Reflection and Analysis of Implementing a Free Asynchronous MOOC to Build Competence in Biomedical Research Data Management

Julie Goldman and Nevada F. Trepanowski

This article reports on the development and evaluation of a massive open online course (MOOC) that provides instruction on best practices in research data management (RDM). The course was developed in response to the growing need for data management professional development for LIS professionals and to promote data management to researchers. In just 18 months of the course launch, the course reached more than 1,000 people from across the world and was effective in building student competency in RDM. The success of this course illustrates the value and utility of free online professional development as a tool for both library and research staff.

Introduction

Traditional in-person continuing education is a great resource for professional development. However, the time and expense associated with in-person education can pose barriers to many Library and Information Science (LIS) professionals looking to increase their knowledge of data services. Massive Open Online Courses (MOOCs) offer flexibility and affordability via asynchronous instruction to ensure the LIS professionals can build the skills required to become effective research data management (RDM) partners.

In 2015, the National Institutes of Health (NIH) launched the Big Data to Knowledge (BD2K) Initiative to address data science challenges, including lack of appropriate tools, poor data accessibility, and insufficient training. As a result, multiple groups received grant funding[1] to expand research education: Georgetown University, to develop a MOOC focused on Big Data; Rutgers, to create open educational resources (OER) "Enabling Data Science in Biology"; Johns Hopkins University, to build OER to "Facilitate Sharing of Next Generation Sequencing Data;"[2] and New York University (NYU) School of Medicine, to establish online training for "Medical Librarians to Understand and Teach Research Data Management."[3]

The research presented in this article focuses on the outcomes of a course funded by one of these grants, a library-developed MOOC focusing on comprehensive training for managing biomedical data for a broad research audience. The developed course, *Best Practices for*

*Julie Goldman is the Countway Research Data Services Librarian at Harvard Medical School; email: julie_goldman@harvard.edu. Nevada F. Trepanowski is an Information and Data Management Specialist with the USDA this work was completed when she was a graduate student in the School of Library and Information Science at Simmons University; email: nevada.f.trepanowski@usda.gov.

Biomedical Research Data Management,[4] addresses the learning gaps identified in both the LIS curriculum and at institutions that foster scientific research. This article reports on the analysis of 18 months of course survey and assessment data to examine the effectiveness of the course, understanding the educational and professional diversity of participants, and student success at achieving their personal goals. Three research questions were used to guide the analysis:

1. *What types of professionals and nonprofessionals participate in the course?*
2. *What are the participants' motivations for enrolling in the course?*
3. *Does the course address the participants' data management needs?*

Literature Review

During the past decades, research funders have realized the need for public sharing of funded research results[5] and the importance of research data management planning.[6] However, these data management skills were noticeably lacking in the biomedical education curriculum.[7] In response to shifts in how research is conducted and researcher needs, librarians have emerged as key stakeholders, offering data services and training.[8] This literature review looks at RDM training offerings and how the curriculum can be assessed to meet learners' needs.

In 2010, the National Library of Medicine (NLM) funded the development of the New England Collaborative Data Management Curriculum (NECDMC) for teaching data management best practices to undergraduates, graduate students, librarians, and researchers in the health sciences, sciences, and engineering disciplines.[9] NECDMC includes presentation slides and static documents for activities and research cases, intending to provide curricular content that can be adapted for any discipline and learning environment. Nationwide feedback regarding NECDMC was very positive. Librarians used NECDMC at their institutions and praised the case-study approach, the use of hands-on exercises, and incorporating different settings and populations. Participants felt the examples and case studies worked well but wanted more emphasis on a librarian's role in data management consulting. NECDMC was seen as beneficial to many audiences because of its adaptable and flexible framework.[10] However, pilots of the NECDMC identified areas for improving the curriculum:[11]

- Providing answer keys for cases and activities
- Incorporating encryption and security for biomedical data
- Inviting guest speakers from other campus groups, e.g., Office of Research, Institutional Review Board, and Information Technology
- Offering the curriculum as an online course

Similarly, the National Science Foundation (NSF) funded a collaborative education effort that focuses on the needs and practices of RDM for the environmental sciences. DataONE education modules[12] offer a series of eight modules composed of instructional slides. These modules were reviewed and updated in 2016 and migrated to GitHub to increase community engagement.[13] However, DataONE does not offer interactive activities or quizzes, and modules do not include specific concerns for biomedical data, such as confidentiality of human research data.

Other early online training programs include two well-known Coursera MOOCs. "Research Data Management and Sharing" from The School of Information and Library Science and the Odum Institute at the University of North Carolina-Chapel Hill and EDINA at the University of Edinburgh[14] serves as an introductory course to RDM, only focusing on five broad topics. The "Data Management for Clinical Research" offered by the Department of

Biomedical Informatics at Vanderbilt University[15] focuses on clinical research and is geared toward anyone working in medical research, rather than directly targeting librarians and early-career scientists.

More recent developments have expanded on these early online offerings, in addition to the education created through the BD2K Initiative. In 2019, the National Network of Libraries of Medicine Training Office (NTO) developed an eight-week online course to address key concepts in RDM.[16] Most recently, the Research Data Management Librarian Academy (RDMLA) launched in 2020 as a unique partnership between a LIS academic program, academic health sciences and research libraries, and a publisher.[17]

As RDM training transitions to interactive online platforms, instructors must understand how to develop and assess online courses. MOOCs have expanded in the past eight years, to bring high-level education to a larger and wider audience.[18] Although MOOCs provide open and accessible education, they still face some critiques; most notably, they enhance the digital divide and are dominantly Western-centric.[19] Additionally, low student retention rates[20] may not outweigh the cost of MOOC production. These types of courses traditionally have high drop-off rates,[21] and the majority have completion rates of less than 10 percent, not improving during a six-year period.[22] To combat these known limitations, Koutropoulos and Hogue[23] offer recommendations for designing MOOCs that facilitate student interaction at all stages:

- Pre-Course: Provide a clear website with essential information about the course. This ensures that participants understand the course objectives, and commitment, enabling them to make a well-informed decision on joining the course.
- During Course: The instruction platform should be easy to use and understand, allowing participants to focus their time on the content and building connections with other participants instead of troubleshooting technological issues.
- Post-Course: Students should be encouraged to maintain a connection with the materials and anyone they interacted with during the course.

Additionally, effective instruction hinges on understanding how students engage in online learning and how to evaluate the success of their course experiences. One way to evaluate course success is to evaluate student knowledge before taking the course and then again after taking the course. For example, Macleod et al.[24] suggest the inclusion of pre-course and post-course standardized questions. Additionally, data derived from pre- and post-course surveys can help evaluate the success of a MOOC by using the evaluation method developed by Douglas et al.[25] of determining whether participants have achieved their unique learning goals.

Through this literature review, it is clear there is a demand for library and research professionals to develop data skills. This demand has driven the expansion of free online training in these areas. Despite this growth, there are still limited formal proofs of competency. Therefore, this article serves as one example for building a MOOC based on gathered practices for teaching RDM and a framework for evaluating the effectiveness of a course through participants' behaviors.

Course Development

Through funding from the NIH BD2K Initiative Research Education MOOC on Data Management for Biomedical Big Data, the NECDMC curriculum was transformed from static documents into an open online course.[26] To convert these fixed materials into dynamic online

content, instructors were identified to record video presentations, and the online platform Canvas was chosen to facilitate interactive learning activities. Complete course development can be found on the project OSF site: https://osf.io/ac9kg.

To incorporate the suggested improvements of NECDMC[27] and to address new opportunities in biomedical research, the pre-existing seven NECDMC modules were expanded into nine modules for the Canvas course. For example, the course added a tour of a biomedical engineering research laboratory, an example of implementing electronic lab notebooks in a research lab setting, a presentation detailing specific legal policies related to biomedical data, demonstrations of digital tools for data sharing and reusability, and testimonials highlighting research projects that support the discoverability of biomedical digital data.

