

Applying Front-end Analysis Insights in the Development of a Data Management Course for Biomedical Researchers

Gronseth, S.L. , Ugwu, L. O. , Beech, B. M. , Bruce, M. A. , Siddiqui, A. , Itani, A. , & Kakadiaris, I. A.

curriculum development

design precedent

FAIR principles

learner analysis

research data management

In this article, we describe how our design team used front-end analysis to identify the training needs, gather design precedents, conduct a learner analysis study, and apply the insights in designing a data management course for biomedical researchers. In-person and online versions of the course were developed and implemented with cohorts of a research training and mentorship program for biomedical researchers. The course development project adds to the emerging literature base of design precedents for instruction

on research data management guided by the FAIR principles of findability, accessibility, interoperability, and reusability of research data.

Introduction

Storing and sharing data securely has become essential for current researchers and those desiring to enter the field. Securely sharing data, or making it more *open*, increases transparency, reduces redundancy, and enables other researchers to confirm and refine study methods and findings (Dijkers, 2019; Eaker, 2020). Managing large datasets poses challenges for researchers, and researchers are increasingly turning to data science and artificial intelligence (AI/ML) tools to harness capabilities for expansive processing and analysis (Miksa et al., 2019). However, to harness AI/ML tools for data analysis, researchers need to prepare data to be computer-read and interpreted while also thoughtfully addressing the extent to which data will be made openly available for research community access. These considerations are formalized through developing a research project data management plan.

Data management plans have become an expected component of external funding applications (National Institutes of Health, 2020; Williams et al., 2017). As part of the more significant open science movement, researcher competency in data sharing can foster “a scientific ecosystem in which research gains increased visibility, is shared more efficiently, and is performed with enhanced research integrity” (Open Science Skills Working Group of the European Commission, 2017, p. 5). That is why funding agencies and the research community have recognized the educational need for courses and other designed learning experiences in research data management (Gargiulo et al., 2021; Majid et al., 2018; Palsdottir, 2021).

Identified Instructional Need

While the need for instruction on data management planning has been identified more broadly, a performance gap observed in a research training and mentorship program in the south-central US fueled the development of the data management course described in this article. The program, called the Obesity Health Disparities Program to Increase Diversity among Individuals Engaged in Health Related Research (OHD-PRIDE), is funded through the National Heart, Lung and Blood Institute (NHLBI) and provides research training, mentoring, and professional development for early-career faculty members and transitioning postdoctoral fellows from underrepresented groups in biomedical, behavioral, and social sciences fields (Beech et al., 2020; Thorpe et al., 2020). The program’s conceptual framework addresses challenges for early career faculty, many of whom have large teaching loads and limited research infrastructure in their institutions (Beech et al., 2018). OHD-PRIDE

aims to equip individuals with essential scientific writing and grantsmanship skills to obtain extramural funding successfully. Fueled in part by recent requirements from funding agencies to include data management plans in grant proposals, program organizers recognized the need for the OHD-PRIDE fellows to be equipped in this area.

Designing the course brought together individuals with expertise in instructional design, computer science, population health, research training, sociology, ethics, and data science. To understand the scope of content for the course, the team explored current literature describing data collection, storage, preservation, and availability; data management planning; defining metadata using common standards and practices; and maintaining data *provenance* [the historical information related to the handling of the data] (Demchenko et al., 2017; Wolff et al., 2021). The FAIR principles are the preferred guidance for deciding where and how data can be placed to support access, reuse, and analysis (Bezuidenhout, 2021; Higman et al., 2019; Wilkinson et al., 2016). The acronym FAIR stands for findability, accessibility, interoperability, and reusability. Infusing FAIR into data management practices enables “the machine knows what we mean,” referring to machine *actionability*; this makes it possible for machines to read and process the data successfully (Jacobsen et al., 2020, p. 13). The four FAIR foundational principles contain 15 FAIR guiding principles that provide greater specificity to support interpretation and implementation by individual stakeholder communities. The FAIR guidance encompasses metadata schemes, structured vocabularies for annotations, and data repository storage and access that support greater efficiency and effectiveness of data use through computer-executed automation of data indexing, retrieval, and analysis (Boeckhout et al., 2018; Persaud et al., 2021; Wise et al., 2019). The framework provides direction for communities (such as those from particular specialty areas) to evaluate their metadata practices, determine needed improvements, and adapt the implementation of the FAIR principles to their needs.

Synthesizing the performance needs and content demands, the team articulated the following four learning goals for the course:

1. Learners will recognize the importance of data reproducibility in scientific research.
2. Learners will gain a deeper understanding of data management and its relevance to research.
3. Learners will gain a deeper understanding of the FAIR principles and their relevance to research.
4. Learners will become aware of the practical skills in data management and FAIR implementation.

The team analyzed the goal statements further and gathered information resources about the associated concepts and skills for each goal statement. More specific learning objective statements were then drafted and refined as the curriculum developed. For example, the learning objectives connected to goal 4 are as follows:

- 4.1.1 Learners will identify the components of rich metadata.
- 4.1.2 Learners will explain the purpose of a persistent identifier.
- 4.1.3 Learners will register meta(data) in a searchable source.
- 4.2.1 Learners will explain the importance of open, free, and universally implementable protocols.

