2019-07-31


Philip Espinola Coombs  
*Boston University*

Christine Malinowski  
*Massachusetts Institute of Technology*

Amy Nurnberger  
*Massachusetts Institute of Technology*

**Corresponding Author(s)**  
Philip Espinola Coombs, Boston University Alumni Medical Library, 72 East Concord, L-120SB, Boston, MA 02118; pcoombs@bu.edu

Follow this and additional works at: [https://escholarship.umassmed.edu/jeslib](https://escholarship.umassmed.edu/jeslib)  
Part of the [Information Literacy Commons](https://escholarship.umassmed.edu/jeslib) and the [Scholarly Communication Commons](https://escholarship.umassmed.edu/jeslib)

This work is licensed under a Creative Commons Attribution 4.0 License.

**Recommended Citation**  

This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in Journal of eScience Librarianship by an authorized administrator of eScholarship@UMMS. For more information, please contact Lisa.Palmer@umassmed.edu.

Keywords
data management, assessment, competencies, instruction, data literacy

Creative Commons License
This work is licensed under a Creative Commons Attribution 4.0 License.

Rights and Permissions
Copyright © 2019 Coombs, Malinowski, and Nurnberger

Acknowledgments
The work presented in this manuscript did not benefit from any outside funding.
Full-Length Paper


Philip Espinola Coombs¹, Christine Malinowski², and Amy Nurnberger²

¹ Boston University, Boston, MA, USA
² Massachusetts Institute of Technology, Cambridge, MA, USA

Abstract

Objective: To evaluate library workshops on their coverage of data management topics.

Methods: We used a modified version of Sapp Nelson’s Competency Matrix for Data Management Skills, a matrix of learning goals organized by data management competency and complexity level, against which we compared our educational materials: slide decks and worksheets. We examined each of the educational materials against the 333 learning objectives in our modified version of the Matrix to determine which of the learning objectives applied.

Conclusions: We found it necessary to change certain elements of the Matrix’s structure to increase its clarity and functionality: reinterpreting the “behaviors,” shifting the organization from the three domains of Bloom’s taxonomy to increasing complexity solely within the cognitive domain, as well as creating a comprehensive identifier schema. We appreciated the Matrix for its specificity of learning objectives, its organizational structure, the comprehensive range of competencies included, and its ease of use. On the whole, the Matrix is a useful instrument for the assessment of data management programming.

Correspondence: Philip Espinola Coombs: pcoombs@bu.edu

Keywords: data management, assessment, competencies, instruction, data literacy

Rights and Permissions: Copyright Coombs, Malinowski, and Nurnberger © 2019

Disclosures: The authors report no conflict of interest.
Introduction

Workshops and instruction sessions are a very common research data management (RDM) service provided by academic libraries (Thielen et al. 2017). Methodical and thoughtful assessment is integral to the success of any library service, therefore, the development of assessment methods specific to data management services must exist alongside the development of the services themselves.

To evaluate the quality of our workshop offerings, MIT Libraries’ Data Management Services (DMS) employed a novel assessment instrument. We used a modified version of Sapp Nelson’s Competency Matrix for Data Management Skills to compare our educational material—slide decks and worksheets—against a matrix of learning goals organized by data management competency and complexity level (Sapp Nelson 2014; Sapp Nelson 2016). We sought to identify data management competencies which were well represented in the curriculum, as well as competencies which needed additional coverage.

MIT Libraries has offered data management workshops since 2009 and launched its formal data management service in 2010. The portfolio of workshop offerings have spanned from overview sessions (e.g., Data Management 101) to deeper dives on specific data management topics (e.g., File Organization, Strategies for Data Sharing & Storage) to emerging challenges in the data landscape (e.g., Managing your Research Code). With the increase in our DMS bandwidth, following the addition of two full-time staff positions, and the instructional and service opportunities presented by the opening of our GIS & Data Lab, we now have the opportunity to reflect on our current offerings and consider how we may evolve our workshop portfolio to close topical gaps and better align with the needs of our community.

Literature Review

Assessing library workshops or instruction sessions should address multiple angles. One, assessing the performance of the instructor. Two, assessing whether students absorbed the content, thereby meeting the learning goals of the session(s). Three, assessing whether the content met the students’ information needs. This last angle is often bolstered by an established standard of competencies, such as the ACRL’s Information Literacy Competency Standards for Higher Education (Association of College & Research Libraries 2000).

