



(RESEARCH ARTICLE)



Data management plans facilitate effective STEM career development programs

Lawrence O. Flowers *

Department of Biological and Physical Sciences, Saint Augustine's University.

International Journal of Science and Research Archive, 2025, 17(01), 1290–1295

Publication history: Received on 05 September 2025; revised on 10 October 2025; accepted on 13 October 2025

Article DOI: <https://doi.org/10.30574/ijrsra.2025.17.1.2807>

Abstract

Data management plans are indispensable for undergraduate career development programs. Guided by FAIR principles, data management plans provide the blueprint specifying how data derived from programmatic implementation will be collected, managed, disseminated, shared, and stored. Science, technology, engineering, and mathematics (STEM) career development programs in college have been instrumental in helping undergraduate and graduate students achieve occupational goals. STEM career development program objectives typically include enhancing students' academic and professional skills by increasing their exposure to real-world experiences in graduate school and the workforce. Data management plans are essential elements for career development programs. Data management plan characteristics must be resolved in the early stages of program development by the program team. Three major types of data originate during the implementation of many STEM career development programs: program data, social science data, and STEM research data. Effective data management plans lead to the production of reusable data that informs evidence-based practices. Include data management plan operating costs in the program budget and assign project team members to carry out specific components of the data management plan. Formative evaluations enable modifications to the data management plan, improving efficiency. Advances in artificial intelligence (AI) platforms reduce problems associated with data management plan administration and enhance collaboration among program personnel responsible for the successful execution of the plan. The purpose of this article is to present practical information for developing a data management plan for STEM career development programs.

Keywords: Data management; FAIR; Metadata; STEM career development program; Institutional review board

1. Introduction

Higher education STEM career development programs are initiatives designed to prepare college matriculants for challenging and rewarding careers [1]. Data management plans are critically crucial for the success of career development programs and are thus typically required components of funding organizations [2-3]. First and foremost, FAIR principles must guide the formation and operation of a data management plan. FAIR implies that protocols are in place to make the data findable, accessible, interoperable, and reusable [4]. While the specifics of a STEM intervention program's data management plan will differ, all data management plans must operate within the FAIR framework. FAIR principles enable the project team and the external research community to quickly find and utilize archived data, allowing for the reproduction of studies and thereby validating data-derived conclusions. Interoperability is a key tenet of the FAIR data policy, referring to data that is curated using optimal software, systems, and canonical procedures to enhance data exchange and utilization, and mitigate system and software compatibility issues that hinder data sharing goals. Most federal funding agencies require a comprehensive data management plan and expect researchers to freely share data with colleagues and interested parties rather than hoard data in perpetuity.

* Corresponding author: Lawrence O. Flowers

Reaching consensus in the data management plan development process is essential. Data management plans suffer when the construction process is limited to one or two members of the STEM career development program team. Data comes in many forms. The type of data collected and managed is determined in part by the career development program's goals and objectives. Three main broad types of data exist for STEM career development programs, including social science data, program data, and STEM research data. Program data is the most comprehensive and implementation-specific category. Any aspect of the program, such as recruitment data, evaluation data, program website statistics, pedagogical products, and student deliverables, is characterized as program data. Social science data refers to any data produced from qualitative or quantitative research studies. Most career development programs at the college level collect data on program participants' academic and professional outcomes to ensure the program is achieving its stated goals and to share the data with various stakeholders. STEM research data refers to data generated from laboratory experiments, engineering and math projects, or coding projects (e.g., artificial intelligence, software development, website design, robotics). STEM career development programs often include faculty-mentored research training components, which can produce copious amounts of STEM research data. If stored and preserved correctly, STEM research data can facilitate the dissemination of publications, scientific presentations, and potential use by other researchers. Since the program team may not be experts on every experimental or developmental protocol employed during the career development program, it is prudent for the program director to consult each faculty undergraduate research mentor to identify expected types of STEM research data. Consider employing digital object management systems to enhance the management and storage of digital information. A Digital Object Identifier (DOI) is a persistent, unique code of numbers, letters, and other characters assigned to physical or virtual content (e.g., articles, books, abstracts, documents) that aids in the identification of digital content.