The nine modules contain a combination of the following elements (see table 1): ungraded pre-module Practice Quiz; video lectures on various data management topics; case study that addresses certain aspects of data management; short activities for the hands-on experience; required and supplemental readings and resources; discussion forum; and post-module Concept Quiz to measure short-term learning outcomes (a full course outline is available on the Open Science Framework, https://osf.io/q4czf). Students could only attempt the ungraded Practice Quiz once but were given two attempts on the Concept Quiz since it contributed to their final grade.

TABLE 1 Best Practices for Biomedical Research Data Management MOOC Curriculum Outline			
Module Topic	**Case Study**	**Activity**	**Assessment**
Course Introduction	–	–	Welcome Survey[†]
1: Introduction and Overview	Identifying Types and Stages of Data*	Research Lifecycle	Practice Quiz Concept Quiz
2: Research Lifecycle	Regeneration of Functional Heart Tissue in Rats*	Data Types, Formats, and Stages*	Practice Quiz Concept Quiz
3: Contextual Details	Combining Data from 10 Years of Research*	Identify Metadata*	Practice Quiz Concept Quiz
4: Data Storage and Security	Studying Vitamin D*	Data Checklist*	Practice Quiz Concept Quiz
5: Data Management Policy	Who Owns Research Data?*	Data Policy Examples	Practice Quiz Concept Quiz
6: Biomedical Ethics	Share and Share Alike?*	De-identifying Data*	Practice Quiz Concept Quiz
7: Data Sharing and Reuse	–	Sharing and Citing Data*	Practice Quiz Concept Quiz
8: Curation and Preservation for Data	Enumeration and Gene-Sequencing*	–	Practice Quiz Concept Quiz
9: Scientific Research Team	–	Apply for an Informationist Grant	Practice Quiz Concept Quiz
Course Conclusion	–	–	Course Assessment Survey User Experience Survey[†]
*Indicates materials already developed for NECDMC. †Indicates survey content provided by Canvas (included in appendices A and D).			

Methods

To evaluate the success of a MOOC, Douglas et al.[28] recommend focusing on whether learners achieved their own learning goals. Descriptive statistics of course analytics alone do not capture the success of a MOOC since both student characteristics (such as learning goals and demographics) and course characteristics (such as instructional design) influence a learner's behaviors.

Course data were analyzed using the recommendations of Douglas et al.[29] for the enrollment period of January 8, 2018, to July 10, 2019. The quantitative approach of using nonparametric descriptive statistics to generate micro and macro analytics from participant responses to course surveys and pre- and post-assessments supported the evaluation of student success and experiences. A quantitative-based method was used over a qualitative approach because it is a cost- and time-efficient method for analyzing the large amounts of student activity that spans an asynchronous MOOC. The evaluation of a single MOOC is, by nature, a nonrandom sample of convenience in that one can analyze the data only from students in the course. However, since the goal of this analysis was only to determine the success of the course, there is no pressing concern over not being able to generalize the analysis to other MOOCs.

Macro-level course analytics were used to examine total enrollment, active participation percentage of total enrollment, diversity of students reached by country and education level, and knowledge achievement. Micro-analytic summaries[30] were used to contextualize the course's impact on students' goals via pre- and post-course surveys as a method to interpret both motivations for enrollment and knowledge achievement. The process of using learner goals and motivation to contextualize course outcomes is suggested by Koller et al. and Douglas et al., who propose that learner intentions provide valuable context to traditional macro-level MOOC metrics.[31] Additionally, micro-analytics are used to capture participant feedback on the course.

Three core questions guided the analysis of the course survey and assessment data:
1. *What types of professionals and nonprofessionals participate in the course?*
2. *What are the participants' motivations for enrolling in the course?*
3. *Does the course address the participants' data management needs?*

Research questions were explored by compiling and visualizing data summaries using Tableau. Anonymized course data was extracted from the Canvas platform; data did contain unique participant ID allowing for response and assessment data for individual students to be linked across course modules and surveys. Participants agreed to data use by Canvas and the instructors through the course service agreement. Exported data were validated and coded in Microsoft Excel. The data validation process is detailed in the Data_Validation_Methods document included in the data analysis files in OSF.[32] Open response data from course surveys were coded to respect participant privacy. Both authors reviewed response coding for congruence. Survey questions are available in appendices A and D, and categories derived from coding are available in appendices B and C. All analysis data files have been deposited in the Open Science Framework.[33]

Macro-Level Analytics
Total Enrollment and Active Participation Percentage of Total Enrollment
Course participation was classified into three categories based on participant activity: 1) *No course content completed*; 2) *Completed one or more modules but not all course content*; and 3) *Com-*

pleted all course content. Participants were classified as having *No course content completed* if all scores in the course grades data set were blank or 0. As stipulated by the category name, participants classified as *Completed one or more modules but not all course content* means the participant may have skipped one or multiple course activities, discussions, or assessments. Additionally, it should be noted that, because the course runs on an asynchronous and continuous schedule, this category does contain participants who are still actively completing the course.

Diversity of Students Reached by Geographic Location and Education Level

Choice-based response data from the course Welcome Survey was used to determine the geographical and educational diversity of participants who enrolled in the course. Responding to the course Welcome Survey was optional; the visualizations produced on geographic and educational diversity only used active response data. Geographic diversity was assessed with the question "Where do you live?" and participants were given the choice of six geographical regions—North America, Asia/Pacific, Europe, Sub-Saharan Africa, the Middle East/North Africa, Latin America—as response options. The educational background of the participants was determined by asking, "What is your highest level of education?"; seven choices—High School or College Preparatory School, Some College but Have Not Finished a Degree, Completed 2-year College Degree, Completed 4-year College Degree, Some Graduate School, Master's Degree (or equivalent), PhD/JD/MD (or equivalent)—were provided.

Knowledge Achievement

Participant knowledge achievement was assessed at a macro level using course assessment data that was analyzed for each of the nine modules and the final course survey. Module assessment data was used to produce score distributions for the practice quiz and the concept quiz, both attempt one and attempt two. Additionally, visualizations were created to capture the overall question success, the percentage a question was answered correctly, for the 10 quiz questions of each module (attempt one and attempt two, independently), and for the final course assessment (attempt one and attempt two, combined). Concept quiz attempts were treated as independent sets for question success, as participants were given all 10 questions in each attempt. Question success data were combined for the final course assessment attempts as participants only completed a random selection of questions from each module in the final assessment. To provide a more robust picture of the participants' mastery of the content, final assessment attempts were combined. Final course assessment data was used to create score distributions for both attempt one and attempt two. Together, the score distribution and question success provide a picture of whether the course content supports participants' gaining knowledge of best practices for data management.

Micro-Level Analytics
Motivations for Enrollment

Data on participants' motivations for enrollment was derived from course Welcome Surveys. Both multiple-choice and open survey questions were used as using data from multiple-choice questions, and open response data together can provide a fuller view of participants' motivations for enrolling in the course. Participant motivation for enrollment was assessed using a choice-based Welcome Survey question: "Question ID 141829: What is your primary reason

for taking an open online course?" Participants were provided with 10 response options: 1) I enjoy learning about topics that interest me; 2) I hope to gain skills for a new career; 3) I like the format (online); 4) I hope to gain skills for a promotion at work; 5) I enjoy being part of a community of learners; 6) I want to try Canvas Network; 7) I am curious about MOOCs; 8) I am preparing to go back to school; 9) No answer provided; 10) I am preparing for college for the first time.

Participants' specific goals for enrolling in the course were assessed with an open response question: "ID 141832: How will this course help you meet your personal or professional goals?" Response data were evaluated and organized into 12 categories. A table of the 12 categories and a brief explanation of the parameters of each are provided in appendix B.

