version <sup>2</sup>	Qualitative data (text)	UIC Institutional Repository/Project website
Datasets and codebooks	Pilot phase (Year 1): Quantitative and qualitative data (numerical data, open-ended responses): Student Academic Engagement Success survey dataset and Online Weekly Journal dataset Implementation phase (Year 2): Quantitative data: SAES survey dataset and OWJ dataset	UIC Institutional Repository/Project website
Recordings (instructional video)	Qualitative data (video and text) Research design Survey tools and data collection Data cleaning, analysis, and sharing  All three recordings are posted on the project website	Project website
Conference posters/publications	Qualitative data (text, visual content, and text descriptions)	Project website

## Results

Table 2 presents the details of how the data were evaluated based on the FAIR principles. Among the 15 Principles, 8 were fully met, 7 were partially met, and none were categorized as “not met.” The results are presented according to the four foundational FAIR principles.

### Findable:

This criteria evaluates whether the data have a globally unique identifier, are described with clear and rich metadata, and are indexed in a searchable repository. Under the Findable principle, F1 (assignment of globally unique and persistent identifiers) and F4 (registration or indexing in a searchable resource) were fully met, while F2 (rich metadata) and F3 (metadata including the data’s identifier) were partially met. For example, survey instruments and datasets/codebooks were assigned a unique identifier generated by the University’s IR, and the recorded workshops were hosted on YouTube, with the URL links being shared on the project website for easy access. For data related to conference presentations and publication documents, their citation records were shown on the project website. However, regarding whether the data were described with clear and rich metadata, they lacked key details for each data, such as a comprehensive project description. For example, when accessing the DOI for survey instrument, the instrument description included only the instrument name, funding source, and history (see Figure 1). Similarly, for the pilot and implementation phases,

each of the two datasets and codebooks was deposited separately in the institutional repository. While they displayed the survey name, citation information, authors, and funding information, but they did not provide essential metadata such as the project's purpose, timeline, response rate, or required software for analysis (see Figure 2 for pilot phase data). The lack of metadata and the names of datasets and phases differed, which made it difficult to recognize that they were part of related research projects. For the recorded workshops, only the title was provided, and there was no description of the content. However, the video was accompanied by the presentation slides, which could help serve as a substitute for more detailed context.

Figure 1. Example of survey instruments (assessment tools) shared in UIC IR

The screenshot shows the UIC IR interface. At the top, there is a navigation bar with the UIC logo, a search bar, and a 'Log in' link. The main content area features a title 'Student Success Assessment Toolkit - Assessment Tools' with a 'Follow' button and a 'Version 2' dropdown. Below the title, there is a description of the project's funding and purpose, a list of tools created by the project, and a link to the survey tool. On the right side, there are 'USAGE METRICS' (385 views, 0 citations), 'AUTHORS (4)' (Jung Mi Scoulas, Sandra De Groote, Kimberly Shotick, Nestor Osorio), 'CATEGORIES' (Education assessment and evaluation), and 'KEYWORDS' (student success, Academic libraries). At the bottom, there is a 'CITE THIS COLLECTION' section with a 'DataCite' dropdown and a 'Copy citation' button. The DOI link is provided as <https://doi.org/10.25417/uic.c.7096387.v2>.

Figure 2. Example of datasets of Student Academic Engagement and Success Survey during the Pilot Phase at UIC shared in UIC IR

The screenshot displays a dataset page with the following elements:

- Title:** Student Academic Engagement and Success Survey - Pilot Data
- Actions:** Cite, Download all (1.2 MB), Share, Embed, + Collect
- Version:** Version 2 (dropdown arrow)
- Metadata:** Dataset posted on 2024-03-25, 09:20 authored by Jung Mi Scoulas, Sandra De Groote, Kimberly Shotick, Nestor Osorio
- Usage Metrics:** 620 views, 110 downloads, 2 citations
- Description:** A pilot of the The Student Academic Engagement and Success Survey was distributed in Spring 2023. The anonymized data set and supporting documentation are attached. Data is an .sav file (SPSS).
- Funding:** IMLS LG-252338-OLS-2
- History:**
  - 2024-03-23 - Publication date
  - 2024-03-24 - First online date
  - 2024-03-25 - Posted date
- Categories:** No categories selected
- Keywords:** student success, academic libraries, pilot data
- Licence:** In Copyright
- Exports:** Select an option (dropdown menu)

### Accessible:

This criterion primarily focuses on data (metadata) being retrievable via a standardized, open, and universal protocol that supports authentication, with metadata remaining accessible even if the data are no longer available. The results indicate that all meta(data) and protocol meet the accessibility criteria, with three of the four principles fully met. However, one principle (A 1.2: the protocol allows for authentication and authorization where necessary) was only partially met. When clicking on the provided URLs, both the survey instruments and datasets were accessible. Nonetheless, it is important to explicitly state any access restrictions for data deposited in the IR. Although no such restrictions are currently mentioned, users can download the available datasets without authentication.

