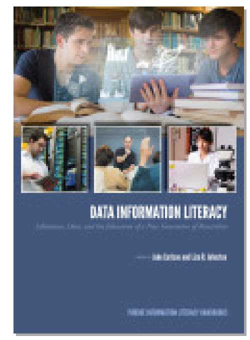
Data Information Literacy

Carlson, Jake , Johnston, Lisa



Published by Purdue University Press

Carlson, Jake and Lisa Johnston.
Data Information Literacy: Librarians, Data and the Education of a New Generation of Researchers .

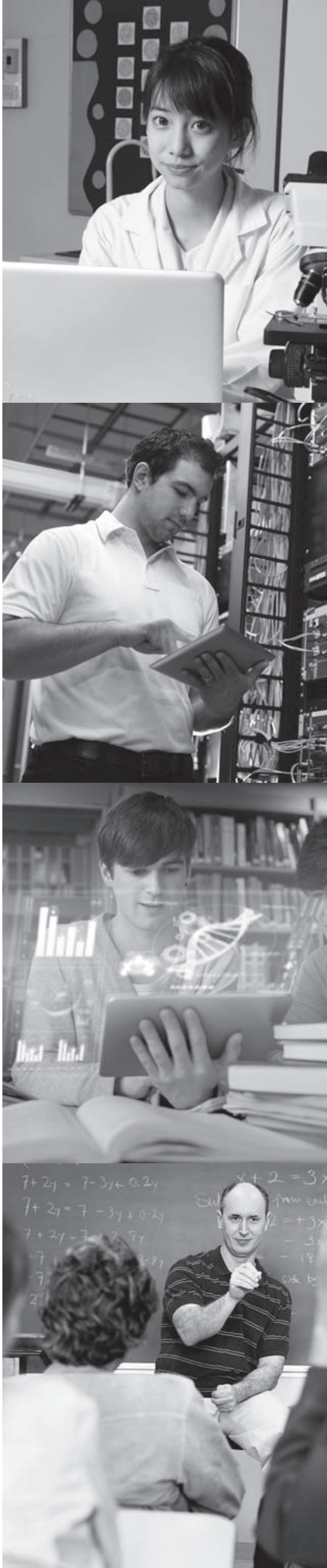
Purdue University Press, 2014.
Project MUSE.muse.jhu.edu/book/42542.

CHAPTER 10

WHERE DO WE GO FROM HERE?

Further Developing the the Data Information Literacy Competencies

Megan Sapp Nelson, Purdue University



INTRODUCTION

Chapter 1 provided a description of the DIL competencies as they were initially conceived (Carlson, Fosmire, Miller, & Sapp Nelson, 2011). Chapter 3 discussed how the competencies were modified and used as a means of gathering information for the DIL project. A primary objective of the DIL project was to create instructional interventions based on these competencies and to explore data-related educational needs within the lab environment. Faculty partners informed this process through in-depth interviews and by responding to the instruction proposed.

As we were conducting the DIL project we recognized a need for continued development of the DIL competencies. Through the interviews, faculty responded to the competencies in light of their own experiences of data management. For each specific competency faculty interviewees were asked, "Are there any skills that are not listed in this competency that you think should be included?" The responses provided guidance for how the DIL competencies might be enhanced, altered, or removed altogether in future versions.

This chapter explores the faculty-proposed changes to the DIL competencies, which are listed here in an order that follows an approximate relationship to the data life cycle.

- Discovery and acquisition of data
- Databases and data formats
- Data conversion and interoperability
- Data processing and analysis
- Data visualization and representation
- Data management and organization
- Data quality and documentation
- Metadata and data description
- Cultures of practice
- Ethics and attribution
- Data curation and reuse
- Data preservation

Following the suggested changes and a discussion on their implications, this chapter will describe future research areas that would enhance understanding of disciplinary practices and curriculum design for these competencies.

