5-17-2016

Pilot Data Information Literacy Competencies Matrix Scaffolded Across Undergraduate, Graduate and Data Steward Levels

Megan R. Sapp Nelson
Purdue University, mrsapp@illinois.edu

Data Information Literacy Competencies Scaffolding
2016/05/16

Level/Domain		Undergraduate Education			Scaffold	Graduate Education			Scaffold	Data Steward (Post Doctorate; Research Supervisor)			
Competency Area		Knowledge (Cognitive)	Skills (Psychomotor)	Attitudes (Affective)		Knowledge (Cognitive)	Skills (Psychomotor)	Attitudes (Affective)		Knowledge (Cognitive)	Skills (Psychomotor)	Attitudes (Affective)	Competency Area
Databases and Data Formats	U1	*Learners leverage Boolean logic and other available tools such as faceted searching to access data sets.*			G1	*Learners learn and use Access and SQL to create tables, queries and forms.* [1]			S1	*Learners understand relational databases and use that understanding to set specifications for relational databases related to the research being performed and communicate the implications of those specifications to relevant database designers and users.*			Databases and Data Formats
		Use Boolean logic to create search strings and faceted searching to refine search strings.	Explain the sequence of steps in refining a search string.	Discuss merits of altering search strategy for accessing data sets in different databases.		Create new relational databases using Access and SQL command line.	Construct databases using tables and forms.	Demonstrate relationships in database through constructed queries.		Design written specifications for relational databases used in research projects.	Compose specifications based upon understanding of relational database design and needs assessment for research project.	Verify research team members understanding of the implications of relational database features and functionality for the research endeavor.	
	U2	*Learners identify standard data formats for their academic disciplines and select appropriate data formats for a given task.*			G2	*Learners identify which data formats are appropriate for their specific research question and design their data collection/generation accordingly.*			S2	*Learners recommend data formats to students they supervise as best practices for the research group that they are in, taking into consideration disciplinary norms and practices.*			
		Identify standard data formats.	Save files in an appropriate format.	Critically evaluate formats given criteria of a specific task.		Identify appropriate formats for a given situation.	Generate data in a pre-selected format.	Justify the data format selection by the research design.		Select appropriate formats for the local research group and educate others.	Compose guidance for research group members regarding data selection.	Verify adherence of research group members to proscribed data formats.	
	U3	*Learners identify proprietary data formats relevant to their academic disciplines, articulate potential problems with the use of proprietary data formats, and select alternative formats when appropriate.*			G3	*Learners contrast proprietary and non proprietary formats for specific criteria and select the most accessible format that meets the criteria for their research data generation.*			S3	*Learners critique software packages based upon the formats the software outputs for criteria based upon factors influencing accessibility, sharing and preservation.*			
		Identify proprietary data formats.	Save files in different formats as needed.	Appreciate that all data formats have different strengths and weaknesses.		Contrast proprietary and non-proprietary formats on given criteria.	Adapt format selection to a research project given objective criteria such as performance or cost.	Compare formats given subjective criteria including usability and credibility.		Judge the impact of the available output formats on accessibility within the discipline, deposit to repositories or forward migration.	Make educated decisions on software choice while considering impact on future data management implications.	Revise format choices as technological and disciplinary criteria alter over time.	
Discovery and Acquisition of Data													Discovery and Acquisition of Data
	U4	*Learners comprehend that data sets are collected into groups by topic or by commonality such as place of origin, that the technology housing these groups are called repositories, and that they can be used to access data sets relevant to a research topic.*			G4	*Learners use tools use as RE3DATA.org and disciplinary finding aids to identify and become familiar with the most frequently used repositories in their subject specialization.*			S4	*Learners recommend specific data repositories as likely locations of relevant data to a specific research question.*			
		Explain that data sets are gathered by topic or other commonality into repositories	Proceed to access data repositories based upon understanding of contents of prevalent topics or commonalities of data held in repository.	Justify selection of repository based upon research topic.		Use re3data.org to identify most prevalent repositories for a specific discipline.	Explain relevance of identified repositories to work being done in research group or on specific project.	Relate specific data sets found to existing research questions or emerging research projects.		Recommend specific repositories based upon subject knowledge, research experience, and prior data use.	Change recommendations based upon specific needs of research projects, taking into considerations the specifications and constraints of the required data.	Question the appropriateness of a repository and its offerings to a specific research question, with objective criteria guiding the critique.	