Several key ethical issues are essential when developing a data management plan. The primary concern for STEM career development programs that intend to utilize program participants as human subjects in educational research studies is to seek Institutional Review Board (IRB) approval from your institution. It is a good idea to contact the IRB director at your institution before writing an IRB application and data management plan to seek guidance and avoid potential downstream problems. The IRB on campus is typically composed of faculty, staff, and administrative experts who ensure that the research study complies with ethical requirements, risks and benefits standards, and human subject informed consent regulations. Seek IRB approval early in the career development program planning stage because, depending on the institution, the IRB approval process may take weeks or months to obtain. Focus on the student data anonymization procedures during the data organization and storage phases. Collect data in such a way that it poses no or minimally accepted risks, as determined by the IRB committee, using quantitative and qualitative survey instruments that are high in reliability and validity. The purpose of the article is to present a strategy for the development of a data management plan appropriate for use in STEM career development programs. The article also presents a data management life cycle for new faculty and principal investigators. Data management plans are typically required components for federal grant proposals in the United States (e.g., National Institutes of Health and National Science Foundation) [5]. Despite their importance in grant proposals, there is no precise description detailing the exact components of a comprehensive data management plan. Williams et al. [6] offer helpful insights regarding data management plan development.

2. Data Management Plans for STEM Career Development Programs

Data from STEM career development programs are vital for the reproducibility of student and program outcomes, as well as the sustainability of program aims. Data management plan components may vary depending on the nature of the career development program's elements and goals. However, all data management plans operate on a similar framework. Several personnel from the STEM career development program should supervise the development and maintenance of the data management plan. Program directors must allocate sufficient funding to cover the costs of human and technological (e.g., hardware, software) resources necessary to implement the data management plan. Figure 1 displays a schematic of a generalized data management life cycle. Search for data management life cycle examples and plan templates that may share greater similarity with your specific program.

2.1. Data Plan

The first stage of the data management lifecycle is the planning stage. In this stage, develop a comprehensive understanding of the data management plan requirements (e.g., format and template) unique to the funding organization. Assign program team members to conduct various stages of the data management life cycle and consider adopting project management and data management software. Develop specific timelines and benchmarks for life cycle stages. Artificial intelligence platforms are emerging as helpful supplemental tools for project management, enabling the tracking of roles and responsibilities for each member of the program team, including the administration of data management plans. These platforms are typically equipped with reminders and communication features to

ensure completion of data management tasks. Plan intermittent internal and external evaluations to adjust ineffective data management plan procedures. Create a schematic visualization to illustrate the timeline, stages, and components of the data management plan. Process figures serve as an excellent visual aid to enhance comprehension and keep the team focused on achieving various objectives. During the data management planning stage, utilize the STEM career development program's logic model to explore potential program components and their corresponding data outputs. The logic model, among other benefits, provides a blueprint for communicating career development program resources, activities, and outcomes. It also serves as a framework for identifying data production, which is essential information for the data management plan.

2.2. Data Collection

During the planning stage, outline the data collection procedures, the timeframe, and the program supervisors responsible for data collection. Data collection procedures are specified in the methodology section in STEM career development program proposals that utilize social science research designs to study program participants using qualitative, quantitative, or mixed methods. Program data, including participant characteristics, program deliverables, website analytics, evaluation data, and descriptive photos, are collected from the program team throughout the program's implementation. Faculty research mentors, who supervise student research projects, collect STEM research data during the program.

2.3. Data Preservation

In the data management life cycle presented in this article, data storage and data preservation strategies take place in stage three. Data storage primarily involves the location and deposition of program data files in storage drives or web-based locations, which facilitate the quick recovery of information. Data preservation focuses on conserving data completeness and integrity over long periods. Data preservation techniques provide safeguards for data files and prevent data corruption, ensuring program data is readily available in a user-friendly format for an extended time. Preservation strategies include employing regular data monitoring, automated data redundancy, data migration, and metadata. Select data preservation software and platforms that are readily available and accessible to program team members and collaborators. Due to their accessibility and affordability, Google data preservation platforms are cost-effective, safe from external cyber threats, and maintain data fidelity for a prolonged timeframe. Also, take advantage of trusted electronic data preservation repositories such as Dryad to support data preservation objectives. Metadata is a general term that refers to data used to characterize other data and can fall into two primary categories: descriptive metadata and technical metadata. Descriptive metadata is crucial for STEM career development program directors. It also provides an opportunity for data management plan administrators to label files and folders with additional explanatory information, improving the ability to access data. Descriptive metadata consists of the date of origin, file type, data generation location, program title, administrator name, storage location, and dissemination strategy. Utilization of text files (.txt, .csv), document files (.doc), image files (.jpeg), video files (.mp4, .mov), and Excel files (.xlsx) is widespread and acceptable for data storage and preservation, enhancing downstream accessibility.