DISCOVERY AND ACQUISITION OF DATA

Skills in this competency include the following:

- Locates and utilizes disciplinary data repositories
- Evaluates the quality of the data available from external sources
- Not only identifies appropriate external data sources, but also imports data and converts it when necessary so it can be used locally

Students need critical thinking skills and techniques to retrieve data from a source external to the research laboratory or classroom. Generally, interviewees agreed with the content of the skills list presented in the interview, with a few exceptions. One faculty member focused on using critical thinking to evaluate the contents of an externally produced data set for quality. The faculty member did not describe the actual metrics by which an individual evaluates data quality. However, the need for metrics was implied.

> I think also, the skill to evaluate the quality of data. It's very easy for anyone to publish data

> online and very often when we get it, it's not very useful. So we need to look at this and make a decision and say, "Okay. It's helpful for us," or "It's not useful."
>
> —Electrical and computer engineering faculty member

Another interviewee agreed that an "appropriate level of skepticism of outside data sources" was important. He explained:

> Know your source; know your quality, particularly when we're working with remote sensing GIS data sets. Just understand that they're inaccurate, there's no way around it.
>
> —Agricultural and biological engineering faculty member

This revealed the need for analytical thinking around quality for a specific type of data: GIS (geographic information system). And it raised the question of whether different data types require different metrics of quality and whether they already exist within disciplines. If so, knowledge of the existence of disciplinary measures of data set quality may be an appropriate addition to the *discovery and acquisition of data* competency.

Faculty also raised a concern about negotiating access to externally acquired data sets.

> To evaluate a hypothesis, you need to find data. Data sets like this are not going to be available, so really what you need to be able to do is to understand how to create this data and then to figure out who has the ability to create this data and then who has the authority to allow you access to the data. This is the kind of thing that I would be involved in. My students would figure out how to create the data, and then I would be figuring out who has the ability to collect it and who has the authority to give us the data. And then talking with . . . [the data producers] about how to generate this data, whether they're open to it, and how to generate this data in a way that doesn't impact their business and doesn't expose the privacy of anything that they care about.
>
> —Electrical and computer engineering faculty member

This faculty member identified that needed data may not be publicly accessible. To intuit who might create data, to make inquiries into the existence of the data, and to negotiate access to the data is a complex access process. Extensive knowledge of the literature, the discipline, and institutional structures to identify those who may be collecting data and an introduction to basic usage agreement terms may be appropriate additions to the *discovery and acquisition of data* competency.

DATABASES AND DATA FORMATS

Skills in this competency may include the following:

- Understands the concept of relational databases and how to query those databases
- Becomes familiar with standard data formats and types for the discipline
- Understands which formats and data types are appropriate for different research questions

The critique of the *databases and data formats* competency included a related skills list for this area. The comments focused on decision making in the design of databases.

> The thing that I don't really see included here is an understanding of some of the implications of the different types of databases. . . . There are several different database products, and within those database products there are usually multiple database engines. So, for example, in MySQL you have a choice between the ISAM engine and an[. . .]other one. But one has higher data integrity and is sort of more enterprise ready but requires more memory. So the other one is the default installed engine for MySQL, because it is a lower resource usage.
>
> —Electrical and computer engineering faculty member

This reflection implies the need for nuanced understanding of the strengths and weaknesses of database products and programming languages. It appears that graduate students and faculty researchers use disciplinary expertise, technical information, and research planning and vision to create criteria by which to judge the most appropriate database products and features that will contribute to an efficient, successful research project. The "development of criteria for decision making" may be an overarching competency in DIL, related to the need for critical thinking throughout the research life cycle. Another issue was the time students have to develop these skills.

> Capabilities for statistical analysis are a little weak. And there are courses they can take on campus for the statistical and the relational databases, so maybe it's something that we should be requiring. The problem is that if they're going to do a Master's thesis, they take only seven courses. Two of them have to be outside of the department, so I guess . . . we could ask to make sure that one of those is either a database course or a statistical analysis course.
>
> —Civil engineering faculty member

"Critical thinking about the development of building a database" was also reported as a needed enhancement of the DIL competencies.