The combined goal responses for *General Knowledge* and *Skill Development* were then classified using a separate "Application" category to capture the setting in which the respondent would use the knowledge or skill gained from the course. Four settings were identified after assessing responses: 1) Educational Setting, 2) Professional Setting, 3) Personal growth—response specifically mentioned personal growth, and 4) Unspecified—the response did not include a setting of where the skills or knowledge gained from the course would be applied. For responses categorized as *Skill Development,* the primary skills described were documented as stated for each response. The described skills were then standardized into 17 categories, listed in the table in appendix C.

Course Efficacy at Meeting Participants' Data Management Goals

Data on the course's efficacy at meeting participants' data management goals was assessed at the end of the course User Experience Survey using an open response question: "Question ID 141845: In what ways has this course helped you meet your personal or professional goals?" Responses were grouped using the same method described in the Motivation for Enrollment section. Categorized data was then linked to individual participant responses from the course Welcome Survey for all participants who completed both surveys using the participants' unique ID to examine if participants successfully achieved the goals they identified at the beginning of the course.
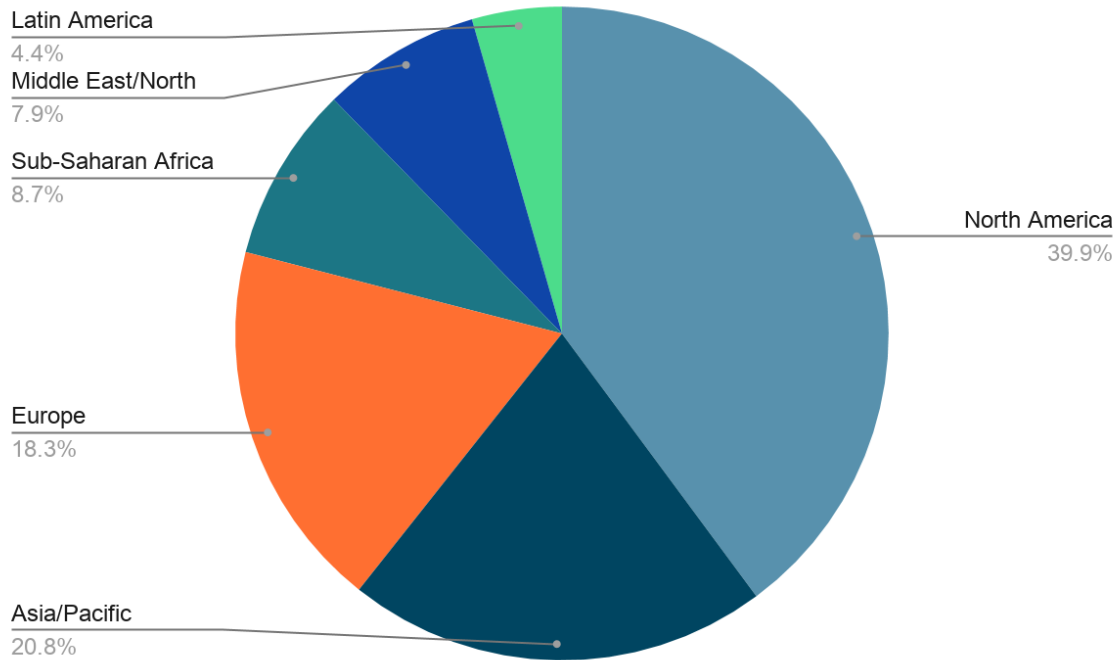
Results
Macro-Level Analytics
Total Enrollment and Active Participation Percentage of Total Enrollment

From January 8, 2018, to July 10, 2019, 1,308 participants enrolled in the course. No course content was completed by 33.87 percent (443 participants) of the total enrolled participants. At least one module was completed by 61.31 percent (802 participants). All course content was completed by 4.82 percent (63 participants) of the 1,308 participants enrolled. The greatest proportion of course participants are from North America (see figure 1), and the majority of course participants hold an advanced degree, with a master's degree being the most common (see figure 2).
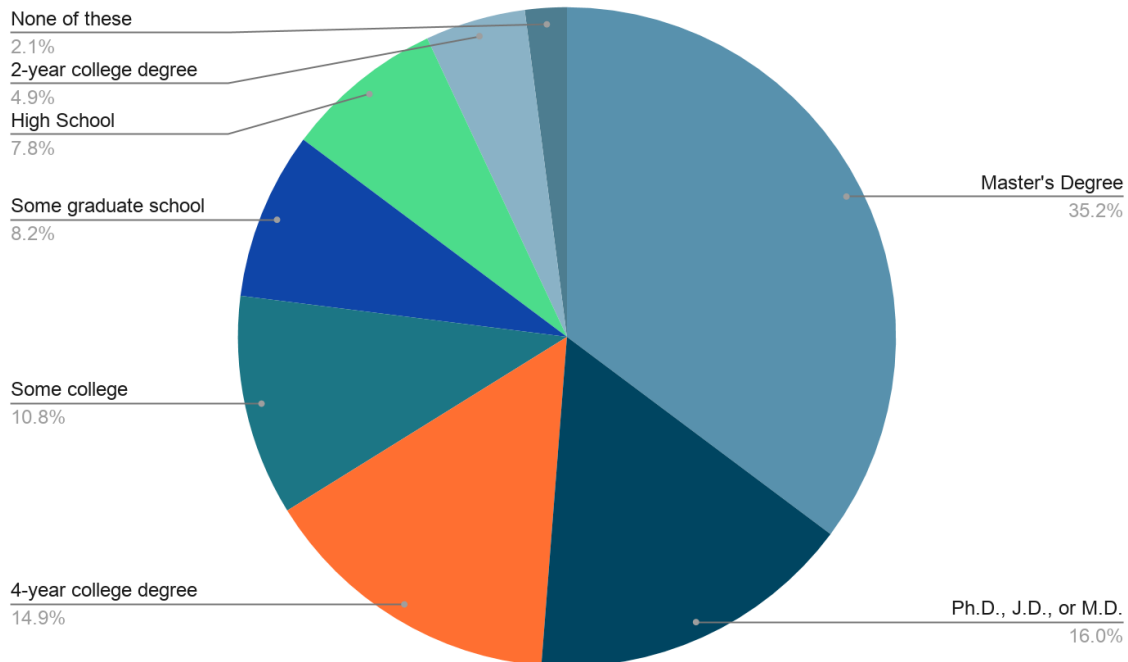
Knowledge Achievement

This was the original research question for module data: *"Does the course address the participants' data management needs?"* This question cannot be answered fully from the dataset alone. Individual participants' specific data management needs are likely too multifaceted to an-

FIGURE 1
Geographic Diversity of Students Reached by Country Who Provided Biographical Data in the Course Welcome Survey*

Latin America
4.4%

Middle East/North
7.9%

Sub-Saharan Africa
8.7%

North America
39.9%

Europe
18.3%

Asia/Pacific
20.8%

*Six geographic locations were provided, and respondents self-reported their location. Responses were received for 728 students.

FIGURE 2
The Highest Level of Education of Course Participants*

None of these
2.1%

2-year college degree
4.9%

High School
7.8%

Some graduate school
8.2%

Master's Degree
35.2%

Some college
10.8%

4-year college degree
14.9%

Ph.D., J.D., or M.D.
16.0%

*Seven education levels were provided, and respondents self-reported their credentials. Responses were received for 730 students.

swer with any granularity without in-depth interviews of each participant. That having been said, the research question intended to examine whether the course content was effective in helping participants develop competency in the best practices for biomedical research data management. Table 4 depicts module summary data for both assessment attempts 1 and 2 for all nine modules, including any modules completed by participants. The module assessment data support the assertion that the course content is effective in helping participants master the course learning objectives. The median scores increase for all modules from attempt 1 to attempt 2. Across modules, attempt 1 median scores have a flat trend, with a range of 40 to 61 percent and an average of ~53 percent. However, attempt 2 median scores have an upward trend with a range of 64.2 to 87.9 percent and an average of ~79 percent. The growth seen across the modules in attempt 2 scores over attempt 1 scores indicates that participants did improve their understanding of the topics presented in the modules. Additionally, the theory of improved understanding building throughout the course is also supported by the final course assessment scores. Both attempts of the final course assessment have a median of ~65 percent, only slightly lower than the 70 percent required for a passing score, with 25 percent of students in both attempts scoring above 80 percent.