	U5	*Learners consider authority of data set, quality of data set as indicated by existing metadata and documentation, and other critical factors as determined by research question and goals when selecting data for reuse.*			G5	*Learners critically examine downloaded data sets to consider the quality of metadata, potential problems with missing or inaccurate data, and problems of data incompletion and adjust research plans accordingly.*			S5	*Learners design protocols that requires research group members to critique downloaded data sets by specific criteria important to the research goals of the group and teach the critique protocol to team members.*			
		Use checklists or other tools to guide critiques of data sets.[2]	Choose a database critically rather than accepting the first available database on a given topic.	Justify selection of database based upon research topic.		Employ statistical, scientific, and/or other relevant literacies to evaluate contents of data set and accompanying documents.	Design and/or organize new data set in usable structure given a specific research plan and the contents of the pre-existing data set.	Explain the potential problems that could arise for new research because of data management decisions that were enacted in the pre-existing research data set.		Recommend and record specific known factors that could positively or negatively impact the usefulness of a specific data set for the research group.	Alter recommendations as new information comes to light and new knowledge is accumulated or as research trajectories change within the research group.	Verify the accuracy of the external data critique protocol on an ongoing basis.	
	U6	*Learners download tabular (.csv, .tsv, .tab) data sets and critically consider the contents of the table with a goal of recombining relevant data into a new data table.* [3]			G6	*Learners import data on relevant topics or of relevant types from repositories and convert file formats when necessary for appropriate usage given the specifications of the local research project.*			S6	*Learners transform data structures of existing data sets based upon the design constraints necessary to successfully investigate a research question or complete their research design (e.g. from a spreadsheet to a relational database).* [4]			
		Demonstrate functional knowledge of tabular file formats and apply that knowledge to download tables relevant to a topic.	Follow instructions for integrating data from multiple tables into a new table.	Justify selection of data for inclusion in reconstituted data table.		Demonstrate functional knowledge of relevant data types in their discipline to download files.	Follow specifications of research and converts file formats in accordance with those specifications.	Justify selection of data based upon the research topic or data types specified in the research design.		Develop a plan for the transformation of the data that maps fields in one data set to the other. [4]	Adapt database structure according to the available design constraints. [4]	Critique the transformations made to ensure that the content of the data hasn't been altered as the data structure has been changed.	
Data Management and Organization													
	U7	*Learners recognize the scholarly communication life cycle and that data is produced during the earliest phases of the scholarly communication lifecycle.*			G7	*Learners recognize that data has a lifecycle of its own, and that specific tasks align with different stages of the lifecycle.*			S7	*Learners use the data lifecycle to guide research data management planning and decision making by recognizing stages as gateways for designated data management tasks.*			
		State the stages of the scholarly communication lifecycle as well as listing scholarly products of each lifestyle stage.	Identify where the scholarly products of each stage of the lifecycle might be retrieved.	Discuss how scholarly products build upon each other throughout the scholarly communication lifecycle.		Apply the data lifecycle to determine the data management stage for pre-existing research projects.	Follow data protocols and explains how different protocols are tied to given stages of the research lifecycle.	Propose tasks to be completed as learner recognizes that data management moves from one lifecycle stage to another.		Design data protocols that reflect the data lifecycle and indicate tasks that should be done at given stages of the data lifecycle.	Reorganize ineffective data management practices to more efficiently leverage the data lifecycle.	Influence member of the research group to consider the data lifecycle as a project management tool for ongoing research data management.	