> For my discipline at least, understanding those concepts of how to build a good database would be important in addition to simply knowing how to create tables and querying. And maybe that's implied by "concept of relational databases," but to me it wasn't there.
>
> —Electrical and computer engineering faculty member

In this case, critical thinking applies to the design of the database. This faculty interviewee called for the addition of "best practices of database design" to the *databases and data formats* competency. This may result in knowledge of the most appropriate, efficient ways to program a database.

Enhanced decision making and critical thinking skills were a necessary addition when choosing appropriate file formats for a given research project.

> I would add to this the skill of understanding the advantages of different types of formats of files. This issue of knowing text files are human readable but not necessarily computer readable; XML files on the other hand are computer readable but bloated and inefficient; binary files are [at risk for having] insufficient documentation of their format, but are generally most efficient. If you're going to work with text files, you have choices. You can do delimiting between fields; you can have things in front that tell you how long

> fields are. Students must understand the tradeoffs in using files in these ways and how it makes them easier or harder to work with.
>
> —Electrical and computer engineering faculty member

Again, the necessity of picking the best tool for the job—this time for choosing file formats—is an important addition. This faculty interviewee considered critically analyzing strengths and weaknesses of available formats and understanding and predicting the consequences of the choice of file format for the long-term management of the research project to be a foundational skill. This choice represents a key decision in the research process that can have impact throughout the research life cycle. Helping students to identify those key decisions and make wise choices for their research can be addressed through the DIL competencies by including "the development of standard operating procedures or decision matrixes."

DATA CONVERSION AND INTEROPERABILITY

Skills in this competency include the following:

- Is proficient in migrating data from one format to another.
- Understands the risks and potential loss or corruption of information caused by changing data formats.
- Understands the benefits of making data available in standard formats to facilitate downstream use.

For the faculty interviewees, this was an area that was crucial but not as explicit as they would have liked. Faculty interviewees focused on the regular replacement of versions of software that leads to problems for future use of data.

> [Students need an] understanding that formats like Microsoft Word .doc files are specific and proprietary to Microsoft and that there is a need to store those in some format which you can be certain that you can open again later. The problem with .doc files is that the only reliable way to open them is to use the version of Word that created them. If that version of Word becomes outdated or runs on machines that are too old, then you never know what you're going to get.
>
> —Electrical and computer engineering faculty member

The concepts of "format obsolescence" and "changes to software over time" need to be addressed in the DIL competency skills list.

There was concern that students do not think critically about the impact data conversion has on the contents of the data. While the competency did address this, faculty interviewees specifically identified student data as potentially problematic because students tended to made conversions without fully understanding the ramifications. One interviewee went so far as to say:

> I don't know that students are aware that they tend to have more faith in their data than I do.
>
> —Agricultural and biological engineering faculty member

The revelation that data conversion may call into question the quality of data ties to the need for students to "think critically throughout the data management process" and "recognize the implications of their decisions."

DATA PROCESSING AND ANALYSIS

Skills in this competency include the following:

- Is familiar with the basic data processing and analysis tools and techniques of the discipline or research area
- Understands the effect that these tools may have on the data
- Uses appropriate workflow management tools to automate repetitive analysis of data

The DIL competencies as originally proposed did not mention *programming* explicitly. In the eyes of faculty interviewees, this was an area that needed to be included.

> [I would add] . . . these quantitative tools that I mentioned about using programming languages and knowing how to automate. Scripts in R, for instance. We're doing a lot of that in my lab now.
>
> —Ecology/landscape architecture faculty member

For the faculty interviewees, the success of a student relies on efficient and proficient use of scripting and programming to process data. One of the faculty members we interviewed highlighted one student who was an excellent data manager because of excellent programming abilities, which allowed him to process data quickly and efficiently through the use of scripting.