The primary issue affecting data accessibility was the inconsistency in the DOIs. For example, the datasets included both pilot data (Year 1) and implementation data (Year 2), with each dataset containing a one-time survey and an online weekly journal. In the pilot stage, separate DOIs were provided for the one-time survey and the weekly journals. In contrast, the implementation stage datasets used a single DOI for both components. This inconsistency may confuse users and hinder their ability to locate and access the correct files.

Additionally, during the evaluation of the dataset, it was identified that one variable had not been removed. While this variable did not contain directly identifiable information, there is a potential risk that, when combined with other data, it could potentially reveal users' identities. The datasets for the pilot phase were immediately corrected and re-uploaded to the IR.

### Interoperable:

This principle emphasizes that (meta)data should use a formal, accessible language for knowledge representation, follow FAIR principles, and include qualified references to other metadata. The results show that two of the three criteria were fully met, while one criterion (I3: metadata include qualified reference to other metadata) was only partially met. That is, all data, including survey instruments, datasets, and conference/publication materials, used standardized taxonomies such as DataCite, with the exception of instructional videos. However, these materials were deposited individually, without appropriate references linking them to related components. For instance, when accessing the datasets, the corresponding survey instruments were not included or referenced. Similarly, the instructional videos created to explain how to use the tools were hosted on the project website but were not linked to the survey datasets and instruments. Furthermore, conference materials and publications, which provide supplementary context, such as project background, tool usage, and analysis, were not cross-referenced within the deposited data files.

### Reusable:

This criterion ensures that metadata are richly described with accurate and relevant attributes, accompanied by a clear data usage license, detailed provenance, and adherence to domain-relevant community standards. Under the Reusable principle, the results showed that three of the four criteria were partially met, while only one was fully met. For example, for datasets and codebooks, all variables were clearly defined with their values and how they are assigned. However, the dataset's purpose was not clearly indicated in the description. Additionally, the data were licensed differently. For example, the survey instruments were shared under a Creative Commons license (CC BY-NC-SA 4.0) (Figure 1), which allows users to adapt and redistribute the materials for non-commercial purposes with proper attribution (Creative Commons, n.d.), whereas the datasets were shared under copyright (Figure 2), which restricts use, adaptation, and distribution without explicit permission from the copyright holder (U.S. Geological Survey, n.d.-b). This difference in licensing was driven by practical considerations during the data sharing process, and reflected an understanding at the time that both approaches would support accessibility. However, it is possible that a distinction in licensing may cause some confusion among users regarding what types of reuse are permitted.

There was also inconsistency in adhering to community standards, particularly regarding standardized survey data formats. For instance, the survey instrument was provided in two versions, CSV and QSF, where QSF allows users to directly upload the survey into Qualtrics, an online survey platform. However, while the datasets for Year Two (implementation stage at UIC and NIU) adhered to community standards by providing both Excel and SPSS formats for data analysis, the Year One (pilot data) datasets were only available in SPSS format. The absence of CSV or Excel versions for the pilot data did not meet the community standards (Figure 2). Additionally, while the instructional recordings are publicly available on YouTube and could be shared for viewing, they lack clear licensing and citation information, which may cause confusion about whether and how the content can be reused or formally cited.

Overall, the meta(data) met most of the FAIR principles, with notable adherence to the findable and accessible criteria, such as the use of unique identifiers and publicly accessible metadata. However, there were gaps in interoperability and reusability, as some metadata lacked comprehensive

descriptions, and the datasets inconsistently adhered to community standards. Additionally, while licenses and version history information (tracking changes over time) were applied inconsistently, there is room for improvement in providing more accessible data formats.

## Critical takeaways

The process of self-evaluating shared data using the FAIR principles provides valuable lessons and helps identify improvements in both the current data sharing practices and future data management strategies. Below are the key takeaways and actionable insights for researchers on strengthening future data sharing efforts and reusability.