2.4. Data Security

It is essential to protect STEM career development program data from prohibited access and unapproved utilization. The use of password-protected computer systems and firewalls can diminish the risk of data theft and cybersecurity threats. Frequent data backup procedures for electronic data files will reduce the likelihood of data loss from external cyber threats. If feasible, use program funds to acquire cybersecurity software and an experienced cybersecurity analyst to protect highly sensitive data. Employing strong passwords with limited access and virtual private networks (VPNs) are helpful strategies to fortify program data security.

2.5. Data Analysis

The methodology section from the career development grant proposal, research proposal, or research publications will inform the data analytic techniques used to analyze program data, social science data, and STEM research data generated during program activities. During this stage of the data management life cycle, use robust, evidence-based measurement techniques to determine whether program data intersects with program objectives and to interpret novel findings that may impact STEM education and research communities.

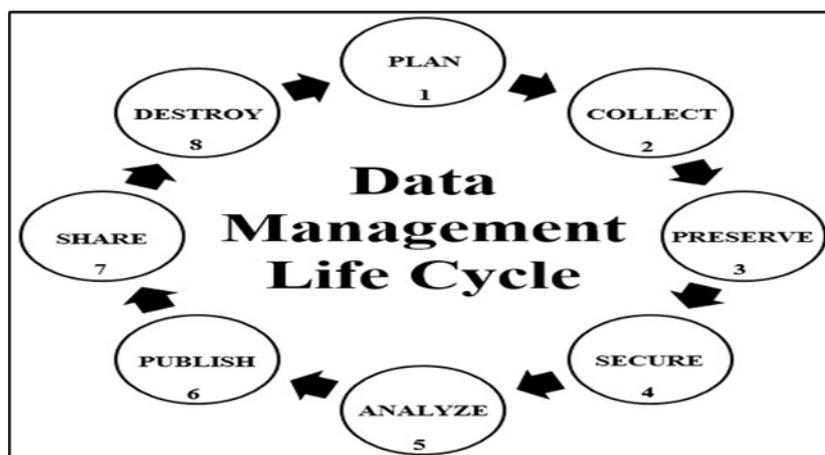


Figure 1 Data management life cycle for STEM career development programs

2.6. Data Dissemination

Publishing program-generated data in research and education journals and newsletters is the most common way to distribute data to a broad audience and inform interested parties. Many STEM career development programs also have a website dedicated to presenting the program's strategic plan, contact information, and deliverables. This website can also serve as a platform to share data with the research community. Presenting data on STEM career development programs at regional, national, and international conferences is another standard method of disseminating interesting program-focused findings. Disseminating program deliverables (e.g., lab exercises, computer code) on the National Science Digital Library (NSDL) is also a safe and secure distribution strategy.

2.7. Data Sharing

Publicly funded grants typically mandate the inclusion of a data-sharing policy in grant proposals. The careful articulation of data sharing provisions is a required component in data management plans. Some programs choose to use their website as a repository for raw data related to the career development intervention program, making that data available to the public. Program websites that serve as data repositories must withstand potential external cyber threats and include data intrusion alerts, data transmission features, and a data request form. Contact legal representation for guidance in the development of the data request form that specifies acceptable uses and intellectual property rights of STEM career development program data.

2.8. Data Destruction

After data has been utilized for its intended purpose or a predetermined time has elapsed, discard unnecessary data from password-protected computers to free up disk space and increase storage capacity. Removing superfluous data will also reduce data storage expenses. Basic data wiping procedures and even physical destruction methods, including shredding and degaussing, are ideal for sensitive data. The National Institute of Standards and Technology (NIST) provides helpful guidelines for the safe destruction of unwanted data.

3. Conclusion

According to extensive literary evidence, STEM career development programs are vastly beneficial workforce preparation opportunities for college students. STEM career development programs help students get jobs via informative seminars, workshops, research experiences, and internships. Students acquire career development skills through scientific presentations, mock interviews, and instruction on writing resumes and cover letters. STEM career development programs generate different types of data of interest to the program team, the funding agency, and other stakeholders, which require the use of carefully constructed data management plans. Data management plans are crucial components of STEM career development programs. Data management plans encompass intricate processes to ensure that data is stored, shared, and disposed of at appropriate time points during and after the career development program. Data management plans describe the who, what, when, where, and how regarding the management, stewardship, and preservation of all types of data generated from programmatic implementation or research experimentation [7]. Before developing the data management plan, check with the funding agency to determine if a specific data management plan template or format is required.