TABLE 2
Data depicted in the table includes the median assessment score and interquartile range (IQR) of the scores for both attempts 1 and 2 of the nine modules of the course
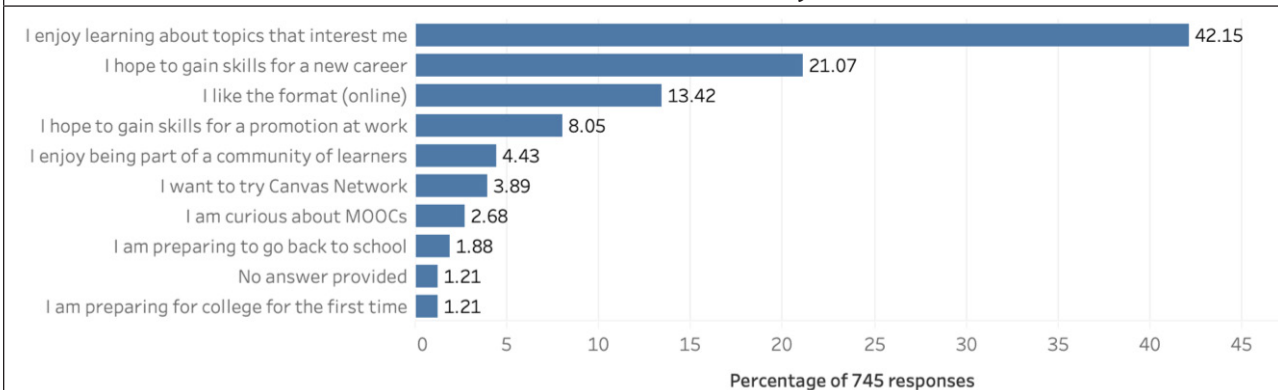
Module	Attempt 1			Attempt 2		
	Total Participants	Median Score	IQR	Total Participants	Median Score	IQR
Module 1: Introduction and Overview	233	51.7	39.1	161	64.2	50
Module 2: Research Lifecycle	164	56	41.3	114	77.5	39
Module 3: Contextual Details	139	61.1	43.7	87	86	23.3
Module 4: Data Storage and Security	130	54.8	33.4	92	83.1	21.7
Module 5: Data Management Policy	126	40.5	44.8	83	80.9	18.5
Module 6: Biomedical Ethics	76	59.6	38.7	47	87.9	20
Module 7: Data Sharing and Reuse	116	41.5	39.4	78	76.8	31.9
Module 8: Curation and Preservation for Data	112	54.4	38	75	75.5	33
Module 9: Scientific Research Team	111	60	43.3	69	82.4	29.6
Final Course Assessment	105	66.7	44.5	38	65.3	32.6

Micro-Level Analytics
Motivations for Enrollment

The course Welcome Survey question "141829: What is your primary reason for taking an open online course?" assessed the students' motivation for enrolling in the course via a multiple-choice question. The three most common participant motivations for enrolling in the course comprise ~76 percent of the total response pool (see figure 3). *The enjoyment for learning about topics that interest the participants* was the most common response. *The hope for gaining skills for a new career* and *a preference for an online instruction format* were the second and third most frequent responses, respectively.

FIGURE 3
Participant Enrollment Motivation Derived from an Open-Response Question in the Course Welcome Survey*

Motivation	Percentage
I enjoy learning about topics that interest me	42.15
I hope to gain skills for a new career	21.07
I like the format (online)	13.42
I hope to gain skills for a promotion at work	8.05
I enjoy being part of a community of learners	4.43
I want to try Canvas Network	3.89
I am curious about MOOCs	2.68
I am preparing to go back to school	1.88
No answer provided	1.21
I am preparing for college for the first time	1.21

Percentage of 745 responses

*Reported as a percentage of 745 total responses

Participant Goals

Participant motivation for enrollment is mirrored in the three most prevalent goals participants believed would be supported by the course. Course goals were assessed via an open response question—"141832: How will this course help you meet your personal or professional goals?"—in the course Welcome Survey. *General knowledge*, *Skill development*, and *Career growth* comprised the top three specific participant goals and account for ~62 percent of responses. The qualifier "specific goals" is used because the second most common response was "no answer." Additionally, it is important to note that this was an open-response question and not a standard-choice question; therefore, some respondents may have been deterred from answering. The goal of *General knowledge* is in alignment with the most common participant motivation from the previous section, an enjoyment for learning. Similarly, *Skill development* and *Career growth* are closely associated with the second most common motivation, *Hoping to gain skills for a new career*.

Setting in Which Course Goals of General Knowledge or Skills Would Be Applied
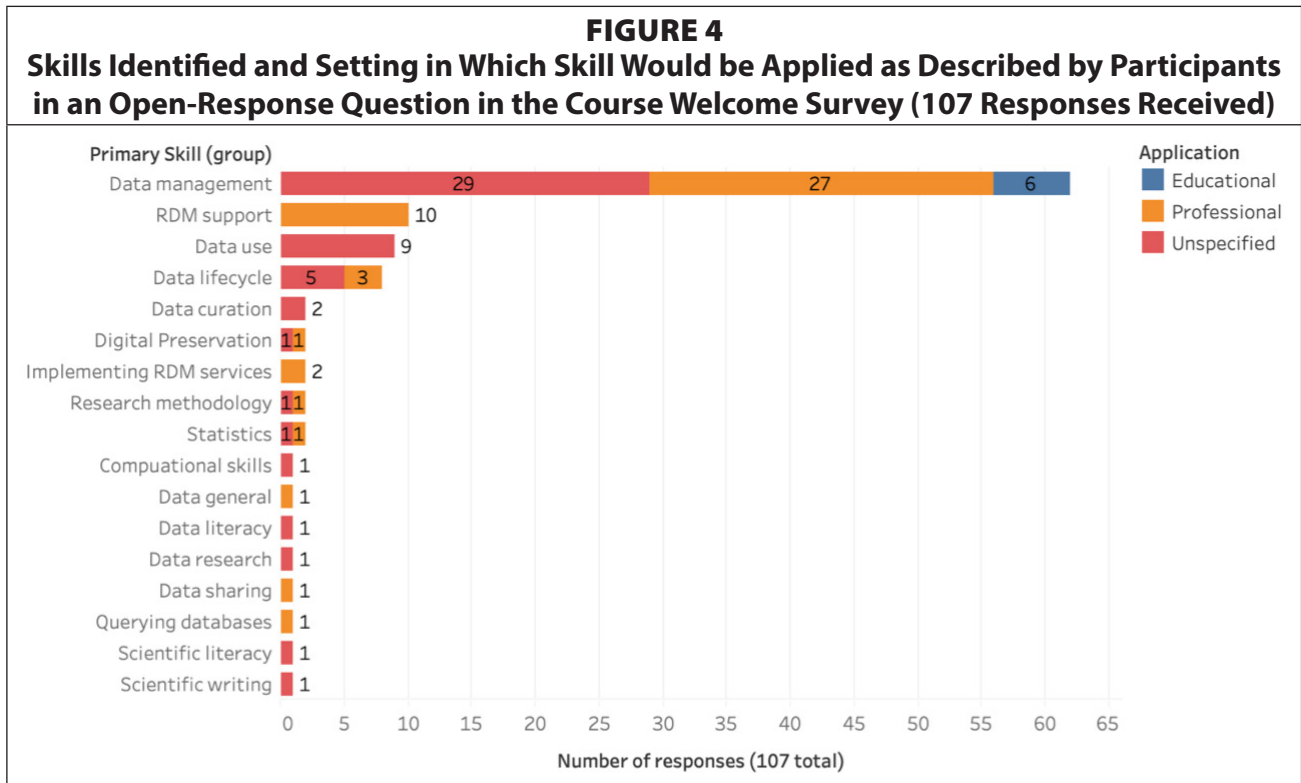
To determine the types of professionals and nonprofessionals enrolled in the course, goal responses for general knowledge and skill development were analyzed to determine the setting in which participants would apply the knowledge or skills gained from the course. It is important to note that participants were not specifically asked about the setting for applying their achieved goals. However, after reviewing the responses, many respondents did include a setting. Thus, the decision was made to pull settings out from the responses when possible, to provide additional depth to the data. Settings identified included *Professional setting* (relating to a work environment), *Educational setting* (related to school at any level of education), and *Personal life* (only used when participants directly specified "personal or private life").