Many in the data management community were inspired by the ACRL’s Information Literacy Competency Standards for Higher Education (2000) to create competency lists focused on data management specifically (Sapp Nelson 2017). As data management rose to the forefront of the minds of educators and researchers over the last decade, the community needed to develop an understanding of the skills which comprised the new field of data management. The increasing professionalization of data workers across the same period provided further motivation to clarify and standardize data management skills. Over the last decade, several competency lists were created: Qin and D'Ignazio (2010), Carlson et al. (2011), Piorun et al. (2012), Prado and Marzal (2013), and Schneider (2013).

However, competency lists could only provide limited guidance for data management educators. There was a need for “more explicit measurable, observable, and transferable learning objectives,” (Sapp Nelson 2017, p. 3). Identifying these needs, Sapp Nelson built on
the Carlson et al. 2011 list to create an expanded matrix of data management competencies which included learning goals: the Competency Matrix for Data Management Skills (Sapp Nelson 2017). Sapp Nelson also identified an additional shortcoming of many of the competency lists. Excluding Schneider 2013, none of the lists provided guidance on which competencies should be taught to specific learner groups (Sapp Nelson 2017). This proved to be an important element of the Matrix.

A review of the literature found no examples of the use of Sapp Nelson’s Matrix thus far. However, the Carlson et al. 2011 list of data management competencies, on which Sapp Nelson’s Matrix was partially based, appears with moderate frequency in the literature. Pouchard and Bracke (2016) used the Carlson et al. 2011 list to develop and organize a survey on data literacy concepts for members of the Purdue University College of Agriculture. Both Theilen (2017) and Whitmire (2015) used the list to develop new credit-bearing courses on data management for graduate students at their respective institutions. Other citations divulge the existence of the list within the literature reviews, such as in Wiley (2017) and Thoegersen (2018).

The Matrix was intended to provide “a lens that gives insight to specific, actionable intelligence on how to improve data services,” (Sapp Nelson 2017, p. 9). We wished to gain such actionable intelligence on how to improve our own data services at MIT Libraries and Sapp Nelson’s Matrix appeared ripe for a first test.

Methodology

Instrument

As noted above, the assessment was conducted using a modified version of Sapp Nelson’s Competency Matrix for Data Management Skills (Sapp Nelson 2016). The original is a matrix of 12 data management competency areas, broken down into 36 competencies. There are 3 domains, which Sapp Nelson also refers to as levels: Personal, Team, and Enterprise. Each competency is represented as a learning goal in each of the 3 domains. The learning goals increase in complexity from the Personal domain through to the Enterprise domain. There are 108 unique learning goals, each of which includes three corresponding behaviors which illustrate completion of the learning goal. The behaviors are organized based on Bloom’s revised taxonomy of educational objectives into the categories of “Knowledge (Cognitive),” “Skills (Psychomotor),” and “Attitudes (Affective)” (Anderson and Krathwohl 2001; Sapp Nelson 2016).

At the start of our assessment, we found that certain modifications to the Matrix were necessary, both to make it more applicable to our institution, as well as to make it more user-friendly. This was unsurprising, as Sapp Nelson cautioned that significant editing would be needed to tailor the tool to each institution (Sapp Nelson 2016). The assessment team, which comprised the authors, met on a weekly basis, and together determined the modifications with consideration towards our institutional curricular needs, as communicated to us by members of the MIT community in prior research consultations. Our modifications, and the reasoning behind them, are outlined in the following sections.
Modification 1: New competency

We created a new competency under the competency area, “Data Management and Organization.” This new competency (DMO-4) focuses on file organization (file naming, versioning, best practices). Despite the name of the competency area, these topics were not represented, implicitly or explicitly, in any of the existing learning objectives.

Modification 2: “Learning goals” and “learning objectives”

We tweaked the terminology of the original Matrix, changing the original’s “learning goals” to what we termed “high-level learning objectives.” In the original, each learning goal had three behaviors, one for each of the general domains of Bloom’s revised taxonomy: cognitive, psychomotor, and affective. We appreciated the specificity of the behaviors, and chose to use the behaviors, rather than the learning goals as our tools for evaluation. This way, we provided ourselves additional granularity and avoided having to speak about fractional fulfillment of a learning goal, where only one or two of the corresponding behaviors was deemed to be represented in a workshop. We considered Sapp Nelson’s “learning goal,” as a broad learning objective, and the behaviors as more specific learning objectives. With this in mind, we chose to refer to learning goals as “high-level learning objectives,” and behaviors as “learning objectives.”

Modification 3: Reconceived the behaviors

In the original Matrix, the behaviors were organized by whether they applied to the cognitive, psychomotor, or affective domains of Bloom’s taxonomy. Upon examination, many did not appear to fit clearly into those categories and instead seemed organized according to levels of increasing complexity within Bloom’s cognitive domain.