- 4.2.2 Learners will identify the types of data access levels.
- 4.2.3 Learners will apply the principles of accessibility to a dataset.
- 4.3.1 Learners will explain the term “qualified references.”
- 4.3.2 Learners will create metadata that includes FAIR vocabularies.
- 4.3.3 Learners will find controlled vocabularies associated with terms in their fields of study.
- 4.4.1 Learners will describe good data reusability practices.
- 4.4.2 Learners will create detailed provenance for a given dataset.
- 4.4.3 Learners will list the minimum metadata requirements.

Design Precedents for FAIR Principles Training

The team gathered design precedents that could provide insights into potential course formats and instructional approaches to teach FAIR principles in data management planning (Boling, 2021). As a relatively new training area, we found varied training approaches ranging from lecture-based sessions to hands-on workshops involving practice with data set labeling and storage. The utility of the design precedents for our instructional design was that they provided the team with ideas for content breadth, sequencing, activities, and organizational structure that we would later use in developing our course. We noted how prior courses have prioritized instructional design strategies that build upon theoretical discussions through activity opportunities for implementing FAIR principles in real-world research contexts.

One design precedent came from a collaborative group of about 40 individuals from the FAIRsFAIR - Fostering Fair Data Practices in Europe project, in which they produced a handbook for college educators (Engelhardt et al., 2022). In a section on “Teaching and Training Designs for FAIR,” the authors acknowledge the diverse learning needs of professionals in this space and suggest considering a variety of training format possibilities

FAIR training could be delivered as part of a formal course, part of a training or promotional event, or it can be embedded in managerial processes (e.g., grant application support, ethical review process, basic training for new affiliate researchers, etc.). It could also be a lecture, a workshop, a series of events, an online course, self-learning materials, or training interventions. (p. 37)

The handbook writers specifically articulated each training format type’s pros and cons regarding the FAIR content and contextualized the format decision considering the training goals, available resources, and anticipated constraints. For example, FAIR training “events” provide advantages of networking opportunities and promotion of open data science. However, they tend not to provide enough time to delve deeply into this content.

FAIR (2022) also offers an example of a self-learning material – a question-based online tool called “FAIR-Aware.” The tool is designed for users who have a particular dataset and are working to prepare it for uploading to a data repository. Available in English and French, the tool presents users with ten self-assessment questions that generate targeted guidance texts in the FAIR areas they most need. It also provides a glossary of commonly used FAIR-

related terms. For trainers, it offers capabilities for downloading learner self-assessment results and utilizing analysis templates to provide follow-up feedback and support to learners.

We found workshops and courses to be another recognized approach for the FAIR content. For example, the German National Library of Science and Technology (Technische Informationsbibliothek [TIB]) conducted a week-long in-person workshop that included didactic presentations of the content via lectures ranging from 17 minutes to just under two hours and hands-on coding sessions using practice materials in various STEM disciplines (Leinweber et al., 2018). The scope of content was organized across the five days as (a) FAIR introduction and findability, (b) accessibility, (c) interoperability, (d) reusability, and (e) wrap-up. The lectures were recorded and released as a nine-part series available under Creative Commons licensing for free access. The hands-on coding sessions align with case-based learning (CBL) techniques, which have been used in biomedical education to enable professional learners to consider target objectives within real-world scenarios and resources (Greenberg-Worisek et al., 2019). The training immerses learners in the content and facilitates knowledge development through immersive and meaningful learning experiences.

The team also considered design precedents from their own experiences. One author with previous experience teaching an undergraduate course for pre-service teachers described some of the hands-on activity strategies she has used for engaging learners with new vocabulary. Card sort is one such activity in which students are provided with a stack of cards or post-it sticky notes and a list of terms. Working in small groups of 2-3, students write the terms on the cards and then ideate categories and ways to create groups of terms. Following the sorting, groups share their category headings, organizational approaches, and relationships that they view among the terms. Think-Pair-Share is another activity strategy this author has used to prompt learners to consider their prior knowledge and experiences with target concepts and socially engage in conversation with a classmate about their thoughts.

Two other authors have built prior online, on-demand short courses on the Coursera and Blackboard platforms, including courses about instructional technologies for teachers, COVID-19 contact tracing, and value-based care. They reflected on structural aspects of the materials, including format types, lengths, graphical elements, interactive components, assessment approaches, and connections to real-world applications. For example, the courses largely present the content through short (about 3-7 minutes) videos that were edited to appear professional and unified across each course. The videos incorporate intro/outro audio/video sequences, on-screen text to reinforce key terms, diagrams that illustrate relationships between concepts, and high-quality stock and original photographs. Videos and activities in these courses were often grouped into modules that could be completed in about 1-2 hours. Embedded video questions and closed-response quiz items were used to assess learner recall of concepts. Several of the courses culminated in peer-graded assignments that prompted learners to apply the target skills in the development of a written submission that would then be reviewed by several of their course mates. The peer-graded assignment instructions were highly detailed, with specific instructions on what the learners were to produce and how peer graders should score and provide feedback on their classmates' work.