A major aspect of being an excellent programmer is the ability to learn new languages and techniques for the processing of data and then implement those techniques in an appropriate way.

> So the other piece is learning how to learn new techniques, right? How to go to the literature, how to go to the web, how to pick up and teach yourself new tools.
>
> —Ecology/landscape architecture faculty member

This "lifelong learning ability" (to understand and add tools to a personal research repertoire) facilitates the graduate student's ability to manage research data. The need for lifelong learning skills is imperative for all disciplines in scientific research, but it is rarely explicit. The need for making this long-term acquisition of skills apparent and built into the research experience and courses emerged consistently across the interviews.

> Certainly most . . . basic use of the tools [is learned] in a statistics class or a methodology class or something like that. But to me what happens is that . . . [students] tend to learn fairly basic application in those classes and then the transference of learning those tools to applying them toward a specific research project, critically, are very different skills. And they get that mostly in one-on-one mentorship. . . . I mean any faculty member working with a graduate student on their thesis. To me, that's the mentorship.
>
> —Ecology/landscape architecture faculty member

Some faculty expressed concern that analysis tools changed the data. A faculty interviewee was adamant that students understand that raw data should be kept in an unaltered state.

> I mean, we don't change the data. Once the data is there, I don't want them changing the data. . . . This is very important.
>
> —Natural resources faculty member

The researchers alter and analyze data sets, but the raw data should be preserved. "Keeping the raw data" should be added to this competency to underpin conversion and interoperability.

DATA VISUALIZATION AND REPRESENTATION

Skills in this competency include the following:

- Proficiently uses basic visualization tools of discipline
- Avoids misleading or ambiguous representations when presenting data in tables, charts, diagrams, and so forth
- Chooses the appropriate type of visualization, such as maps, graphs, animations, or videos, based on an understanding of the reason/purpose for visualizing or displaying data

Data visualization and representation received the most feedback by far. Faculty agreed that this was a fundamental competency for which the vast majority of graduate students needed to develop advanced skill sets.

> I'd say it's essential because it's communication. If we don't communicate, we haven't done much in the long run.
>
> —Ecology/landscape architecture faculty member

The suggestions fell into a broad spectrum of interests and concerns. A frequent refrain was the need to identify data that tell a story.

> "Avoids misleading or ambiguous representations when presenting data," I'd also put in there saying what data not to show.
>
> —Electrical and computer engineering faculty member

> I would say the thing that I don't see here that's most important is being able to evaluate, "Does this graph show what I expected it to show?"
>
> —Electrical and computer engineering faculty member

For a graduate student to gain this competency, he or she must have a clear understanding of what to communicate, and then, what the visualization is actually communicating. Critical analysis identifies which data fields heighten understanding or increase explication of the findings. This necessitates a higher level of understanding about the content of the data set and the research project as a whole.

Faculty interviewees took that need for critical thinking further, to address the use of data visualizations to make conclusions that are valid.

> I think some of these ideas are introduced in courses and probably they see it in practice. I think it is something—at a basic level—used within the discipline. But I don't think that the part of understanding how these [visualizations] can be used to support the decision making process [is present]. And that may be a skill—you know—of connecting it to that. So if they understand the reason or purpose for visualizing, then [they can] utilize . . . [visualizations] in support of making decisions.
>
> —Electrical and computer engineering faculty member

The need for informed decision making based upon visualization and representation again ties to the need for critical thinking across the data life cycle. In this case, critical thinking extends to asking questions of the data for the purpose of understanding it and making informed decisions.

Finally, the concept of students "learning multiple representation tools and choosing the most appropriate tool for the story they wish to tell" was a necessary addition to the DIL competency.