### Takeaway 1: FAIR principles apply to all data types and support the inclusion of diverse research outputs.

Originally, the author's intention was to examine only datasets that contained both qualitative and quantitative data, were collected during the pilot phase, and were shared, guided by the FAIR guidelines in this paper. However, during the process, it became clear that the datasets were connected to various other data types, such as survey instruments, recorded videos explaining data analysis procedures, and conference/publication materials, that provide insights into the data analysis and key findings. These connections underscore the importance of considering a broader range of data types, aligning with the intent of the FAIR principles. Although this case study primarily involves quantitative data, the principles and challenges addressed, such as ensuring data usability, transparency, and long-term access, are equally applicable to qualitative datasets. The findings illustrate how FAIR principles can guide ethical and effective data sharing practices across data types and research outputs.

### Takeaway 2: The importance of clear metadata and data linking for enhanced discovery and usability.

One major takeaway is the importance of providing clearly described metadata for all shared data. Although metadata was provided separately for each data type, such as datasets and survey instruments, it was often insufficient for users to fully understand the context, intended use, and background of the data. For example, while survey instruments and datasets included descriptions of their content and usage, they lacked essential details, such as the purpose of the research project, the development of the instruments, and the methods used for data analysis. Although this information was available on the project website, the metadata for survey instruments was not linked to the corresponding datasets. This disconnection of descriptions and data may hinder users, especially those accessing data from multiple sources, from effectively using the data.

In addition to traditional data types (survey instruments and datasets), supplementary materials are often overlooked despite their potential value. In this case, instructional recordings were created to guide other researchers on using the assessment tools and analyzing the data. However, they lacked accompanying metadata that clearly described their content and purpose. As a result, users may need to watch the videos to understand their purpose, as the videos did not provide clear descriptions of their content. While these videos provide valuable explanations of the project background, survey instrument usage, and data analysis, they were not linked to other datasets. These videos could serve as an alternative form of metadata, but clearer, more concise descriptions are needed to enhance

their usefulness for researchers. Without proper descriptions, the potential for researchers to fully recognize the value of data reuse is diminished, limiting both discoverability and practical application.

### **Takeaway 3: Using consistent DOIs across platforms to enhance user experience.**

Another key takeaway is the importance of using a consistent DOI for data across various channels, such as publications and project websites, to avoid confusing users. The evaluation revealed that the DOI for datasets in publication differed from the one on the project website. This issue underscores the critical importance of maintaining consistent DOIs across all research outputs. The DOI discrepancy on the project website was addressed during the evaluation process, highlighting the need for greater attention to detail in linking data sources consistently. Even with a recorded revision history, some users may have already accessed the data before updates were made. For instance, more than 400 users had already downloaded the assessment tools as of March 18, 2025. These users likely accessed the data through a workshop or conference presentation (Soulas et al., 2024), where they received guidance on how to use the data. Although these attendees may not require detailed information, it is still important to provide updates or clarification especially for users accessing the data through other channels.

### **Takeaway 4: Careful examination of the data for confidentiality.**

Ensuring respondents' confidentiality is crucial when sharing data. During a dataset review, it was found that respondents could potentially be identified because one variable was identifiable when combined with other demographics, such as class level. Upon identifying this potential privacy risk, the author took immediate action to remove the variable, preventing any unintended exposure. Additionally, qualitative data from open-ended responses were carefully reviewed for privacy concerns and were confirmed to meet confidentiality standards. All updated datasets and codebooks for the pilot phase were updated. This highlights the need for careful consideration of data anonymization to ensure respondents' privacy and confidentiality.

Taken together, the author's primary focus in this case study on revisiting shared data, identifying gaps, and addressing those issues demonstrates that data sharing is not a one-time task. Rather, it involves ongoing reflection and improvement. The author critically examined their own practices by working with the research team and made necessary corrections with future and potential users in mind, emphasizing a user-centered approach. After evaluating the shared data using the FAIR principles, the research team made a significant change: all relevant data were consolidated into a collection (Soulas, De Groote, Shotick, Nestor, 2024). The collection includes assessment tools, datasets and codebooks, instructional videos, and publications as a complete package. Within the collection, each item (such as assessment tools and datasets) is assigned its own DOI, and all associated metadata, including project descriptions and relevant references, are clearly documented. This structure allows users to view all related materials in one place while still enabling access to and citation of individual datasets when sharing or using the data.