Learner Analysis Study

Study Context

We then conducted a learner analysis study with a cohort of target learners from OHD-PRIDE. Learner analysis can inform the appropriateness and feasibility of the intended training for the target audience (Agic et al., 2023; Dick et al., 2015). We constructed a learner analysis survey instrument to gather data on learners' demographics, prior familiarity with FAIR, and learning preferences that could be used to inform the instructional design project plan. The intention of gathering information about learning preferences was to understand how learners might desire content representations, learning activity options, and assessment formats, as was found useful by Rohani et al. (2024) in the development of health data science education programs. The instrument underwent expert review for content validity by two mentors from the program, and review feedback was used to simplify and clarify the items. The resulting instrument consists of demographic, FAIR principles knowledge, and learning preferences items (see Appendix A). We attended a January 2022 cohort meeting and discussed the educational need in light of upcoming federal grant proposal expectations for implementing FAIR principles into data management planning. Following Institutional Review Board approval, an invitation to participate in the study was distributed to members of the program cohort.

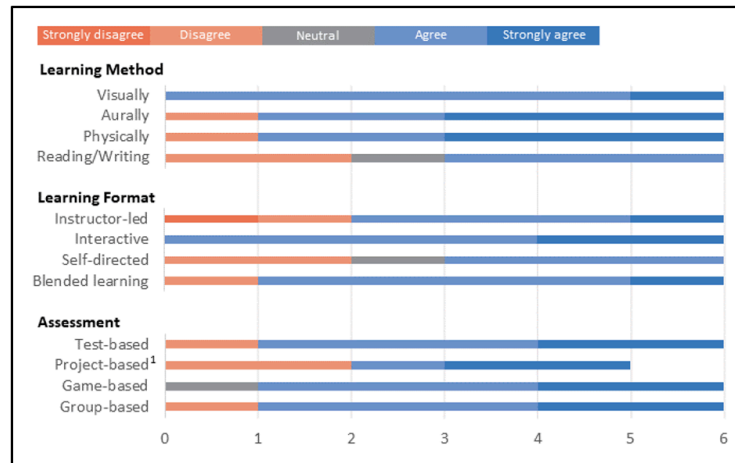
All six cohort members consented in writing to participate and completed the survey via Qualtrics XM™ online survey software (100% response rate). Respondents included three males and three females, primarily in the age range of 35-44 years, with one individual younger and one older than that range. Five respondents described themselves as Black or African American, and one as Hispanic/Latino. All respondents indicated that their primary research area was obesity health disparities. The collected survey data were analyzed using descriptive statistics for the items in the FAIR principles knowledge and learning preferences sections, and summary themes were identified and discussed within the research team to identify connections to considerations for the course design.

Findings

Respondents indicated minimal prior awareness of the FAIR principles overall. Figure 1 illustrates the target learners' preferences for learning about using FAIR. The average ratings were similar (Agree) for three of the four preferred learning methods: visually, such as viewing diagrams, videos, or presentations; aurally, such as listening to audio recordings or participating in group discussions; and physically, such as participating in hands-on activities. The fourth option, reading and writing, was indicated as their least preferable.

Figure 1

Target Learner Preferences for Learning about FAIR



¹Note: There was one missing response in the project-based assessment item.

The target learners were divided in their views about the preferred learning format for the FAIR content. For instance, four respondents agreed and strongly agreed that instructor-led/traditional training would be preferable, while the other two disagreed and strongly disagreed. Responses regarding self-directed learning (e.g., critical reflection, mind mapping, independent learning activities) were mixed. This category yielded slightly lower ratings, with three agreeing, two disagreeing, and one indicating neutral. The one format that respondents unanimously preferred was interactive training activities, such as group activities, games, and simulation. Blended learning was the second most preferred format.

Regarding preferred assessment formats, interactive game-based assessments were preferred mainly, followed by group-based assessments. Respondents had mixed views on the project-based format option. The test-based assessments category was the least preferred assessment format. An additional assessment format of the "adaptive learning method" was suggested in the "Other" option.

Implications for the Data Management Course Design

Informed by the identified training need, design precedents, and learner analysis findings, the course development team applied these insights to design the course content scope and instructional strategy. The team first developed and piloted an in-person version of the course in June 2022, and an online, on-demand version was developed and offered through Coursera the following year. Considering the contextual need to connect FAIR-based applications to the appropriate communities of practice, the course provided obesity health disparities research implications, such as using an example dataset from a study on the impacts of obesity and weight loss for hands-on practice activities. Figure 2 provides an example learning activity that used the obesity and weight loss dataset. This activity also incorporated technical skills relating to searching for controlled vocabularies with keywords in the National Library of Medicine MeSH (Medical Subject Headings). Given the target learners' preference for visual learning methods, video vignettes were developed to correspond with the activity. The videos discuss theoretical concepts and provide examples

from practice. Videos corresponding to the activity featured in Figure 2 introduce controlled vocabularies and explain how to find and link to controlled vocabularies (see Figure 3).

Figure 2

Example Learning Activity Using Obesity and Weight Loss Dataset

Learning Activity: Using Controlled Vocabularies

Objectives:

- Learners will create metadata that includes FAIR vocabularies.
- Learners will find controlled vocabularies associated with terms in their fields of study.

Instructions:

Using the [example dataset from a study on the impacts of obesity and weight loss](#) or your dataset, write four or more keywords to describe it. Search for controlled vocabularies related to your keywords in the [National Library of Medicine MeSH \(Medical Subject Headings\)](#).