General Knowledge Application Setting

The majority (~52%) of goal responses did not identify a specific setting in which they would apply the *General knowledge* gained from the course. However, of those who did provide a setting, a *Professional setting* (31.78%) was the most common by nearly a factor of 3 from the second most common response of an *Educational setting* (10.85%), followed by *Personal life* (4.65%).

Skill Application Setting

Figure 4 depicts the skills participants hoped to gain by completing the course and the setting in which they would apply the skill. Since setting was not directly asked for in the Welcome Survey question, we derived this from responses that explicitly stated a place they would use the skill. Similar to the general knowledge goal data, many of the skill-oriented responses did not identify a setting (49.53%) for applying the specific skills gained from the course. Likewise, the respondents who did include a setting also identified a *Professional setting* (44.86%) as the most common response by a factor of nearly 8 to that of an *Educational setting* (5.61%).
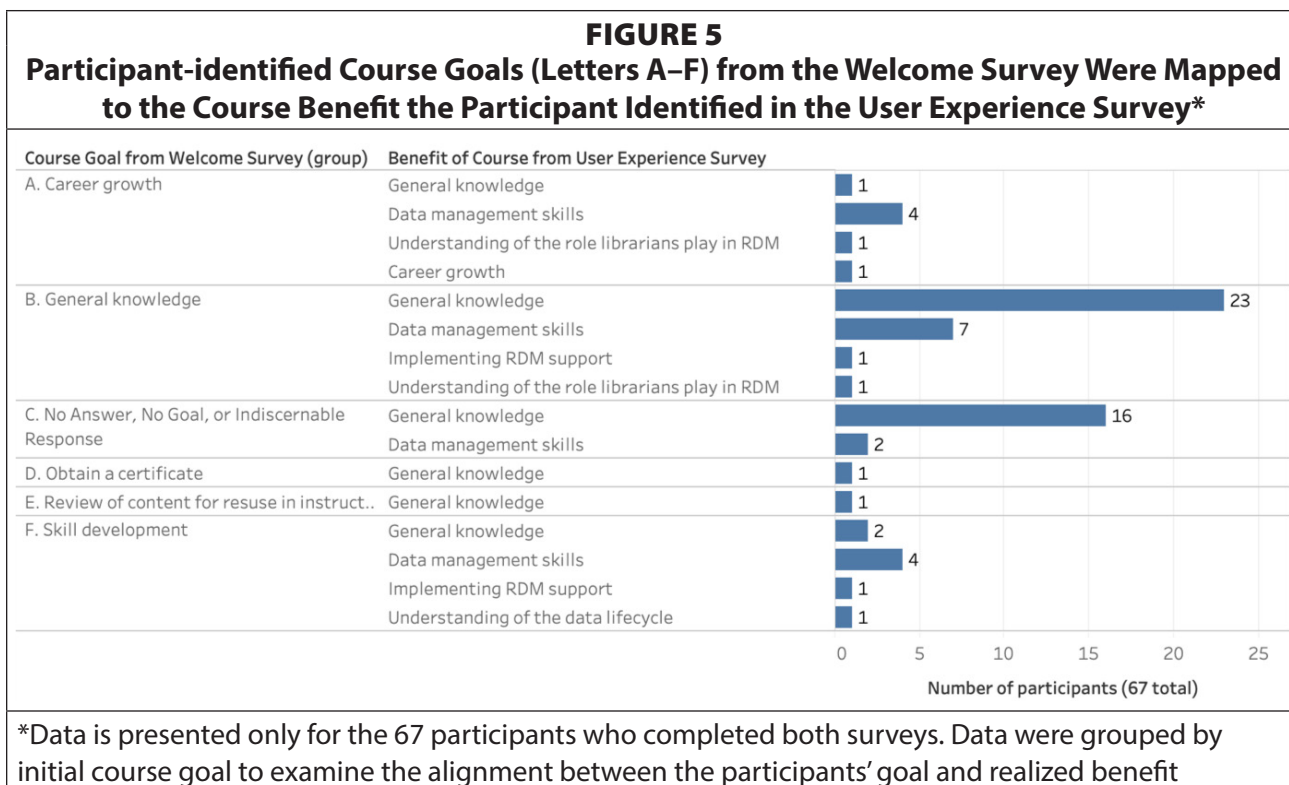
FIGURE 4
Skills Identified and Setting in Which Skill Would be Applied as Described by Participants in an Open-Response Question in the Course Welcome Survey (107 Responses Received)

Skills Identified by Participants

Specific skills were also identified and grouped for respondents specifying skill development as a course goal to determine the types of skills participants hoped to gain from the course. Additionally, the identified skills were then linked to the setting in which the skill would be used as described by the participant. The most common skill identified in all settings (professional, educational, and unspecified) was *Data management*. All *Educational setting* responses identified *Data management*. The second most common skill for a *Professional setting* response was *Research data management (RDM) support*.

Course Efficacy at Meeting Participants' Data Management Goals

To determine whether the course content supported participants in achieving the goals they identified at the beginning of the course, goals from the Welcome Survey were linked to participant-identified benefits in the User Experience Survey. The goals that would be supported by the course included: A) *Career growth*; B) *General knowledge*; C) *No answer, no goal, or indiscernible response*; D) *Obtain a certificate*; E) *Review of content for reuse in instruction*; and F)

Skill development. Nonspecific goals of "*No answer, no goal or indiscernible response*" were kept in the analysis because the linked participants provided benefits in the User Experience Survey. In the User Experience Survey, participants identified six benefits derived from completing the course: *General knowledge, Data management skills, Understanding of the role librarians play in RDM, Career growth, Implementing RDM support,* and *Understanding the research data lifecycle.* For 67 participants who completed both surveys, combined open-answer responses from the course Welcome Survey and the User Experience Survey have been linked (see figure 5).

FIGURE 5
Participant-identified Course Goals (Letters A–F) from the Welcome Survey Were Mapped to the Course Benefit the Participant Identified in the User Experience Survey*

Course Goal from Welcome Survey (group)	Benefit of Course from User Experience Survey	Number of participants (67 total)
A. Career growth	General knowledge	1
	Data management skills	4
	Understanding of the role librarians play in RDM	1
	Career growth	1
B. General knowledge	General knowledge	23
	Data management skills	7
	Implementing RDM support	1
	Understanding of the role librarians play in RDM	1
C. No Answer, No Goal, or Indiscernable Response	General knowledge	16
	Data management skills	2
D. Obtain a certificate	General knowledge	1
E. Review of content for resuse in instruct..	General knowledge	1
F. Skill development	General knowledge	2
	Data management skills	4
	Implementing RDM support	1
	Understanding of the data lifecycle	1

*Data is presented only for the 67 participants who completed both surveys. Data were grouped by initial course goal to examine the alignment between the participants' goal and realized benefit

Group A's goal is *career growth* and matches four benefits: *General knowledge, Data management skills, Understanding of the role librarians play in RDM,* and *Career growth.* The three benefits that do not specifically state career growth are still easy to associate with professional development that would support career growth.

Group B's goal is *General knowledge* and corresponds to four benefits: *General knowledge, Data management skills, Implementing RDM support,* and *Understanding of the role librarians play in RDM.* This group has the largest alignment between initial course goal and derived course benefit, with 23 participants setting and achieving a goal of general knowledge on the best practices of biomedical research data management. Data management skills had the next largest benefit response in the group. While this response is not explicitly oriented toward general knowledge, it could be argued that skills development requires general knowledge and the more advanced task of applying knowledge.