> I wish my graduate students had much more background in a broader array of representational tools[,] . . . [like] the students I teach in landscape architecture. We teach them those skills explicitly in our coursework; we spend huge amounts of time. We teach them representation. We have courses where they learn that it's not just a media skill, but . . . [they] are representing information and data, and how to do that compellingly. The science students get very little of that.
>
> —Ecology/landscape architecture faculty member

The idea that successful future scientists need new media skills is an area in which there is a need for additional research before proposing revisions to the competencies. The DIL project has just scratched the surface of the needs in visualization.

DATA MANAGEMENT AND ORGANIZATION

Skills in this competency include the following:

- Understands the life cycle of data, develops data management plans, and keeps track of the relation of subsets or processed data to the original data sets
- Creates standard operating procedures for data management and documentation

This competency was generally supported by the faculty interviewees.

> I'd say part of the standard operating procedures are to have very high levels of annotation, whether it be in your programming files or in your data sets themselves about what the data is, what the units are, when it was collected, where there might be errors. I mean I keep extensive records of my own notes so that at the end, I look back and I've crossed off every single thing I've needed to do for every single line of data. I don't think most students are that thorough.
>
> —Ecology/landscape architecture faculty member

However, establishing best practices and using them presented a difficulty arising from the lack of consensus regarding what the best practices entail.

> When you say "utilizes best practices and understands the importance of frequently updating their understanding of what best practices are," in order to do that, one has to have readily available sources that tell you what they are.
>
> —Ecology/landscape architecture faculty member

Standard operating procedures vary across disciplinary practice and research methodology. Designing and consistently using a standard operating procedure requires in-depth knowledge of how the different types of equipment and techniques used within the laboratory impact the collection of data. This systems thinking may be intuitive to the faculty interviewees. However, it is unclear how graduate students design research methodologies that may be based on an entire laboratory of methodologies and equipment without documentation about those environments. This need for specifying not only what skills the graduate students need, but also at what point in their research careers they are likely to learn these skills, points to a larger issue—namely that not all data competencies may be needed or learned during the graduate research phase of a scientific researcher's career.

DATA QUALITY AND DOCUMENTATION

Skills in this competency include the following:

- Recognizes, documents, and resolves any apparent artifacts, incompletion, or corruption of data
- Utilizes metadata to facilitate an understanding of potential problems with data sets
- Documents data sufficiently to enable reproduction of research results and data by others
- Tracks data provenance and clearly delineates and denotes versions of a data set

Few faculty interviewees wanted to augment or change this area. In one case, however, the faculty member sought to clarify a type of documentation that he felt was important but was not referenced in the DIL competencies. The need for a "story" of the changes that a data set goes through was the primary concern.

> *Interviewer:* So you're saying that the amount of documentation and description is good for your purposes—you can get a sense of what they're doing—but it wouldn't be enough for someone else outside of your lab to make sense of it. What is the gap there? What would be needed for somebody else to understand?
>
> *Faculty:* I think in my lab we have a cumulative knowledge. So from the beginning, we know . . . for example, that we have a research proposal. So we know the basic idea of what we want to do. And then we do some experiments and then the next experiment, we write the difference from the previous one. Then, so if you accumulate knowledge, then you understand the difference. Then you look at something and say, "Okay, I understand where it comes from." But for somebody else, just by looking at the difference, it does not make enough sense. . . . To understand what this means. Unless you understand the history.
>
> —Exchange with electrical and computer engineering faculty member

A major concern of professors was that individuals outside their laboratory may misunderstand, misrepresent, or misuse their data, if shared. The importance of providing not simply the context for an individual data set but also the context for the data set in relation to the entire project is a nuanced change that needs to be included in the competencies.

Another nuance regarding data quality and documentation is the use of externally written documentation (such as documentation for a software programming language) when creating new data products. Particularly with the reuse of software code, faculty found that the successful use of outside documentation is a necessary skill for students. The use of outside documentation reflects the need to establish the context of the production of a new data object, regardless of the origin of those context documents.