Note: If you are using the example dataset, you can explore the keywords and MeSH terms in these articles:

- ["Substitution of sugar-sweetened beverage with other beverage alternatives..."](#)
- ["Substituting sugar-sweetened beverage with water or milk..."](#)
- ["Associations between sugar intake from different food sources..."](#)

Optional:

Practice working with a data repository (such as the [Texas Data Repository](#)) and create an entry for your dataset. Edit the metadata fields of key terms and top classification to include controlled vocabularies.

Figure 3

Video Vignette of Data Management Expert

The screenshot shows a video player interface. On the left is a sidebar with a list of videos under the heading 'Overview: Making your Data Interoperable'. The selected video is 'Video: Finding Controlled Vocabularies' (2 min). The main video area displays a screen recording of the National Library of Medicine MeSH website. The video title 'Finding Controlled Vocabularies' is shown below the video player. Below the video player is a transcript area with the text: 'Mesh is a big choice. Mesh is related to medicine and it's broadly applicable for medicine as well as to biomedical sciences but it's really broad. It's not as broad as Library of Congress subject headings which are probably too broad. When I used to work at Johns Hopkins Institute of the...'. A 'Save note' button is visible in the top right corner of the video player area.

With the target learner base having minimal prior familiarity with the principles of FAIR, contextualizing concepts were incorporated into the introductory module. These concepts included explanations about the circumstances that gave rise to the development of the FAIR principles, the link between data reproducibility and scientific discovery, and the role of

researchers in the data ecosystem. To support the building of the necessary background knowledge for learners to understand the rationale for implementing FAIR into their practices, foundational concepts of data reproducibility, open science, and data management were explained. Essential vocabulary terms associated with the content, such as *metadata*, *persistent identifiers*, and *controlled vocabularies*, were defined when first used in the modules and then curated into a complete glossary reference.

To address the target learners' preferences for interactive activities and group-based assessments, hands-on and collaborative activities that situated concepts within real-world scenarios were designed for the course. Activities included group discussions, Think-Pair-Share, sorting, concept mapping, and exploratory checklists (see Figure 4). The activities were designed at varying levels of thinking (e.g., Bloom's taxonomy), from information recall and comprehension to implementation and evaluation (Adams, 2015). Learner-to-learner collaboration was supported through instructional technologies such as cloud-based editable document activity scaffolds.

Figure 4

Example Exploratory Checklist Learning Activity

Learning Activity: Reusability Checklist	
<p><i>Objective:</i></p> <ul style="list-style-type: none"> • Learners will create detailed provenance for a given dataset. <p><i>Instructions:</i></p> <ol style="list-style-type: none"> 1. Review the data info page from a study on the impacts of obesity and weight loss. 2. Complete the checklist below by seeing if you can find each item listed in the left column in the provided data. You can mark an "X" in the right column for the items you find. 	
Checklist items	Mark "X" if found in the data info page
Metadata includes general information such as data title, PI contact and affiliation, collaborators, date of collection, and geographic location.	
Metadata provides methodological information such as data collection and processing methods.	
Metadata provides an overview of data and files, such as file names, the date of file creation, and whether data is used in raw or transformed form.	

Files are saved in standard formats, e.g., CSV for tabular data.	
Metadata provides data-specific information such as variables, units of measurement, and code definitions.	
The repository has assigned a license to the dataset.	
The repository has allowed the assigning of access levels to the dataset, e.g., open, restricted, or embargoed.	
Metadata includes suggestions for citation.	
Metadata includes publication details.	
Metadata conforms to community standards.	

Suggested starter texts for articulating applications of FAIR in different portions of a data management plan were embedded into the course to enable learners to more readily implement the concepts learned into their scholarly projects (Figure 5).

Figure 5

Suggested Data Availability Starter Text for Learners

Examples of Data Availability Statements	Statement type	Example statement
	Openly available data	"All data underpinning this publication are openly available from the University of FAIR-Data Repository at http://doi.org/10.15000/a789457 "
	Embargoed data	"All data underpinning this publication will be available from the University of FAIR-Data Repository at http://doi.org/10.15000/a123456 from 01/02/2019 onwards, following the cessation of an embargo period."
	Restricted data	"Due to ethical/commercial issues, data underpinning this publication cannot be made openly available. Further information about the data and conditions for access are available from the University of FAIR-Data Repository at http://doi.org/10.15000/a123456 "
	Partially restricted data	"Due to the sensitive nature of this research, only a subset of the participants consented to their anonymised data being retained and shared. Anonymised interview transcripts and survey results from participants who provided consent, other supporting data, and further details relating to the restricted data, are available from the University of FAIR-Data Repository at http://doi.org/10.15129/a123456 "
	Physical data	"Physical data supporting this publication are stored by the University of FAIR-Data. Details of the data and how it can be accessed are available from the University of FAIR-Data Repository at http://doi.org/10.15129/a123456 "
	Secondary data	"Pre-existing data underpinning this publication are openly available from UKDS at http://doi.org/10.12345/54321 . Further information about data processing, and additional new supporting data are available from the University of FAIR-Data Repository at http://doi.org/10.15129/a123456 "
	No new data created	"No new data were created during this study. Pre-existing data underpinning this publication were obtained from NPL and are subject to licence restrictions. Full details on how these data were obtained are available in the documentation available from the University of FAIR-Data Repository at http://doi.org/10.15129/a123456 "
	No data	"This work is entirely theoretical, there is no data underpinning this publication."