Group C included all nonspecific goals, including responses with answers (no answer), no goal specified but contained answers (no goal), and answers that could not be discerned because they were in another language or included nonstandard characters (indiscernible response). Two benefits were mapped to this initial course goal including *General knowledge*

and *Data management skills*. These responses were included in the visualization because they illustrate that, even when the participant began the course without a specific goal, they were still able to identify a benefit after the course.

Group D is made up of a single participant whose initial course goal was to *Obtain the course certificate*. The benefit identified at the end of the course was *General knowledge*; while this is not a perfect match, one could make the argument that the certificate of completion is proof that the participant has amassed a general knowledge on the best practices of biomedical research data management.

Group E, like group D, is made up of a single respondent. The participant's course goal was *a Review of content for reuse in instruction*, and the participant identified the benefit of *General knowledge* at the conclusion of the course. While there is no way to know if the participant's need for course instruction was met, having a general knowledge of the course would give the best information for making any decisions on reuse.

Group F's initial course goal was *Skill development*; participants identified four benefits after the course: *General knowledge*, *Data management skills*, *Implementing RDM support*, and *Understanding the research data lifecycle*. Similar to groups A and B, the benefit that matched the goal for the category had the largest number of participants' responses. Additionally, the benefits of *Implementing RDM support* and *Understanding the research data lifecycle* can certainly be argued to have a connection to skills development in that the benefit would either actively require the use or understanding of data management skills.

The alignment seen in the course goals and benefits across the groups makes it clear that the course is effectively supporting the participants' identified research data management needs.

Discussion

Based on the module and assessment data, it is reasonable to assume that participants were successful in internalizing and connecting course material across modules to develop an understanding of the best practices for biomedical research data management. A total of 1,308 participants enrolled in the course during that period. Active enrollment accounts for 865 participants (~66.2% of total enrollment) who completed at least one module of the course; 443 participants (~33.8% of total enrollment) enrolled but did not complete any course content. Approximately 18 months after the course was first launched, 99 participants have completed the course; 75 of those participants achieved a passing score of 70 percent or higher. The data indicates that the course curriculum was effective at supporting participants in building competency in biomedical data management practices. Furthermore, the final course pass rate among participants who completed at least one module was 17 percent, a value that is above the standard 10 percent seen in MOOCs.[34]

Analysis of the course data indicates that course participants are primarily professionals from North America, Asia, and Europe who hold at least four-year college degrees, with approximately half holding an advanced secondary degree. Primary motivators for enrolling in the course are the enjoyment of learning followed by skills for a new career and a preference for the online format. The most common course goals were centered around advancing the participants' professional lives, either by gaining general knowledge of biomedical research data management or by building data management skills.

Analysis of the course assessments, final grades, participant-identified goals, and derived course benefits indicate that the course is meeting the data management needs of participants

who complete the course. Participants' scores across modules support the assertion that participants are increasing their understanding of the topics presented in the modules. Additionally, the question success rate from the modules is sustained for the majority of questions in the final course assessment, indicating that participants retained the knowledge gained in the modules. Of the participants who completed the course, three quarters achieved a passing score. Mapping of initial course goals to participant-identified derived course benefits indicates that the majority of participants who provided this data gained the skills or knowledge required to support their goal. Thus, it is reasonable to surmise that the course is effective in providing participants with tools to build an understanding of the best practices for biomedical research data management.

In keeping with the recommendations from Koutropoulos and Hogue,[35] participants retain access to the course on Canvas, even after they have completed the course. This allows participants to access valuable reference materials to help them in their work or additional studies and interact with other students from the class. Additionally, a Course Wiki Guide has been created and shared with anyone interested in referencing or reusing the curriculum in the future.[36]

Limitations
As mentioned earlier, the evaluation of this individual course is an instance of convenience sampling, in that only data from students in the course were analyzed. However, since the goal of this research was to determine the educational success of a single sample group, there is no pressing concern over not being able to generalize this analysis to other MOOCs. Instead, the goal of the work presented here is to serve as an example of how MOOCs on research data management can be developed. Additionally, this paper serves as a model for evaluating the success of a MOOC through the lenses of traditional macro-level analytics such as course completion rate and micro-level analysis of whether students achieved their self-identified course goals.

Another limitation relates to the course development and curriculum focus of the course. This course was developed by librarians to transform current standalone RDM training materials into a more comprehensive online course. Repurposing the existing curriculum limited the topics covered in the course. When the course was created, there was a substantial gap in the availability of online RDM training for the biomedical sciences; designing a new curriculum would have increased development time and inadvertently contributes to the void in resources.

Follow-up Research
The substantial monetary and time investment for developing and managing a MOOC, coupled with a course's ability to extend the reach of the university's mission to a wide audience of learners, highlights the necessity of thorough evaluation. However, in the process of planning the evaluation methodology for this course, it became clear that the majority of literature on MOOCs centers largely from a research perspective where multiple MOOC courses are evaluated to explore elements related to knowledge creation and learner retention. Future research efforts will be directed at expanding the methodology used in this paper into an evaluation framework designed for use by practitioners interested in assessing the efficacy of a single MOOC. The need for this type of evaluation framework stems from three

key factors: 1) MOOCs continue to grow in popularity in the modern education landscape; 2) MOOCs require extensive institutional resources for successful development and maintenance; 3) MOOCs have institutional value and impact as tools for extending the reach of their sponsoring organization's mission.

Conclusions

The content in *The Best Practices for Biomedical Research Data Management* addresses the variety of areas involved in data management. The course was successful in using feedback from previous RDM training to offer a free online course aimed at building the data skills of the research community. In 18 months, the course successfully reached more than 1,000 people around the world, increased their understanding of data management topics, and successfully supported participants' learning goals. The success of this course shows how free online content can have an impact on data services knowledge.

Since the release of this course in 2018, additional reports have been published on the skills that librarians should develop to provide data services. While technical skills are important for offering targeted and in-depth services, soft and traditional library skills allow for high-quality and successful implementation of data services.[37] This landscape is still evolving. As the library's role related to data services evolves, it is more important than ever to find ways to develop effective free, online, interactive professional development opportunities and to evaluate the success of those courses in helping participants achieve their goals.

Declarations

APPENDIX A

Welcome Survey (Required by Canvas Network)

141829: What is your primary reason for taking an open online course?
- ☐ I like the format (online)
- ☐ I enjoy learning about topics that interest me
- ☐ I enjoy being part of a community of learners
- ☐ I hope to gain skills for a new career
- ☐ I hope to gain skills for a promotion at work
- ☐ I am preparing to go back to school
- ☐ I am preparing for college for the first time
- ☐ I am curious about MOOCs
- ☐ I want to try Canvas Network

141830: Not everyone has the same participation and learning goals. We welcome the diversity. Which type of online learner best describes you?
- ☐ An observer. I just want to check the course out. Count on me to "surf" the content, discussions, and videos but don't count on me to take any form of assessment.
- ☐ A drop-in. I am looking to learn more about a specific topic within the course. Once I find it and learn it, I will consider myself done with the course.
- ☐ A passive participant. I plan on completing the course but on my own schedule and without having to engage with other students or assignments.
- ☐ An active participant. Bring it on. If it's in the course, I plan on doing it.