	U8	*Learners explain the concept of a data management plan, its constituent parts and its use as an organizational document of research endeavors.*			G8	*Learners proactively seek out data management plans as sources of information about research groups and projects and evaluate their own data practices by the expectations spelled out within data management plans.*			S8	*Learners develop data management plans for their research teams that serve as guiding project management documents, are revised as changes are made to research practice and that adequately balance responsibility to funding agencies and local research practice.*			
		List parts of a data management plan and what those parts generally contain.	Describe how a data management plan is related to a research project, and why it is created.	Discuss how a data management plan may be used in a research project.		Apply existing data management plans to their own work on a research project.	Organize research data according to those management practices specified in a data management plan.	Initiate engagement with data management plans that are specific to research groups or projects as the learner moves into that environment.		Develop data management plan that reflects local, disciplinary and data management best practices	Update data management plan as changes are made to research practices at the local or disciplinary level.	Modify data management plan based upon responsibility to funding agency, disciplinary best practices, local best practices and emerging technologies in data management.	
	U9	*Learners use a rubric to identify the credibility of databases that they find online in order to verify currency, relevance, authority, accuracy, and purpose.* [2, 6]			G9	*Learners use a rubric and/or template to prepare a data set package for sharing and reuse that describes the data set in ways that allow users to determine currency, relevance, authority, accuracy, and purpose of their dataset.*			S9	*Learners use a rubric to identify areas of incompleteness or inaccuracy in a data management plan, whether their own or that of another researcher whose proposal they are reviewing for a funding agency.*			
		List the criteria of the CRAAP test.[7]	Explain the criteria by which they believe the data sets meet or do not meet each factor of the CRAAP test.[2,6]	Discuss the similarities of evaluating information between scholarly articles, internet content and data sets.		Apply critical thinking to elements of descriptive metadata needed for end users to be able to assess the contents of a data set.	Follow a rubric or template for the preparation of metadata.	Appreciate the value of metadata for the evaluation of the contents of data set.		Judge completeness and accuracy in light of DMP best practices.	Recommend or make revisions based upon knowledge of DMP best practices.	Discriminates between differing approaches to data management as described in DMPs, identifying those that are preferred or best practices.	
Quality Assurance													
	U10	*Learners state that data errors can be prevented by stating measurement units, data fields to be captured and formats for data captured prior to starting data capture.* [8]			G10	*Learners apply best practices for data collection such as double entry or designing data storage to atomize records in order to prevent errors in data from arising.* [9]			S10	*Learners perform quality control checks on data sets and teach others to do so as well in order to ascertain that their data is accurate and free from errors.* [8]			
		State that data errors exist.	Recognize proactive steps that can be taken to prevent data errors.	Appreciate that actions taken early in the data capture process can positively impact data quality throughout the data lifecycle.		Create research plan that integrates best practices throughout data collection to minimize errors.	Organize data collection to minimize error generation.	Intitiate critical discussions of data collection procedures from the lens of error generation with the intent of identifying better practices for future projects.		Critique data sets based upon acknowledged best practices of quality control appropriate to their discipline.	Develops a systematic training program for others to ensure that data generated within the research enterprise is subject to quality control measures.	Revise practices and training programs based upon emerging technologies and practices to further improve quality control measures.	
	U11	*Learners state that data corruption can be caused by any number of interactions with data, including human error, computer file creation or transformation problems.* [8]			G11	*Learners think critically about the contents of the data set at all phases of data collection and analysis in order to identify errors at the point of generation in a timely manner and to appropriately address the errors.* [8]			S11	*Learners guide all members of the research group in regular and structured investigations of data collected, to identify errors; points of generation or introduction for those errors; and to create plans to mitigate existing errors and prevent future error introduction.*			
		State that data corruption happens.	Display a desire to prevent data corruption.	Discuss the types of data manipulations that can lead to data corruption, including human error, file creation, or transformations.		Analyze contents of data as the set is generated for potential data errors.	Assess data using metrics (based upon scientific and data literacy) to identify errors.	Synthesize information from variety of data set, metrics, and external sources to create plan to address errors.		Recommend procedures for regular and structured investigation of errors in data sets.	Initiate discussions and instruction on identification of errors, and points of generation or introduction of errors.	Synthesize best practices and local norms into plans to mitigate existing errors and to prevent future errors.	