UNIVERSITY of HOUSTON

FAIRsFAIR

23

The team recognized that a range of assessment items would be needed, from recall/recognition multiple-choice items to evaluative/demonstrative authentic tasks. The activities served as formative assessments, providing opportunities for feedback and reflection on learning progress within each module. In the online version of the course, closed-response, auto-graded quizzes were created to serve as knowledge checks at the end of each module. A culminating activity of developing a data management plan incorporating FAIR principles addressed higher levels of thinking and application. The assessment was

trialed as an instructor-reviewed assignment initially in-person, and then a peer-graded version was developed for online, on-demand with the aim that it could guide learners to support their classmates through review and feedback (see Figure 6). In both versions, learners needed additional time to complete the assignment than what was planned in the course, and the timeliness of feedback on their work was not optimal. Future iterations could explore its use as a scaffold for discussion of the learner's data management plan within the broader research training and mentorship program.

Figure 6

Culminating Peer-review Assessment of Data Management Plan

The screenshot shows a Coursera interface for a peer-graded assignment. The title is 'Peer-graded Assignment: Data Management Plan'. Below the title are tabs for 'Instructions', 'My submission', and 'Discussions'. There are links for '(Admin-Only Link) Preview Rubric' and 'Back To Assignment'. The main content is a 'Rubric Preview' table with two columns: 'PROMPT' and 'RUBRIC'.

PROMPT	RUBRIC
<ol style="list-style-type: none"> 1. Open the draft of your data management plan that you created using the DMPTool in Module 3. Finish completing it. 2. Check to make sure that your data management plan addresses the following 10 elements: <ol style="list-style-type: none"> a. a brief description of the types and amount of scientific data that will be generated in the project b. which scientific data will be preserved and shared, c. the types of documentation and metadata that will accompany the data, d. what format(s) the shared scientific data will be made available and any specialized tools, software, and/or code that will be needed for access, e. when the data will be shared and for how long, f. which repository or other mechanisms will be used to store the data, g. who will have access to the data and under what conditions, h. justifications for any data access and re-use limitations, such as those that might be related to issues of informed consent, privacy, and confidentiality protections, i. how the data will be findable by other researchers, and j. how and by whom the oversight of data management activities will be handled. 3. Download your plan in PDF format and submit it for peer review. 	<p>Does the data management plan include a brief description of the types and amount of scientific data that will be generated in the project?</p> <p><input type="radio"/> 1 pt Yes <input type="radio"/> 0 pts No</p> <p>Does the data management plan mention which scientific data will be preserved and shared?</p> <p><input type="radio"/> 1 pt Yes <input type="radio"/> 0 pts No</p> <p>Does the data management plan describe the types of documentation and metadata that will accompany the data?</p> <p><input type="radio"/> 1 pt Yes <input type="radio"/> 0 pts No</p> <p>Does the data management plan identify the format(s) the shared scientific data will be made available and any specialized tools, software, and/or code that will be needed for access?</p> <p><input type="radio"/> 1 pt Yes <input type="radio"/> 0 pts No</p>

Conclusions and Future Work

The data management course described in this article focuses on applications of FAIR principles in obesity health disparities research. It aims to expand participant awareness of open science, close skill gaps in FAIR-grounded research data management practices, situate participant implementation of FAIR within their associated obesity health disparities community of practice, and improve participant skill in writing FAIR-informed data

management plans as part of external funding applications. Participant demographics in the learner analysis study reflect the context of this project being situated within the OHD-PRIDE research training and mentorship program for early-career faculty and post-doctoral fellows from underrepresented groups in biomedical, behavioral, and social sciences fields, and these demographic variables minimally impacted design decisions for the curricular planning. Additionally, though learners were asked about their learning preferences prior to the development of the course, the team found in the piloting of the program that presenting course concepts through multiple modalities, including visual presentations, videos, text-based guides and templates, and exploratory learning experiences provided representations of the concepts and skills in supportive ways. The training design incorporates a single example dataset for hands-on practice activities to contextualize learner applications of the course content. The course's in-person and online, on-demand versions offer instructional format flexibility, and developing both versions met the expected deliverables of our design project.

The continued expansion of open science efforts and external funder expectations for incorporating data management plans in grant proposals will likely further fuel the need for educational programming in this area. A future iteration could merge the in-person and online formats to create a blended version, which may utilize the online content delivery as pre-class work and maximize the in-person training time for interactive learning activities. The incorporation of game-based assessments could also be explored further. Applying gamification elements would involve the need for incremental assessment points. For example, adding embedded questions in the videos could provide game-based point opportunities, foster retrieval practice, and support learner motivation and engagement (Hopkins et al., 2016).

In sum, the design team found the front-end analysis insights helpful in guiding design decisions about delivery format, scope of content, and activity structure for a data management course. Exploring design precedents spurred ideas about how the technical FAIR content could be organized into a formal instructional experience. Gathering input from target learners in the design's initial phase confirmed the course's instructional need. It also enabled the design team to understand target learner views about the proposed course aspects. The resulting course reflected proactive responses to the front-end analysis insights as the learning needs of the early-career biomedical researchers were acknowledged and addressed in teaching conceptual knowledge and relevant applied skills in FAIR principles for data management. While the course was piloted as a standalone professional development that introduces FAIR-grounded concepts and skills, future iterations could utilize the mentorship affordances of a broader research training and mentorship program, such as OHD-PRIDE, to provide extended support of skills application into the development of grant proposal data management plans.