141831: How many hours a week are you planning to spend on this course?
- ☐ Less than 1 hour
- ☐ Between 1 and 2 hours
- ☐ Between 2 and 4 hours
- ☐ Between 4 and 6 hours
- ☐ Between 6 and 8 hours
- ☐ More than 8 hours per week

141832: How will this course help you meet your personal or professional goals? [open ended]

141833: What is your highest level of education?
- ☐ High School or College Preparatory School
- ☐ Some college, but have not finished a degree
- ☐ Completed 2-year college degree
- ☐ Completed 4-year college degree
- ☐ Some graduate school
- ☐ Master's Degree (or equivalent)
- ☐ PhD, JD, or MD (or equivalent)
- ☐ None of these

141834: Is English your primary spoken language?
- ☐ Yes
- ☐ No

141835: Where do you live?
- ☐ North America
- ☐ Latin America
- ☐ Europe
- ☐ Middle East/North Africa
- ☐ Sub-Saharan Africa
- ☐ Asia/Pacific

141836: What is your gender?
- ☐ Male
- ☐ Female
- ☐ Other

141837: How old are you?
- ☐ 13–18
- ☐ 19–24
- ☐ 25–34
- ☐ 35–44
- ☐ 45–54
- ☐ 55–64
- ☐ 65 or older

141838: How did you hear about this Canvas Network Course? (select all that apply)
- ☐ Through a social media site (like Facebook or Twitter)
- ☐ From a news story (print, online, radio, or TV) that mentioned the course and/or Canvas Network
- ☐ From a friend or colleague
- ☐ I clicked on an ad
- ☐ From a web search
- ☐ From the instructor
- ☐ From a Canvas or Canvas Network communication
- ☐ From the sponsoring institution (newsletter, institution's website/blog, or flyer)

141839: Where have you taken an online course before? (Select all that may apply)
- ☐ Never taken an online course
- ☐ At school
- ☐ Canvas Network
- ☐ Coursera
- ☐ EdX
- ☐ Udacity
- ☐ FutureLearn
- ☐ Other

141840: If you have any general feedback you'd like to provide, please do so here: [open ended]

APPENDIX B
Categorization of Participant Open Response Answers to question 141832: How will this course help you meet your personal or professional goals?

#	Category	Parameters
1	Career growth	The response specifically mentions taking the course to advance the respondent's current or future career goals.
2	Gain experience with Canvas Network	The response explicitly states Canvas as being the motivation for taking the course rather than the course content.
3	General knowledge	The response includes themes of learning and/or mastering content without focusing on a specific skill.
4	Indiscernible answer	The response was either already categorized as indiscernible from the validation stage or was found to be irrelevant to the question being asked.
5	No answer	The response was left blank.
6	No goal specified	The response did not specify a goal or fit within the parameters of the other categories.
7	Not in English	The response was in a language other than English.
8	Obtain a certificate	The response states the end of course certificate of completion as their primary goal.
9	Required to take course	The response stated that taking the course was a required activity for another course or work.
10	Review of content for reuse in instruction	The response stated that the goal of taking the course was to assess it for inclusion in another course or training.
11	Reviewing course design	The response stated that their primary goal was to learn and/or assess the course design.
12	Skill development	The response identified developing a specific skill as being their primary goal.

APPENDIX C
Skill Categories Identified from Open Participant Responses

#	Skill	#	Skill
1	Computational skills	10	Digital preservation
2	Data curation	11	Implementing RDM services
3	Data general	12	Querying databases
4	Data lifecycle	13	RDM support
5	Data literacy	14	Research methodology
6	Data management	15	Scientific literacy
7	Data research	16	Scientific writing
8	Data sharing	17	Statistics
9	Data use		

APPENDIX D

User Experience Survey (Required by Canvas Network)

141842: How strongly do you agree or disagree with the following statement: The course materials (lectures, videos, documents) have a positive impact on my learning experience.

- ☐ Strongly Disagree
- ☐ Disagree
- ☐ Neither Agree nor Disagree
- ☐ Agree
- ☐ Strongly Agree

141843: How strongly do you agree or disagree with the following statement: The course activities (discussions, assignments, projects, quizzes) have a positive impact on my learning experience.

- ☐ Strongly Disagree
- ☐ Disagree
- ☐ Neither Agree nor Disagree
- ☐ Agree
- ☐ Strongly Agree

141844: How many hours a week are you spending on this course?

- ☐ Less than 1 hour
- ☐ Between 1 and 2 hours
- ☐ Between 2 and 4 hours
- ☐ Between 4 and 6 hours
- ☐ Between 6 and 8 hours
- ☐ More than 8 hours per week

141845: In what ways has this course helped you meet your personal or professional goals? [open ended]

141846: How likely are you to recommend a course on Canvas Network to a friend?

- ☐ 0 – Not Likely
- ☐ 1
- ☐ 2
- ☐ 3
- ☐ 4
- ☐ 5 – Neutral
- ☐ 6
- ☐ 7
- ☐ 8
- ☐ 9
- ☐ 10 – Extremely Likely

141847: Please give this course an overall rating on a scale of 1 to 5 with 1 being the lowest and 5 being the highest rating.

☐ 1 star
☐ 2 stars
☐ 3 stars
☐ 4 stars
☐ 5 stars

141848: How much instructor involvement do you like to have in your online learning experiences?
☐ I like to learn on my own
☐ I prefer peer-to-peer interactions with my classmates (social learning)
☐ I prefer to communicate only with the instructor
☐ I like variety
☐ I do not interact with my instructor

141849: Ideally, how long should Canvas Network Course last?
☐ 0–2 weeks
☐ 2–4 weeks
☐ 4–6 weeks
☐ 6–8 weeks
☐ 8 weeks or more

141850: How strongly do you agree or disagree with the following statement? I have a positive user experience when I access my course on my smartphone (e.g., iPhone, Android phone).
☐ I do not use a smartphone to access my course
☐ Strongly Disagree
☐ Disagree
☐ Neither Agree nor Disagree
☐ Agree
☐ Strongly Agree

141851: How strongly do you agree or disagree with the following statement? I have a positive user experience when I access my course on my tablet device (e.g., iPad, Nexus).
☐ I do not use a tablet device to access my course
☐ Strongly Disagree
☐ Disagree
☐ Neither Agree nor Disagree
☐ Agree
☐ Strongly Agree

141852: If you'd like to provide any general feedback on the course, please do so here: [open ended]

Notes

1. National Library of Medicine, "NIH Big Data to Knowledge (BD2K) Grants" (n.d.), https://www.nlm.nih.gov/ep/AwardsBD2K.html#2015 [accessed 25 May 2021].

2. Robert A.Wright, "Developing a Suite of Online Learning Modules on the Components of Next-Generation Sequencing Projects," *Medical Reference Services Quarterly* vol. 39, no. 1 (2020): 90–99, https://doi.org/10.1080/0276 3869.2020.1688623.

3. Kevin B. Read et al., "A Two-Tiered Curriculum to Improve Data Management Practices for Researchers," *PloS One* 14, no. 5 (2019): e0215509, https://doi.org/10.1371/journal.pone.0215509.

4. Canvas Network, "Best Practices for Biomedical Research Data Management" (n.d.), https://www.canvas. net/browse/harvard-medical/courses/biomed-research-data-mgmt [accessed 26 April 2021].

5. National Institutes of Health, "Final NIH Statement On Sharing Research Data" (n.d.), https://grants.nih. gov/grants/guide/notice-files/NOT-OD-03-032.html [accessed 26 April 2021].

6. National Science Foundation, "Scientists Seeking NSF Funding Will Soon Be Required to Submit Data Management Plans" (2010), https://www.nsf.gov/news/news_summ.jsp?cntn_id=116928 [accessed 26 April 2021].

7. Nicholas R. Anderson et al., "Issues in Biomedical Research Data Management and Analysis: Needs and Barriers," *Journal of the American Medical Informatics Association* 14, no. 4 (2007): 478–88, https://doi.org/10.1197/ jamia.M2114; Lisa M. Federer, Ya-Ling Lu, and Douglas J. Joubert, "Data Literacy Training Needs of Biomedical Researchers," *Journal of the Medical Library Association* 104, no. 1 (2016): 52–57, https://doi.org/10.3163/1536-5050.104.1.008.

8. Tania P. Bardyn et al., "Health Sciences Libraries Advancing Collaborative Clinical Research Data Management in Universities," *Journal of eScience Librarianship* 7, no. 2 (2018), https://doi.org/10.7191/jeslib.2018.1130; Kevin B. Read, "Adapting Data Management Education to Support Clinical Research Projects in an Academic Medical Center," *Journal of the Medical Library Association* 107, no. 1 (2019): 89, https://doi.org/10.5195/jmla.2019.580.