Table 2. Self-evaluation of Shared Data guided by FAIR principles

Principles	Criteria	Self-evaluation
<b>FAIR principles:</b> to improve the reusability of scientific data, ensuring it is both discoverable by machines and reusable by other researchers.		
<b>Findable</b>		
	F1. (meta)data are assigned a globally unique and persistent identifier	Met
	F2. data are described with rich metadata (defined by R1 below)	Partially Met
	F3. metadata clearly and explicitly include the identifier of the data it describes	Partially Met
	F4. (meta)data are registered or indexed in a searchable resource	Met
<b>Accessible</b>		
	A1. (meta)data are retrievable by their identifier using a standardized communications protocol	Met
	A1.1 the protocol is open, free, and universally implementable	Met
	A1.2 the protocol allows for an authentication and authorization procedure, where necessary	Partially Met
	A2. metadata are accessible, even when the data are no longer available	Met
<b>Interoperable</b>		
	I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation.	Met
	I2. (meta)data use vocabularies that follow FAIR principles	Met
	I3. (meta)data include qualified references to other (meta)data	Partially Met
<b>Reusable</b>		
	R1. meta(data) are richly described with a plurality of accurate and relevant attributes	Partially Met
	R1.1. (meta)data are released with a clear and accessible data usage license	Partially Met
	R1.2. (meta)data are associated with detailed provenance	Met
	R1.3. (meta)data meet domain-relevant community standards	Partially Met

FAIR Principles adapted from Wilkinson et al. (2016)

## Conclusion

Conducted under a tight project timeline (July 2024) and prior to the effective date of the IMLS Public Access Policy (applicable to awards made one or after October 1, 2025), this project originally focused on meeting general data management requirements rather than full FAIR principles implementation. Unlike projects funded by agencies that adopted these standards earlier, applying the FAIR framework retrospectively in this study demonstrates how such principles can be meaningfully integrated even when they were not originally required. As David et al. (2020) emphasize, FAIR principles do not explicitly address data quality, which remains a persistent challenge for data reuse. In this project, however, FAIR principles were deliberately applied not only to ensure findability, accessibility, interoperability, and reusability, but also to enhance the overall quality, usability, and reusability of the data. This experience demonstrates that, even when FAIR compliance is not initially required, these principles can meaningfully guide efforts to make data more robust and reusable for future research. It highlights that FAIR guidelines function as a tool for refinement rather than a constraint, even for qualitative data. Ensuring that various types of data are easily discoverable, accessible, and accompanied by clear descriptions increases the likelihood that more researchers will engage with the findings and ultimately reuse them.

The gaps identified through the FAIR principles' evaluation have been addressed, including improving metadata clarity, specifying project goals, and consolidating relevant links. As the project director, the author's goal is to encourage researchers to reuse the data and contribute to ongoing findings, in alignment with the project's original aim, supporting others in using these tools to address student success within their own institutional contexts.

Under a subsequent grant funded by the Institute of Museum and Library Services during 2024-2026 (Grant Number: LG-256587-OLS-24), a workshop series for academic librarians on how to apply these tools effectively was offered from April through June 2025. During this workshop, the updated data were introduced to participants, with the aim of enhancing their ability to understand, access, and apply the data according to their needs. Additionally, this updated data will be presented at an upcoming conference to reach a broader audience. While careful preparation of data before sharing is ideal, this case study demonstrates that it is still possible to identify and correct gaps for future users. Additionally, it serves as a resource for information professionals navigating the complexities of data sharing and the responsibilities of maintaining data quality, transparency, and usability over time.

## Acknowledgement

The research project introduced in this paper was made possible in part by the support of the Institute of Museum and Library Services (Grant Number: LG-252338-OLS-22). The author would also like to acknowledge the Co-Principal Investigator, Sandra De Groote, whose contributions were essential to implementing the changes throughout the project.

## References list

Creative Commons. (n.d.). *About the licenses*. <https://creativecommons.org/licenses/>