Figure 1: Bloom’s taxonomy of the cognitive domain (Vanderbilt University Center for Teaching, 2016).
With this in mind, we reconceived the behaviors as learning objectives of increasing complexity corresponding to levels of Bloom's cognitive domain.

- “Knowledge (Cognitive)” became “Basic,” corresponding to 1: Remember and 2: Understand
- “Skills (Psychomotor)” became “Working,” corresponding to 3: Apply and 4: Analyze
- “Attitudes (Affective)” became “Advanced,” corresponding to 5: Evaluate and 6: Create

**Modification 4: Unique identifiers**

We created a unique identifier for each learning objective. This allows for easier identification and for future additions to the Matrix, should they be needed. The original Matrix had a basic numbering scheme for the learning objectives, but this was not broken up by competency area. Under that scheme, one could not add competencies without throwing off the numbering for every other competency.

**Table 1:** Key for the new identifier scheme.

<table>
<thead>
<tr>
<th>Example: DSF-P-1-B</th>
<th>Identifier</th>
</tr>
</thead>
<tbody>
<tr>
<td>DSF</td>
<td>Three letter code corresponding to the competency area (Databases, Data Sets, and Data Formats)</td>
</tr>
<tr>
<td>P</td>
<td>Letter corresponding to the domain (Personal)</td>
</tr>
<tr>
<td>1</td>
<td>Number corresponding to the high-level objective</td>
</tr>
<tr>
<td>B</td>
<td>Letter corresponding to the level (Basic)</td>
</tr>
</tbody>
</table>

**The Modified Matrix**

Our modified version became a matrix of 12 data management competency areas, broken down into 37 competencies, which include 333 unique learning objectives, as we termed Sapp Nelson’s “behaviors.” The learning objectives within each competency are arranged in order of increasing complexity according to Bloom's revised taxonomy of educational objectives (Anderson and Krathwohl 2001). The modified Matrix is openly available in Harvard Dataverse at [https://doi.org/10.7910/DVN/RIMKTJ](https://doi.org/10.7910/DVN/RIMKTJ) (Coombs 2019).

**Application**

We limited the scope of the assessment to nine workshops which are regularly offered by DMS, as this presented an accurate representation of the subjects in which we provide instruction. The assessment did not include workshops taught only once, such as by request of
a department. The educational material examined consisted primarily of slide desks, with transcripts and worksheets included when available.

We examined each of the workshop presentation materials against the 333 learning objectives in the assessment matrix to determine which of the learning objectives applied. This included examining the content on each slide or page to determine which competency area was relevant, then analyzing the content to determine the domain (personal, team, or enterprise), and level (basic, working, or accomplished) at which learners were expected to gain knowledge. We made basic counts of the results to determine which competencies were covered by these workshops and to what extent.

**Results**

**Summary**

The following workshops were assessed. Descriptions of each workshop are included in the Appendix.

- Data Management: 101
- File Organization
- Version Control
- Strategies for Data Sharing & Storage
- Using Metadata to Find, Interpret, & Share Your Data
- Data Management Plans & the DMP Tool
- Data Management for Postdocs and Research Scientists
- Quick and Dirty Data Management
- Managing Your Research Code

The best represented competency areas refers to the competency areas which have the highest percentage of learning objectives represented in at least one workshop. The best represented competency areas are: Data Preservation, Data Conversion and Interoperability; Data Curation and Re-Use; and Data Management and Organization. The competency areas which are least represented in the current education offerings are: Data Visualization; Ethics; Discovery and Acquisition of Data; and Databases, Data Sets, and Data Formats. The full table is included below (Table 2).

Data Management for Postdocs and Research Scientists had the highest number of learning objectives represented (135), followed by Quick and Dirty Data Management (58), and Managing Your Research Code (52). Version Control had the least number of learning objectives represented (1). The table with the full dataset appears in Appendix 1.

**Competency Area Coverage**

The following table shows the percentage of learning objectives within each competency area.
which are represented in at least one workshop. This illustrates the competency areas which are particular strengths and weakness of the current curriculum of DMS.

**Table 2**: Shows the percentage of learning objectives covered at least once in any workshop, for each competency area.