> Maybe include outside documentation. This seems to imply that it's all about organizing the data itself, putting things in the data. But I've found oftentimes outside documentation

> is actually . . . more helpful than just looking at code.
>
> —Electrical and computer engineering faculty member

"Knowledge of tools to assist with the creation of documentation" was brought up as a necessary addition, particularly in the context of software developers. For specific software programming languages, there are software tools that collate and/or create documentation from the software as it is written.

METADATA AND DATA DESCRIPTION

Skills in this competency include the following:

- Understands the rationale for metadata and proficiently annotates and describes data so it can be understood and used by self and others
- Develops the ability to read and interpret metadata from external disciplinary sources
- Understands the structure and purpose of ontologies in facilitating better sharing of data

Metadata and data description were generally accepted as necessary, but not very well understood, aspects of data management by the faculty interviewees. While the competencies as written were generally held to be accurate by the faculty, one interviewee felt that an even more basic need existed.

> Almost maybe even a basic level of understanding the rationale for metadata, but even just . . . a basic understanding, basic knowledge of [the concept of] metadata. And examples.
>
> —Electrical and computer engineering faculty member

The authors of the competencies assumed that there was a need for all graduate students to understand that such a thing as metadata exists, that it provides some basic level of function to a data set, and that it can be useful during research projects. However, this assumption may have been unrealistic.

> In the GIS world, you're at the mercy of other people's data. You're the beneficiary and at their mercy. I don't know how many students go into metadata. I mean certainly when I was learning it until probably like four or five years ago, I didn't go into the metadata that much. But I now use it as the source to describe what I think about this data, what are the caveats to it.
>
> —Ecology/landscape architecture faculty member

This concept may be outside of graduate students' previous experiences with data management. A missing step might be to explain what metadata is and why it is useful.

CULTURES OF PRACTICE

Skills in this competency include the following:

- Recognizes the practices, values, and norms of his or her chosen field, discipline, or subdiscipline as they relate to managing, sharing, curating, and preserving data
- Recognizes relevant data standards of his or her field (e.g., metadata, quality, formatting) and understands how these standards are applied

There were mixed responses to this competency. No one rejected or augmented any of the skills listed. However, many of the respondents focused on the idea that cultures of practice remained unformed within their discipline. Even so, some thought that it was important.

> This is really important, and I think that it's such a changing target right now. I think it's the journal requirements and the funding requirements that are making it important and making it essential. . . . They're absolutely right to do so.
>
> —Ecology/landscape architecture faculty member

But respondents were unclear as to what comprised *cultures of practice*. The ecology/landscape architecture faculty member went on to say: "But it's something that most of us are ill-prepared for. We're just sort of like, 'Oh, okay. What do we do?' And we ourselves have had very little training in this."

On the other hand, one faculty member described this as not being critical for the students to do their work.

> It's probably . . . you know, they can do their work without understanding this. It's not essential that they have this. It's best if they do, but they don't. I convey it to them just simply through our discussions of what we're doing, why we're doing it, and so on. I guess I could be doing more, but we don't talk about all of these functions. I mean we talk about some of them, but not all of them.
>
> —Civil engineering faculty member

Given this lack of clarity with a simultaneous indication of importance, this competency needs to be investigated further. It is unclear that the definition of *cultures of practice* proposed is reflective of current scientific research practice.

> I don't even know if there are practices, values, and norms. I would love . . . guidance. So the question is, do our graduate students know these things? I mean, I'm between "I don't know" to "I guess," because we've sort of ignored it. . . . So I'm not sure which is the right way to frame that. But no, we're clueless about this. How's that?
>
> —Natural resources faculty member

ETHICS AND ATTRIBUTION

Skills in this competency include the following:

- Develops an understanding of intellectual property, privacy and confidentiality issues, and the ethos of the discipline when it comes to sharing and administering data
- Acknowledges data from external sources appropriately
- Avoids misleading or ambiguous representations when presenting data

The *ethics and attribution* competency briefly mentions intellectual property. This emerged as a problem. Given the complex nature of intellectual property in the research software field, a faculty interviewee spelled out what a graduate student who creates software should know about intellectual property.