Data Conversion and Interoperability													Data Conversion and Interoperability
	U12	*Learners move files from one format to another as part of their workflow.*			G12	*Learners recognize that moving data from one file format to another can introduce errors in the data and check for errors.*			S12	*Learners anticipate the need to move data from one format to another while creating their research plan and build in checks for error and data loss at points of data conversion.*			
		Explain the process by which a file can be saved or forward converted to another file format.	Select relevant file formats appropriate to their current project.	Demonstrate conversion process from one format to another.		Demonstrate awareness that conversion between file formats can induce errors.	Check for errors immediately upon creation of files.	Compare pre-existing file and newly created file to determine if errors have been introduced.		Design research plans with points of conversion identified as points of weakness/points of potential data loss.	Initiate specific protocols for checking error and data loss at points of data conversion.	Verify that data has forward converted successfully with a mimimum of data loss and/or error.	
	U13	*Learners recognize and prefer to save files in standard file formats.*			G13	*Learners use file formats that are standard within their disciplinary field whenever possible and understand the cons to using non-standard file formats for disciplinary research.*			S13	*Learners specify the use of standard file formats in their research group whenever possible and provide explanation to research personnel regarding preferred file formats for specific research projects.*			
		List standard file formats.	Choose relevant files formats from a list of standard file formats.	Justify selection of standard file format over non-standard file format.		State standard file formats within their disciplinary field.	Follow disciplinary best practices for file formats.	Compare advantages and disadvantages to using non-standard file formats in disciplinary research.		Categorize preferred file formats in a group protocol document for the use of research personnel.	Revise protocol as new standards emerge within the discipline.	Influence research personnel to adopt specific file formats for research project through instruction, conversation, etc.	
	U14	*Learners are aware of the difference between proprietary and non-proprietary formats and understand why using non-proprietary formats for files that they want to access in 5-10 years is best.*			G14	*Learners select non-proprietary formats whenever possible and convert proprietary files into non-proprietary formats for long term access and sharing.*			S14	*Learners identify equipment (sensors, electronics, etc.) in the research facility that output proprietary formats in the research group protocol and recommend specific non-proprietary formats to convert those files to for longterm preservation and access.*			
		Explain the difference between proprietary and non-proprietary file formats.	Differientiate between file formats as proprietary or non-proprietary.	Justify selection of non-proprietary file formats as a result of access longevity criteria.		Use non-proprietary formats whenever appropriate for the research design and methodology.	Alter file formats through conversion at key points in the research process for the purpose of data access and sharing.	Share information about the conversions with others to facilitate data access and sharing.		List equipment that outputs in proprietary formats in a shared document that is accessible by research personnel.	Originate recommendations for specific non-proprietary formats for those outputs to be converted to.	Act to ensure that conversions happen at key points in the research process for the purpose of data access, sharing and preservation.	
Metadata													Metadata
	U15	*Learners understand that metadata describes the context in which a data set was created an answers the questions of Who, What, When, Where, and Why about that data set.* [9]			G15	*Learners describe how metadata can be used by a research group and/or discipline to communicate about research findings and to assist in reproducing research results.* [10, 11]			S15	*Learners develop a metadata protocol for their research project in which metadata standards or expected levels of metadata creation (fields, units of measures, etc.) are identified.*			
		Summarize the function of metadata in terms of journalism questions Who, What, When, Where, and Why.	Identify the presence of fields within a metadata record that answer the 5W questions.	Discuss the function of the fields and why researchers might find them valuable for re-using or sharing data.		Describe metadata requirements for the research group and discipline of which they are a member.	Follow research group and disciplinary guidelines for metadata implementation.	Discuss the use of metadata in communicating research findings and reproducing research results.		Create a document with expectations for completion of metadata to existing standards or with specified fields.	Adapt metadata to research group and specifies completion at key points in research life cycle.	Verify completion of metadata at key points in research life cycle.	