References

- Adams, N. E. (2015). Bloom's taxonomy of cognitive learning objectives. *Journal of the Medical Library Association*, 103(3), 152–154. <https://doi.org/10.3163/1536-5050.103.3.010>

- Agic, B., Fruitman, H., Maharaj, A., Taylor, J., Ashraf, A., Henderson, J., Ronda, N., McKenzie, K., Soklaridis, S., & Sockalingam, S. (2023). Advancing curriculum development and design in health professions education: A health equity and inclusion framework for education programs. *Journal of Continuing Education in the Health Professions*, 43(4S), S4-S8. <https://doi.org/10.1097/ceh.0000000000000453>
- Bezuidenhout, L. (2021, April 6). *Open and responsible research: Roles and responsibilities for data stewards*. <https://zenodo.org/record/4665390>
- Beech, B. M., Bruce, M. A., Thorpe, Jr., R. J., Heitman, E., Griffith, D. M., & Norris, K. C. (2018). Theory-informed research training and mentoring of underrepresented early-career faculty at teaching-intensive institutions: The obesity health disparities PRIDE program. *Ethnicity & Disease*, 28(2), 115-122. <https://doi.org/10.18865%2Fed.28.2.115>
- Beech, B. M., Norris, K. C., Thorpe, R. J., Heitman, E., & Bruce, M. A. (2020). Conversation Cafés and conceptual framework formation for research training and mentoring of underrepresented faculty at historically black colleges and universities: Obesity health disparities (OHD) PRIDE program. *Ethnicity & Disease*, 30(1), 83-90. <https://doi.org/10.18865%2Fed.30.1.83>
- Boeckhout, M., Zielhuis, G. A., & Bredenoord, A. L. (2018). The FAIR guiding principles for data stewardship: Fair enough? *European Journal of Human Genetics*, 26(7), 931–936. <https://doi.org/10.1038/s41431-018-0160-0>
- Boling, E. (2021). The nature and use of precedent in designing. In J. K. McDonald & R. E. West (Eds.), *Design for Learning*. EdTech Books. <https://edtechbooks.org/id/precedent>
- Demchenko, Y., Belloum, A., & Wiktorski, T. (2017). *EDISON data science framework: Part 1. Data Science Competence Framework (CF-DS)*. <https://edison-project.eu/data-science-competence-framework-cf-ds/>
- Dick, W., Carey, L., & Carey, J. O. (2015). *The systematic design of instruction* (8th ed.). Pearson.
- Dijkers, M. P. (2019). A beginner's guide to data stewardship and data sharing. *Spinal Cord*, 57(3), 169-182. <https://doi.org/10.1038/s41393-018-0232-6>
- Eaker, C. (2020). *Open research toolkit*. <https://osf.io/a4ftw/>
- Engelhardt, C., Biernacka, K., Coffey, A., Cornet, R., Danciu, A., Demchenko, Y., Downes, S., Erdmann, C., Garbuglia, F., Germer, K., Helbig, K., Hellström, M., Hettne, K., Hibbert, D., Jetten, M., Karimova, Y., Kryger Hansen, K., Kuusniemi, M. E., Letizia, V., ... Zhou, B. (2022). *D7.4 How to be FAIR with your data. A teaching and training handbook for higher education institutions* (V1.2.1 DRAFT) [Computer software]. Zenodo. <https://doi.org/10.5281/zenodo.6425568>

- FAIRsFAIR. (2022). *FAIR-Aware: Assess your knowledge of FAIR*. FAIRsFAIR.
<https://www.fairsfair.eu/fair-aware>
- Gargiulo, P., Galimberti, P., Tammara, A. M., & Zane, A. (2021). *FAIR RDM (Research Data Management)*: 11.
- Greenberg-Worisek, A. J., Campbell, K. A., Klee, E. W., Staff, N. P., Schimmenti, L. A., Weavers, K. M., Ekker, S. C., & Windebank, A. J. (2019). Case-based learning in translational biomedical research education: Providing realistic and adaptive skills for early-career scientists. *Academic Medicine*, 94(2), 213–216.
<https://doi.org/10.1097/ACM.0000000000002470>
- Higman, R., Bangert, D., & Jones, S. (2019). Three camps, one destination: The intersections of research data management, FAIR and Open. *Insights*, 32(1), 18.
<https://doi.org/10.1629/uksg.468>
- Hopkins, R., Lyle, K., Hieb, J., & Ralston, P. (2016). Spaced retrieval practice increases college students' short- and long-term retention of mathematics knowledge. *Educational Psychology Review*, 28(4), 853–873. <https://doi.org/10.1007/s10648-015-9349-8>
- 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
- Leinweber, K., Kraft, A., Kuzak, M., Luke Johnston, Martin Hammitzsch, & Förstner, K. (2018, July 16). *FAIR data and software: A carpentries-based workshop at TIB, Hannover*.
<https://zenodo.org/record/3707745#.YmmGaOfMJEY>
- Majid, S., Foo, S., & Zhang, X. (2018). Research data management by academics and researchers: Perceptions, knowledge and practices. In M. Dobrev, A. Hinze, & M. Žumer (Eds.), *Maturity and Innovation in Digital Libraries* (pp. 166–178). Springer International Publishing. https://doi.org/10.1007/978-3-030-04257-8_16
- Miksa, T., Simms, S., Mietchen, D., & Jones, S. (2019). Ten principles for machine-actionable data management plans. *PLOS Computational Biology*, 15(3). 2019;15(3):e1006750.
<https://doi.org/10.1371/journal.pcbi.1006750>
- National Institutes of Health. (2020). *NOT-OD-21-013: Final NIH policy for data management and sharing*. <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html>
- Open Science Skills Working Group of the European Commission. (2017). *Providing researchers with the skills and competencies they need to practise Open Science*. Publications Office. <https://data.europa.eu/doi/10.2777/121253>
- Palsdottir, A. (2021). Data literacy and management of research data – A prerequisite for the sharing of research data. *Aslib Journal of Information Management*, 73(2), 322–341.