> *Interviewer:* What do you think your graduate students should know regarding intellectual property and these sorts of issues?
>
> *Faculty:* There are two answers to that. My first answer is that he shouldn't worry about it. As a Ph.D. student, he should focus on doing the research and publishing that research and graduating. Now, clearly that answer is not complete, because it ignores all of the problems that come with ignoring intellectual property and it got us to where we are today. But that is probably what's best for him

> in the short-term. Second, what is best for the lawyers especially who will have to help us deal with it if we ever had to deal with it, is for him to understand that when he's working on something he needs to be cognizant of whose resources is he using, who is paying for his time, and who currently owns what he is doing. Right, so he should be aware of Purdue's policies on work that he's doing and who owns the work that he's doing.
>
> —Electrical and computer engineering faculty member

The faculty interviewee was ambivalent about the utility of this type of knowledge to a student while explicitly listing this as a skill that the student needed to successfully manage data. There was an acknowledgment that lack of knowledge of intellectual property issues can lead to problems of data management. However, this took time that in the professor's viewpoint needed to be spent on primary research and the development of a dissertation project. This ambivalence is representative of how faculty members felt about a number of DIL competencies. They listed many as very important or essential while simultaneously agonizing about how little time was available to teach students the competencies while meeting research deadlines.

The same faculty interviewee clarified the reference to patents in the *ethics and attribution* list.

> They should probably be taught the pros and cons of patents on hardware and software and inventions and what that means. And given the concept of what it means to invent something. They should understand something about this issue of "first to invent" versus "first to file" and therefore the importance of documenting everything that you think of. Although, did the system change? It used be first to invent, and I think it may have switched to first to file?
>
> —Electrical and computer engineering faculty member

It is clear that the proposed competency is important but that more detail would reflect a nuanced understanding of the needs of graduate students as they transition to researchers. Disciplinary researchers are sometimes unclear themselves on the terms and conditions under which it is important to file for patents. This is a clear opportunity in which libraries may contribute to the DIL of graduate students. Patent librarians and copyright librarians both have expertise to teach developing researchers in these areas.

DATA CURATION AND REUSE

Skills in this competency include the following:

- Recognizes that data may have value beyond the original purpose, to validate research, or for use by others
- Is able to distinguish which elements of a data set are likely to have future value for self and for others
- Understands that curating data is a complex, often costly endeavor that is nonetheless vital to community-driven e-research
- Recognizes that data must be prepared for its eventual curation at its creation and throughout its life cycle
- Articulates the planning and activities needed to enable data curation, both generally and within his or her local practice

- Understands how to cite data as well as how to make his or her data citable

Interviewees commented that they were satisfied with the list as it was given; however, faculty might perceive this topic to be outside of their domain. One faculty member commented: "So, what is data curation?" This needs to be explored with a broader group of disciplinary faculty members.

DATA PRESERVATION

Skills in this competency include the following:

- Recognizes the benefits and costs of data preservation
- Understands the technology, resources, and organizational components of preserving data
- Utilizes best practices in preparing data for its eventual preservation during its active life cycle
- Articulates the potential long-term value of his or her data for self or others and is able to determine an appropriate preservation time frame
- Understands the need to develop preservation policies and is able to identify the core elements of such policies

Interviewees rarely augmented the topic of data preservation. Faculty were less experienced with it. The few critiques elucidate possible ways of describing the competency that researchers may respond to more readily.