9. Donna Kafel, Andrew Creamer, and Elaine Martin, "Building the New England Collaborative Data Management Curriculum," *Journal of eScience Librarianship* 3, no. 1 (2014): 60–66, https://doi.org/10.7191/jeslib.2014.1066.

10. Mayu Ishida, "The New England Collaborative Data Management Curriculum Pilot at the University of Manitoba: A Canadian Experience," *Journal of eScience Librarianship* 3, no. 1 (2014): 80–85, https://doi.org/10.7191/ jeslib.2014.1061.

11. Ishida, "The New England Collaborative Data Management Curriculum Pilot at the University of Manitoba"; Christie Peters and Porcia Vaughn, "Initiating Data Management Instruction to Graduate Students at the University of Houston Using the New England Collaborative Data Management Curriculum," *Journal of eScience Librarianship* 3, no. 1 (2014): 86–99, https://doi.org/10.7191/jeslib.2014.1064; Amanda L. Whitmire, "Implementing a Graduate-Level Data Information Literacy Curriculum at Oregon State University: Approach, Outcomes and Lessons Learned," University of Massachusetts Medical School (n.d.), https://doi.org/10.13028/351S-G605.

12. Heather Henkel et al., "DataONE Education Modules," *Data Observation Network for Earth,* https://www. dataone.org/education-modules [accessed 5 June 2020].

13. Heather Soyka et al., "Using Peer Review to Support Development of Community Resources for Research Data Management," *Journal of eScience Librarianship* 6, no. 2 (2017): e1114, https://doi.org/10.7191/jeslib.2017.1114.

14. Robin Rice, "New MOOC! Research Data Management and Sharing," Edinburgh Research Data Blog (2016), http://datablog.is.ed.ac.uk/2016/02/24/new-mooc [accessed 5 June 2020].

15. Stephany Duda and Paul Harris, "Data Management for Clinical Research" (n.d.), https://www.coursera. org/course/datamanagement [accessed 26 April 2021].

16. Jessica Van Der Volgen and Shirley Zhao, "Building a National Research Data Management Course for Health Information Professionals," *Journal of eScience Librarianship* 8, no. 1 (2019): e1160, https://doi.org/10.7191/ jeslib.2019.1160.

17. Research Data Management Librarian Academy, "RDMLA" (n.d.), https://rdmla.github.io [accessed 5 June 2020].

18. Justin Reich and José A. Ruipérez-Valiente, "The MOOC Pivot: From Teaching the World to Online Professional Degrees," *Science* 363, no. 6423 (2019): 130–31, https://doi.org/10.1126/science.aav7958.

19. Taskeen Adam, "Digital Neocolonialism and Massive Open Online Courses (MOOCs): Colonial Pasts and Neoliberal Futures," *Learning, Media and Technology* 44, no. 3 (2019): 365–80, https://doi.org/10.1080/17439884 .2019.1640740.

20. Daphne Koller et al., "Retention and Intention in Massive Open Online Courses: In Depth," *Educause Review* 48 (2013): 62–63, https://er.educause.edu/articles/2013/6/retention-and-intention-in-massive-open-online-courses-in-depth.

21. Apostolos Koutropoulos and Rebecca Hogue, "How to Succeed in a Massive Online Open Course (MOOC)," *Learning Solutions Magazine* (2012), https://learningsolutionsmag.com/articles/1023/how-to-succeed-in-a-massive-online-open-course-mooc [accessed 5 June 2020].

22. Tharindu Liyanagunawardena, Shirley Williams, and Andrew Adams, "The Impact and Reach of MOOCs: A Developing Countries' Perspective," *eLearning Papers* 33 (2013), https://centaur.reading.ac.uk/32452/; Reich and

Ruipérez-Valiente, "The MOOC Pivot."

23. Koutropoulos and Hogue, "How to Succeed in a Massive Online Open Course."

24. Hamish Macleod et al., "Emerging Patterns in MOOCs: Learners, Course Designs and Directions," *Tech-Trends* 59 (2014): 56–63, https://doi.org/10.1007/s11528-014-0821-y.

25. Kerrie A. Douglas et al., "Meaningful Learner Information for MOOC Instructors Examined through a Contextualized Evaluation Framework," *International Review of Research in Open and Distributed Learning* 20, no. 1 (2019), https://doi.org/10.19173/irrodl.v20i1.3717.

26. Julie Goldman, "Pitfalls and Positives: Developing a Massive Open Online Course," presented at North Atlantic Health Science Libraries 2017, Waltham, MA, Open Science Framework, https://doi.org/10.17605/OSF.IO/FMD4C; Julie Goldman and Elaine Martin, "Best Practices for Biomedical Research Data Management," Open Science Framework (2017), https://doi.org/10.17605/OSF.IO/VRNFX; Julie Goldman, "Designing Collaborative Online Training for Research Data Management," poster presented at the University of Massachusetts and New England Area Librarian eScience Symposium 2018, Worcester, MA, Open Science Framework, https://doi.org/10.17605/OSF.IO/NQM72; Julie Goldman and Allison Herrera, "MOOC: Miscalculations, Oversights, Opportunities and Celebration," poster presented at ACRL New England Chapter Annual Conference 2018, Plymouth, MA, Open Science Framework, https://doi.org/10.17605/OSF.IO/M3UC2.

27. Ishida, "The New England Collaborative Data Management Curriculum Pilot at the University of Manitoba"; Peters and Vaughn, "Initiating Data Management Instruction to Graduate Students at the University of Houston Using the New England Collaborative Data Management Curriculum"; Amanda L. Whitmire, "Implementing a Graduate-Level Data Information Literacy Curriculum at Oregon State University."

28. Douglas et al., "Meaningful Learner Information for MOOC Instructors Examined Through a Contextualized Evaluation Framework."

29. Douglas et al., "Meaningful Learner Information for MOOC Instructors Examined Through a Contextualized Evaluation Framework."

30. Sheila MacNeill, Lorna M. Campbell, and Martin Hawksey, "Analytics for Education," *Journal of Interactive Media in Education* 1 (2014): 7, https://doi.org/10.5334/2014-07.

31. Koller et al., "Retention and Intention in Massive Open Online Courses"; Kerrie A. Douglas et al., "Board #32: NSF PRIME Project: Contextualized Evaluation of Advanced STEM MOOCs," ASEE Annual Conference & Exposition, Columbus, Ohio (2017), https://peer.asee.org/27830.

32. Julie Goldman and Nevada Trepanowski, "Data from Reflection and Analysis of Implementing a Free Asynchronous MOOC to Build Competence in Biomedical Research Data Management," *Open Science Framework* (2020), Dataset, https://osf.io/vncxq.

33. Goldman and Trepanowski, "Data from Reflection and Analysis of Implementing a Free Asynchronous MOOC to Build Competence in Biomedical Research Data Management."

34. Tharindu Liyanagunawardena, Shirley Williams, and Andrew Adams, "The Impact and Reach of MOOCs."

35. Koutropoulos and Hogue, "How to Succeed in a Massive Online Open Course."

36. Julie Goldman and Elaine Martin, "Biomedical Research Data Management Open Online Education: Challenges & Lessons Learned."

37. Matt Burton et al., "Shifting to Data Savvy: The Future of Data Science in Libraries," Project Report, University of Pittsburgh, Pittsburgh, PA (2018), http://d-scholarship.pitt.edu/id/eprint/33891; Lisa Federer, Sarah C. Clarke, and Maryam Zaringhalam, "Developing the Librarian Workforce for Data Science and Open Science," Center for Open Science (2020), https://doi.org/10.31219/osf.io/uycax.

38. Goldman and Trepanowski, "Data from Reflection and Analysis of Implementing a Free Asynchronous MOOC to Build Competence in Biomedical Research Data Management."