- Persaud, B. D., Dukacz, K. A., Saha, G. C., Peterson, A., Moradi, L., O'Hearn, S., Clary, E., Mai, J., Steeleworthy, M., Venkiteswaran, J. J., Kheyrollah Pour, H., Wolfe, B. B., Carey, S. K., Pomeroy, J. W., DeBeer, C. M., Waddington, J. M., Van Cappellen, P., & Lin, J. (2021). Ten best practices to strengthen stewardship and sharing of water science data in Canada. *Hydrological Processes*, 35(11), e14385. <https://doi.org/10.1002/hyp.14385>
- Rohani, N., Sowa, S., & Manataki, M. (2024). Identifying learning preferences and strategies in health data science courses: Systematic review. *JMIR Medical Education*, 10, e50667, 1-16. <http://dx.doi.org/10.2196/50667>
- Thorpe, Jr., R. J., Beech, B. M., Norris, K. C., Heitman, E., & Bruce, M. (2020). Writing accountability groups are a tool for academic success: The Obesity Health Disparities PRIDE program. *Ethnicity & Disease*, 30(2), 295-304. <https://doi.org/10.18865%2Fed.30.2.295>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. 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). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, 3(1), 160018. <https://doi.org/10.1038/sdata.2016.18>
- Williams, M., Bagwell, J., & Zozus, M. N. (2017). Data management plans, the missing perspective. *Journal of Biomedical Informatics*, 71, 130-142. <https://doi.org/10.1016/j.jbi.2017.05.004>
- Wise, J., de Barron, A. G., Splendiani, A., Balali-Mood, B., Vasant, D., Little, E., Mellino, G., Harrow, I., Smith, I., Taubert, J., van Bochove, K., Romacker, M., Walgemoed, P., Jimenez, R. C., Winnenburg, R., Plasterer, T., Gupta, V., & Hedley, V. (2019). Implementation and relevance of FAIR data principles in biopharmaceutical R&D. *Drug Discovery Today*, 24(4), 933–938. <https://doi.org/10.1016/j.drudis.2019.01.008>
- Wolff, I., Broneske, D., & Köppen, V. (2021). FAIR research data management for learning analytics. *Proceedings of the DELFI Workshops 2021*. DELFI Workshops 2021.

Appendix A

FAIR Training Learner Analysis Survey

Section 1: Demographics

1. Name
2. Email
3. Gender
4. Age
5. Race
6. Primary research area

Section 2: Awareness of FAIR Principles

7. Please choose the option that best represents your familiarity with the FAIR principles.
(Not familiar at all–Extremely familiar)

Section 3: Learning Preferences

8. My preferred method of learning about the FAIR principles would be: (Strongly disagree–Strongly agree)
- a. Visually, such as viewing diagrams, videos, or presentations
 - b. Aurally, such as listening to audio recordings or participating in group discussions
 - c. Physically, such as participating in hands-on activities
 - d. Through reading and writing
 - e. Other _____
9. In which format would you prefer to learn about the FAIR principles? (Strongly disagree –Strongly agree)
- a. Instructor-led/traditional training
 - b. Interactive training (e.g., group activities, games, simulations)
 - c. Self-directed learning (e.g., critical reflection, mind mapping, independent learning activities)
 - d. Blended learning
 - e. Other _____
10. If you were to participate in training on the FAIR principles, what assessment format would you prefer to demonstrate your understanding of the FAIR principles? (Strongly disagree –Strongly agree)
- a. Test-based
 - b. Project-based
 - c. Interactive (e.g., game-based)
 - d. Group-based
 - e. Other _____

Conflicts of Interest and Source of Funding

Research reported in this publication was funded, in part, by the National Institutes of Health (NIH) Agreement No. 1OT2OD032581-01 and the National Heart, Lung, and Blood Institute (NHLBI) of the NIH under Award Numbers R25HL126145 and R25HL126145-08S1. The content is solely the authors' responsibility and should not be interpreted as representing the official policies of the NIH or NHLBI, expressed or implied.

Ethical approval

The methodology and research instruments for this study were approved on February 17, 2022, by the IRB at the University of Houston (STUDY00003499). Informed consent was obtained from all participants in this study.