> The only thing I'd add to this, when you say "utilizes best practices and understands the importance of frequently updating their understanding of what best practices are . . ." well, in order to do that, one has to have readily available sources that tell you what they are.
>
> —Ecology/landscape architecture faculty member

Another faculty member focused in on the long-term, local reuse of the data by

> making sure that any data that you care about is accessible, is replicated, and is in a format that you can still read.
>
> —Electrical and computer engineering faculty member

This response covered several DIL competencies. However, it shows the crucial interconnectedness of data curation, reuse, and preservation in the mind of this faculty member. The roughly linear format in which we presented the competencies did not show their actual roles and interplays. Presenting them in a format that shows their interconnectedness may encourage researchers to perceive them differently.

FURTHER DEVELOPING THE DATA INFORMATION LITERACY COMPETENCIES

The need for critical thinking as a necessary precursor to decision making about research projects and for the design of new research projects emerged as a strong theme. Critical thinking is a fundamental trait of an information literate individual (ACRL, 2014). The heavy focus by the faculty interviewees on this higher order thinking ability implies a need that is not present consistently among graduate students. There is a need for studies on how to instill critical thinking around data management. This would provide welcome insight into a crucial facet of data management that builds a well-rounded scientist, as well as an information literate individual.

Further investigation is needed into whether critical thinking about data management is necessary in disciplines outside of science, technology, engineering, and mathematics (STEM), or if this emphasis is a manifestation of the scientific method that underpins research in the STEM disciplines. Extending DIL to include social scientists and humanities researchers would elucidate whether the need for critical thinking skills with regard to data is truly universal.

The primary addition identified for specific competencies was that of visualization skills. A variety of questions arose as fruitful areas of future study:

- Do all graduate students in the sciences need data visualization skills, or only students in selected disciplines?
- Are there visualization skills or tools that are most appropriate in specific disciplines?
- Do scientific disciplines now prize visualization to the point that credit courses in visualization are logical additions to scientific graduate curricula?
- What role can visualization training play in creating a successful scientist in the long term?

These are areas of investigation that could have long-term impact on professional success for scientists.

The developers of the DIL competencies explored the broad range of competencies STEM researchers need to be successful in working with data. There is a need for investigation regarding how these competencies may be strategically embedded across higher education, from undergraduate programs, into graduate school, and even into postdoctoral programs. Identifying those skills that are appropriate at all stages of the developing researcher's career would help in planning to introduce skills "just in time" and in personally meaningful ways to students.

The next step for this research is curriculum mapping: to identify, within an undergraduate or a graduate curriculum, the courses that are logical places to introduce basic concepts of data management into the curriculum (Harden, 2001). The DIL competencies presume a basic understanding of data management concepts. However, the current curricula in most disciplines do not introduce these basic concepts systematically and progressively. Using educational techniques such as scaffolding, which would incorporate elements of data management bit by bit (Dennen, 2004), from the beginning of undergraduate curricula could help graduate students have stronger preparation for their research responsibilities.

The limitation of the research was the narrow scope of interviews and disciplines asked to respond to the competencies. Intended to be an initial foray into faculty reactions to the competencies and a few ways to teach them to graduate students, the project included only a small number of STEM disciplines. A necessary next step will be to get feedback from many more faculty on their perception of the relevance, utility, and accuracy of the competency list.

REFERENCES

Association of College and Research Libraries (ACRL). (2014). Information literacy competency standards for higher education. Retrieved from http://www.ala.org/acrl/standards/informationliteracycompetency

Dennen, V. P. (2004). Cognitive apprenticeship in educational practice: Research on scaffolding, modeling, mentoring, and coaching as instructional strategies. *Handbook of Research on Educational Communications and Technology,* 813–828. Retrieved from http://www.aect.org/edtech/ed1/31.pdf

Carlson, J., Fosmire, M., Miller, C. C., & Sapp Nelson, M. (2011). Determining data information literacy needs: A study of students and research faculty. *portal: Libraries & the Academy, 11*(2), 629–657. http://dx.doi.org/10.1353/pla.2011.0022

Harden, R. M. (2001). AMEE guide no. 21: Curriculum mapping: a tool for transparent and authentic teaching and learning. *Medical Teacher, 23*(2), 123–137.