The establishment and execution of proven data management plan activities are fundamental to monitoring program participants and primary outputs. One mistake made by program directors is the underappreciation of the importance of the data management plan when planning STEM career development programs at the college level. It is essential to assign clear responsibility roles for the specific data management procedures. It is also necessary to consider strategies for addressing career development program team attrition and assigning backup roles to ensure a continuous data management process. Annual or biannual program evaluations will result in modifications to the data management plan. Utilize metadata and avoid the use of an unsystematic file naming system to improve data organization and data file structure. Descriptive metadata and web-based applications will enhance data recovery after the program has ended [8]. Develop iron-clad contracts regarding the sharing, access, and reuse of data to ensure that intellectual property rights are clearly defined and that external users understand the acceptable platforms for using and disseminating data from program implementation and research projects.

Data management planning is not an intuitive skill for most principal investigators. Even successful principal investigators who have constructed numerous grant applications and secured many grants spend significant time contemplating the most effective strategies for collecting, storing, and sharing data. The development of traditional face-to-face or online workshops to train faculty, principal investigators, and grant writing teams would be instrumental in creating quality data management plans that can withstand scrutiny from grant review panels. Frequent data management training is also essential due to the rapid advancements in research data management procedures, bibliographic technology, and artificial intelligence platforms that can simplify the burden of data curation and community access. While students may not initially see the value of data management training while pursuing an undergraduate degree, college students who are on track to pursue graduate degrees or careers in research would also benefit from data management training.

STEM faculty should consider incorporating data management training elements into courses that include a laboratory component or in cases where students generate original data within the course framework. Incorporate research data management learning modules into course-based undergraduate research experiences (CUREs). Given the reported benefits to student academic outcomes and potential future occupational impacts, teaching students how to manage research data during CUREs projects may enhance learning gains [9-10]. Additionally, an understanding of long-term data preservation techniques would benefit graduate students who wish to use data generated during their graduate studies as early-career faculty to establish a research program or serve as preliminary data in grant proposals. Studies designed to acquire contemporary data management plan strategies from current program evaluators and principal investigators are beneficial due to frequent changes in data management policies and technology. This article outlines a general data management life cycle; however, it is advisable to search for articles and reports that provide more detailed information on the data management plan.

Compliance with ethical standards

Acknowledgments

This work was supported by grants from the National Science Foundation (EES - 2205612, EES - 2232563, EES - 2306512).

References

- [1] White E, Washington E, Flowers L. STEM career readiness at HBCUs to enhance occupational diversity. *Journal of Education & Social Policy*. 2021; 8: 34-38.
- [2] Kanza S, Knight N. Behind every great research project is great data management. *Biomed Central Research Notes*. 2022; 15: 1-5.
- [3] Michener W. Ten simple rules for creating a good data management plan. *PLOS Computational Biology*. 2015; 11: 1-9.
- [4] Wilkinson M, Dumontier M, Aalbersberg I, Appleton G, Axton M, Baak A, et al. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*. 2016; 3: 1-9.
- [5] Gonzales S, Carson M, Holmes K. Ten simple rules for maximizing the recommendations of the NIH data management and sharing plan. *PLOS Computational Biology*. 2022; 18: 1-16.
- [6] Williams M, Bagwell J, Zozus M. Data management plans: The missing perspective. *Journal of Biomedical Informatics*. 2017; 71: 130-142.

- [7] Pellen C, Munung N, Armond A, Kulp D, Mansmann U, Siebert M, et al. Data management and sharing. *Journal of Clinical Epidemiology*. 2025; 180: 1-6.
- [8] Jacob D, Ehrenmann F, David R, Tran J, Mirande-Ney C, Chaumeil P. An ecosystem for producing and sharing metadata within the web of FAIR data. *GigaScience*. 2025; 14: 1-8.
- [9] Zhang J, DeChenne-Peters S, Hecht D, Wolyniak M, Kuhn M, Koletar C, et al. Course-based undergraduate research experience impacts on student outcomes at minority-serving community colleges. *Biochemistry and Molecular Biology Education*. 2025; 53: 253-264.
- [10] Liu J, Benning C. A plant mutant screen CURE integrated with core biology concepts showed effectiveness in course design and students' perceived learning gains. *Biochemistry and Molecular Biology Education*. 2025; 53: 57-69.