<table>
<thead>
<tr>
<th>Competency Area</th>
<th>Percentage of Learning Objectives Covered</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Preservation</td>
<td>93%</td>
</tr>
<tr>
<td>Data Conversion and Interoperability</td>
<td>85%</td>
</tr>
<tr>
<td>Data Curation and Re-Use</td>
<td>85%</td>
</tr>
<tr>
<td>Data Management and Organization</td>
<td>61%</td>
</tr>
<tr>
<td>Cultures of Practice</td>
<td>50%</td>
</tr>
<tr>
<td>Data Analysis</td>
<td>50%</td>
</tr>
<tr>
<td>Metadata</td>
<td>44%</td>
</tr>
<tr>
<td>Quality Assurance</td>
<td>39%</td>
</tr>
<tr>
<td>Databases, Data Sets, and Data Formats</td>
<td>33%</td>
</tr>
<tr>
<td>Discovery and Acquisition of Data</td>
<td>33%</td>
</tr>
<tr>
<td>Ethics, including citation of data</td>
<td>31%</td>
</tr>
<tr>
<td>Data Visualization</td>
<td>0%</td>
</tr>
</tbody>
</table>

**Further Details**

The following subjects are covered in DMS workshops, but not represented in the Matrix:

- Checksums / fixity checks
- ReadMe files
- How to structure data within files
- The difference between transitional storage and long-term preservation options
- Using repositories for preservation
- Institutional vs. disciplinary repositories
- Data journals
Matrix objectives not represented in current educational offerings:

- DSF-1: Boolean searching, using databases, constructing databases
- MET-4: ontologies
- DAV: data visualization
- ETH-4: locating, understanding, and discussing ethical requirements and policies, at the institutional, state, national, and international levels

Discussion

Instrument

Overall, we found the Matrix to be a useful instrument for assessment. However, modifications were necessary to increase the clarity and functionality of the Matrix.

Sapp Nelson had cautioned that it would be necessary to refine the subject matter contained in the Matrix to better represent the unique needs of individual institutions (Sapp Nelson 2017). We found this to be true; before beginning the assessment, we noted the omission of file organization. After deciding that this omission was significant, given our instruction portfolio, and included enough material to warrant a new competency, we constructed DMO-4 to rectify the omission. As we conducted the assessment, we noted other topics not represented in the Matrix, mentioned in section 4.3. Some of these topics may be implicitly included in existing competencies; for example, the topic of institutional vs. disciplinary repositories may be implicitly included in competency DSF-4, which focuses on understanding and using data repositories. Other topics, such as ReadMe files, were not included implicitly or explicitly in existing competencies. Further clarification of the competencies would address these ambiguities and minor omissions.

It is worth noting that the modifications we made to the Matrix went beyond a subject matter enhancement: we changed terminology, reconceived the “behaviors,” and created unique identifiers for each learning objective. We believe that these structural modifications increased the clarity and functionality of the Matrix.

Reconceiving the “behaviors,” was the most significant structural change. In the original Matrix, under each learning goal, were three behaviors, each of which was intended to correspond to each of the three domains of Bloom’s taxonomy: cognitive, affective, and psychomotor. We noticed that the behaviors primarily involved thought processes (cognitive domain), rather than the emotional component of learning (affective domain), or physical skills (psychomotor).

Take, for example, the behaviors included in the Personal Domain’s 14th (P14) “learning goal.” The behavior for Knowledge (Cognitive), “List standard file formats,” maps to the “remember,” category of Bloom’s cognitive domain. The behavior for Skills (Psychomotor), “Choose relevant files formats from a list of standard file formats,” maps to the “apply” category. The behavior for Attitudes (Affective) maps to the “evaluate” category.
As another example, consider the behaviors in the Team Domain’s 2nd (T2) “learning goal.” The behavior for Knowledge (Cognitive), “Discuss advantages and disadvantages of different formats depending upon different data management situation,” maps to the “understand” category of Bloom’s cognitive domain. The behavior for Skills (Psychomotor), “Generate data in a pre-selected format,” maps to the “apply,” category. The behavior for Attitudes (Affective), “Justify the data format selection by the research design,” maps to the “evaluate,” category.

Learning in research data management, like most disciplines, involves multiple domains; however, the majority of the behaviors, as written, were most appropriate for the cognitive domain. Considering this, we chose to reconceive the behaviors instead as learning objectives all within the cognitive domain.

We also changed the terminology. We chose to refer to the original Matrix’s “learning goals,” as “high-level learning objectives,” and its “behaviors,” as “learning objectives.” These changes in terminology were not necessary for the successful application of the Matrix. The changes reflected how we intended to apply the Matrix: by using the “behaviors,” as more granular learning objectives. In this regard, the changes in terminology were useful.