Acknowledgments

The authors would like to acknowledge the Obesity Health Disparities PRIDE research training and mentoring program partnership among the University of Houston, the University of California, Los Angeles (UCLA), Johns Hopkins University, and the University of Texas Southwestern for their contributions to the curriculum. The authors also thank Elizabeth Heitman, Ph.D., Haleh Aghajani, Ph.D., MSc; Anthony Dellureficio, MLS, MSc; Ramkiran Gouripeddi, MBBS, MS; and Katherine Sward, Ph.D., RN, FAAN for their review and guidance during the curriculum conceptualization and Daniel Burgos, Gione-Paul Comeaux, and Khy-Lan Patton for their contributions to the online instructional materials.



Susie L. Gronseth

University of Houston

Susie L. Gronseth, Ph.D., is Clinical Professor in the Learning, Design, and Technology program area in the College of Education at the University of Houston (Houston, Texas, USA). Through interdisciplinary research and community engagement initiatives, she provides expertise in applications of learning technologies, instructional design and evaluation, health sciences education, and inclusive educational frameworks that strategically address diverse learner needs in online, face-to-face, and blended contexts. She has received recognition for her course design, teaching, and university leadership, including the International Society for Technology in Education (ISTE) Online Learning Network Award, the University of Houston Teaching Excellence Award, the Indiana

University School of Education Distinguished Alumni Award,
and University of Houston Faculty Senate President.



Lydia Oluchi Ugwu

University of Illinois College of Medicine Peoria Houston

Lydia Oluchi Ugwu holds a PhD in Curriculum and Instruction with a focus on Learning, Design, and Technology, from the University of Houston. She serves as the Assistant Dean for Faculty Development at the University of Illinois College of Medicine Peoria where she plays a pivotal role in advancing the college's strategic plan to enhance educational quality and teaching excellence. She has presented at various national and international conferences and has published peer-

reviewed articles and book chapters on instructional design and curriculum development.



Bettina M. Beech

Dr. Bettina M. Beech is the Chief Population Health Officer and Clinical Professor at the University of Houston. She has a unique interdisciplinary background that integrates public health, population health, minority health, and health disparities research with an emphasis on chronic disease self-management across the life course. Her secondary line of

research focuses on the science of research training and mentoring, particularly among early career faculty.



Marino A. Bruce

Marino A. Bruce, PhD, MSRC, MDiv is a Clinical Professor of Behavioral and Social Sciences and the Associate Dean for Research and in the University of Houston Tilman J. Fertitta Family College of Medicine. Dr. Bruce is also the Founding Director of UHPH Collaboratories – synergistic interdisciplinary research and training units within UH Population Health. Dr. Bruce is a sociologist and population

health scientist who examines the full range of factors influencing cognitive and physical functioning among Black males over the life course and across generations. He is editor of Research on Race and Ethnic Relations and current associate editor of Ethnicity and Disease. Dr. Bruce earned a PhD in Sociology from North Carolina State University and received postdoctoral training in Family Medicine from the University of Wisconsin-Madison and in Biobehavioral Health from Duke University.



Ayesha Siddiqui

University of Houston

Ayesha Siddiqui is an instructional designer at the C. T. Bauer College of Business, University of Houston, where she assists more than 200 faculty members and 150 teacher assistants. She has a master's degree in Curriculum and Instruction with a focus on Learning, Design, and Technology from the University of Houston and a bachelor's degree in interdisciplinary studies from the University of Houston Downtown.



Amani Itani

Amani is a PhD student at the University of Houston, studying Curriculum and Instruction with a focus on Learning Design and Technology. Amani's research journey is centered around the intersection of education and technology, with a particular focus on integrating Artificial Intelligence into educational contexts (AIEd). Her explorations also encompass curriculum planning methodologies and the cultivation of 21st-century skills within modern educational landscapes.



Ioannis A. Kakadiaris

Ioannis A. Kakadiaris has a long and distinguished track record in biometrics, data/video analytics, and artificial intelligence. His team has made contributions in the areas of face and ear recognition and profile-based face recognition. CBL's 3D-3D face recognition software ranked first in the 3-D-shape section of the 2007 Face Recognition Vendor Test organized by NIST. He proposed 3-D aided 2-D Face Recognition, which created a new research field later called heterogeneous or asymmetric face recognition that, for the first time, enabled matching 3-D data with legacy 2-D data. CBL's research has been supported by federal (NIH, NSF, Army Research Labs, DHS, and National Institute of Justice), state (Texas Higher Education Coordinating Board), industry (SGI, American Honda, Microsoft Research, Unisys, Siemens Medical Solutions, and BP America), foundations (Juvenile Diabetes Research Foundation and Schlumberger Technical Foundation), and international funding organizations (French Partner University Fund). Ioannis' research has been featured on Discovery Channel, National Public Radio, KPRC NBC News, KTRH ABC News, and KHOU CBS News. He was a recipient of the UH Computer Science Research Excellence Award two times. He has been recognized for his research with several distinguished honors, including the NSF Early Career Development Award, the Schlumberger Technical Foundation Award, the UH Teaching Excellence Award, and the James Muller Vulnerable Plaque Young Investigator Prize. He has served as the VP of Technical Activities at the IEEE Biometrics Council, Program Co-Chair for BTAS 2019, General Co-Chair for IJCB 2020, and he is a graduate of the NSF iCORPs program.



This work is released under a CC BY license, which means that you are free to do with it as you please as long as you properly attribute it.