- Dankar, F.K., (2023). *Practices and challenges in clinical data sharing* (arXiv:2304.06509). Cornell University. <https://doi.org/10.48550/arXiv.2304.06509>
- David, R., Mabile, L., Specht, A., Stryeck, S., Thomsen, M., Yahia, M., Jonquet, C., Dollé, L., Jacob, D., Bailo, D., Bravo, E., Gachet, S., Gunderman, H., Hollebecq, J. E., Ioannidis, V., Bras, Y. Le, Lerigoleur, E., & Cambon-Thomsen, A. (2020). Fairness literacy: The Achilles' heel of applying FAIR principles. *Data Science Journal*, 19(1), 1–11. <https://doi.org/10.5334/dsj-2020-032>
- Institute of Medicine (US). (2013). *Sharing clinical research data: Workshop summary*. Washington, DC: National Academies Press (US). <https://www.ncbi.nlm.nih.gov/books/NBK137823/>
- Institute of Museum and Library Services (n.d.) *Public access*. <https://www.imls.gov/about-us/additional-resources/policy-notice/public-access>
- Institute of Museum and Library Services. (2022). *Developing and testing assessment tools for measuring library impact on student academic success* (Grant No. LG-252338-OLS-22). <https://www.imls.gov/grants/awarded/lg-252338-ols-22>
- Institute of Museum and Library Services. (2024). *Implementing assessment tools for library impact on students' academic success* (Grant No. LG-256587-OLS-24). <https://www.imls.gov/grants/awarded/lg-256587-ols-24>
- Jacobsen, A., de Miranda Azevedo, R., Juty, N., Batista, D., Coles, S., Cornet, R., Courtot, M., Crosas, M., Dumontier, M., Evelo, C. T., Goble, C., Guizzardi, G., Hansen, K. K., Hasnain, A., Hettne, K., Heringa, J., Hooft, R. W. W., Imming, M., Jeffery, K. G., ... Schultes, E. (2020). FAIR Principles: Interpretations and implementation considerations. *Data Intelligence*, 2(1–2), 10–29. [https://doi.org/10.1162/dint\\_r\\_00024](https://doi.org/10.1162/dint_r_00024)
- Kemp, D., Nelson, A. H., Jenkins, D., (2023). *Yes, no, maybe? Legal & ethical considerations for informed consent in data sharing and integration*. Actionable Intelligence for Social Policy. University of Pennsylvania, <https://aisp.upenn.edu/wp-content/uploads/2023/05/Consent-Brief-Appx.pdf>
- Mello, M. M., Lieou, V., & Goodman, S. N. (2018). Clinical trial participants' views of the risks and benefits of data sharing. *The New England Journal of Medicine*, 378(23), 2202–2211. <https://doi.org/10.1056/NEJMsa1713258>
- Mons, B., Schultes, E., Liu, F., & Jacobsen, A. (2020). The fair principles: First generation implementation choices and challenges. *Data Intelligence*, 2(1–2), 1–9. [https://doi.org/10.1162/dint\\_e\\_00023](https://doi.org/10.1162/dint_e_00023)
- National Science Foundation. (2024). *Proposal & award policies & procedures guide (PAPPG), NSF 24-1*. <https://www.nsf.gov/policies/pappg/24-1/ch-2-proposal-preparation#ch2D2i-ii>

- National Institutes of Health (n.d). *Scientific data sharing*, <https://sharing.nih.gov/data-management-and-sharing-policy/sharing-scientific-data>
- Piowar, H. A., Day, R. S., & Fridsma, D. B. (2007). Sharing detailed research data is associated with increased citation rate. *PLoS One*, 2(3), e308–e308. <https://doi.org/10.1371/journal.pone.0000308>
- Sadeh, Y., Denejkina, A., Karyotaki, E., Lenferink, L. I. M., & Kassam-Adams, N. (2023). Opportunities for improving data sharing and FAIR data practices to advance global mental health. *Cambridge Prisms: Global Mental Health*, 10. <https://doi.org/10.1017/gmh.2023.7>
- Scoulas, J. M., Shotick, K., De Groote, S. L., & Osorio, N. L. (2024). Developing and validating assessment tools for measuring undergraduate students' academic engagement and academic achievement: lessons learned from a pilot research project. *Journal of Library Administration*, 64(2), 243–251. <https://doi.org/10.1080/01930826.2024.2305075>
- Scoulas, J. M., De Groote, S. L., Shotick, K., & Osorio, N. L. (2024). *Student success assessment toolkit – tools, data and instructional videos*. University of Illinois Chicago. Collection. <https://doi.org/10.25417/uic.c.7241047>
- U.S. Geological Survey. (n.d.-a). *Why share your data?* <https://www.usgs.gov/data-management/why-share-your-data>
- U.S. Geological Survey. (n.d.-b). *Data licensing*. <https://www.usgs.gov/data-management/data-licensing>
- University of Illinois Chicago Library. (n.d.-a). *Student Success Assessment Toolkit*. <https://libscholar.digital.uic.edu/assessment-tools/>
- University of Illinois Chicago Library. (n.d.-b). *INDIGO: Institutional Repository*. <https://indigo.uic.edu/>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. 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). Comment: The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3, 1–9. <https://doi.org/10.1038/sdata.2016.18>
- White House Office of Science and Technology Policy. (2022, August 25). *Ensuring free, immediate, and equitable access to federally funded research*. <https://bidenwhitehouse.archives.gov/wp-content/uploads/2022/08/08-2022-OSTP-Public-Access-Memo.pdf>

## Endnote

---

<sup>1</sup> Associate Professor and Assessment Coordinator, University Library, University of Illinois Chicago, jscoul2@uic.edu

<sup>2</sup> Generic versions can be customized by users to fit their specific needs and institutional contexts.