Lastly, our new identifier schema made it easier to identify and reference each aspect of the Matrix and allowed new competencies to be inserted without disrupting the entire numbering system. Our other changes had shifted the emphasis away from the original’s “learning objectives.” The original Matrix’s identifier schema had only identified the “learning objectives,” which understandably reflected their significance. Typically, one would not be referring to a particular behavior alone; a behavior is significant only in reference to the learning objective for which it is being used to evaluate completion. In our case though, it was important to be able to identify each of the “behaviors.” This need provided the initial impetus to create the new identifier schema.

**Application**

We found that the Matrix is best used to assess data management competencies broadly, and at a high level. Considering the comprehensiveness of competencies included, it is understandable that the Matrix lacks granularity on certain subtopics. For example, our Version...
Control workshop had only one learning objective represented. This was not due to a weakness of the topic-focused workshop; rather, the Matrix had few objectives on the topic. Many of DMS’s regular workshops are similar to Version Control, in that they provide a deep dive into one topic. As such, these workshops had a low number of learning objectives represented in the Matrix.

Results

There were several competencies or skills which were covered by DMS’s regular workshops, but not included in the Matrix, such as writing readme-style metadata (see Results, Further Details for the full list). Additionally, there are several competencies included in the Matrix which are not represented in the regular rotation of workshops, such as the entire competency area of Data Visualization. However, in the case of Data Visualization, another area of the Libraries offered a workshop on the subject.

Many learning objectives in the Matrix focus on discipline-specific knowledge or tasks, while DMS workshops are generally meant for a discipline-agnostic audience. This means that DMS workshops do not often contain material pointing to discipline-specific data management topics or guidelines. This focus on data management for a discipline-agnostic audience resulted in low coverage of certain competency areas, which focused on discipline-specific knowledge or tasks, such as Databases, Data Sets, and Data Formats; and Data Analysis.

Limitations

The results of our assessment must be digested with certain limitations in mind. Most significantly, the assessment is limited to documented aspects of workshops: slides, worksheets, and handouts. The slides sometimes included notes, but did not include full transcripts or learning objectives. Inclusion of learning objectives would have increased the ease and accuracy of the assessment process. The documented aspects may not be fully representative of the topics or concepts included in a workshop. Recordings would have provided a more holistic picture of the workshops’ subject matter coverage.

Additionally, the lack of coverage of a particular competency does not necessarily mean that the competency is missing from the larger institution’s curriculum. Division of tasks among Library departments or larger academic units may result in low coverage of certain competencies when an assessment is conducted on a single department or academic unit. For example, low coverage in the competency area of Discovery and Acquisition of Data reflects divisions in tasks among library staff at MIT. Discovery and acquisition of data is outside the scope of DMS; considering this, we were not surprised to find that this competency was poorly represented in DMS workshops. Additionally, our assessment found the competency area of Data Visualization completely missing from DMS’s regular rotation of workshops. Still, a workshop of the subject is available through another academic unit of MIT. This underscores the importance of doing an environmental scan of one’s institution. This way, one can identify other department or units on campus which may already be providing guidance on data management topics.
Conclusion

The assessment was successful in enabling us to better understand the current topical coverage in DMS workshops. In response to this assessment, we developed two recommendations for the improvement of DMS workshops: 1) weaving ethics into existing workshops, and 2) including learning objectives in workshop materials. DMS is now engaged in a project to revise our curriculum to address these weaknesses, specifically working to incorporate data ethics throughout our individual workshops.

The assessment was also successful in serving as a trial run for Sapp Nelson’s Matrix. Building on the strong foundation of Sapp Nelson’s Matrix, our structural modifications—most importantly, reconceiving the “behaviors,” and creating the new identifier schema—increased its clarity and functionality. We appreciated Sapp Nelson’s Matrix for its specificity of learning objectives, its organizational structure, the comprehensive range of competencies included, and its ease of use. We concur with Sapp Nelson in that the Matrix will need to be adapted to each institution. Our work here demonstrates a methodology for adapting the Matrix to a specific institution. We hope our methodology and the revised Matrix may prove useful to other libraries looking for an instrument for the assessment of data management programming.

Supplemental Content

Appendix 1 & 2
An online supplement to this article can be found at http://dx.doi.org/10.7191/jeslib.2019.1162 under “Additional Files”.

Data Availability

The data that support the findings of this study are openly available in Harvard Dataverse:


Acknowledgments

The work presented in this manuscript did not benefit from any outside funding.

References


Schneider, René. 2013. "Research Data Literacy." In Worldwide Commonalities and Challenges in Information Literacy Research and Practice, edited by Serap Kurbanoğlu, Esther Grassian, Diane Mizrachi, Ralph Catts, and Sonja Špiranec, 134-140. ECIL 2013 Istanbul, Turkey. Springer, Cham https://doi.org/10.1007/978-3-319-03919-0_16