	U16	*Learners define the concept of a standard and recognize that standards govern physical objects, processes and methodologies and that metadata is a methodology governed by standards.* [12]			G16	*Learners recognize that metadata is governed by standards and apply a metadata standard to a data set in order to make their data set available to others and, if applicable, to comply with disciplinary norms.* [13]			S16	*Learners use, critique, and participate in the modification of disciplinary metadata standards to improve the efficacy of existing metadata standards or bring into existence metadata standards appropriate to the disciplinary specifications.*			
		Summarize the function of a standard as a guideline ensuring consistent action or process.	Locate standards and examines the contents of a standard.	Compare a metadata standard (governing a process) to a standard governing a physical object.		Describe a data set with standard metadata.	Follow standard while using best practices to create meaningful, human and computer readable metadata.	Initiate updates to metadata at regular intervals throughout the research life cycle.		Critique metadata standards for gaps or problems that complicate metadata production in their discipline.	Initiate participation in standard issuing bodies' revisions of metadata standards to correct gaps.	Solve lingering issues with metadata standards through standard revision processes to contribute to discipline.	
	U17	*Learners recognize basic fields that frequently appear in metadata records.*			G17	*Learners seek out data sets for re-use and critique the utility of the data set found based in part upon the quality of the metadata available.*			S17	*Learners withhold publication of data sets until an acceptable quality and level of metadata have been produced to accompany the data set so that it can be understood by self and others during future use.*			
		List fields that may commonly appear in metadata records.	Describe the contents of frequently appearing metadata fields.	Discuss where metadata appears in everyday interactions (nutrition labels, media players, etc.)		Discover data sets using scholarly articles or repository directories.	Distinguish criteria for quality metadata.	Compare found data sets' metadata to the quality criteria to determine relative utility/quality of the data.		Judge completion of project or product metadata against disciplinary or standard metadata expectations.	Adapt data release schedule to allow for appropriate metadata completion.	Modify metadata with both internal and external audiences in mind for future reuse.	
	U18	*Learners discuss the concept of ontologies applied to disciplinary applications such as physics, earth sciences, curriculum development, etc.* [14, 15, 16]			G18	*Learners recognize ontologies that apply to their disciplines, major metadata standards for their disciplines or repositories that they frequently use.*			S18	*Learners embrace disciplinary movements to use ontologies as controlled vocabulary and other data exchange technologies as available by discipline.*			
		Define ontology within the context of the information they learn in their disciplinary studies.	Respond to the use of ontologies to change or augment learning within the disciplines.	Discuss that concepts are mapped in ontologies to promotes communication between groups or systems.		Perform a search for "ontology AND *discipline* " to get basic information about the disciplinary culture.	Assemble lists of keywords for metadata from existing ontologies.	Relate ontologies to the discipline as a whole.		Apply ontologies as controlled vocabulary in descriptive metadata.	Respond to disciplinary initiatives to intigrate ontologies in data sharing and exchange technologies.	Influence research personnel to adopt ontology vocabulary as appropriate throughout the research life cycle.	
Data Curation and Re-Use													Curation and Re-use
	U19	*Learners use data from government portals or other sources to download data sets relevant to their disciplinary studies as part of learning their disciplinary culture.* [3,17]			G19	*Learners design research projects and manage research outputs including data with re-use for the purpose of reproducing research as a primary goal.* [18]			S19	*Learners develop and manage research outputs including data with re-use by other researchers for the purpose of reproducing research or creating new knowledge as a specific outcome of research grants and projects.*			
		Use online portals to download a data set.	Construct their disciplinary knowledge with data sets included in that knowledge core.	Understand that data underpins disciplinary knowledge found in scholarly articles and books.		Design research project with best practices of data management, including metadata, formats, etc.	Display awareness of the importance of reproducibility in science and works to ensure reproducibility of work.	Modify work flows to reflect best practices of reproducibility as they are developed.		Design research project for data re-use, using metadata, formats, other best practices to facilitate re-use.	Display awareness of the importance of sharing data for creation of knowledge and to reproducibility.	Modify work flow to reflect best practices of sharing data and reproducibility as they are developed.	
	U20	*Learners understand that there is a cost associated with the production and access of all information.* [19]			G20	*Learners articulate the time and resource costs of managing data in their research plans.*			S20	*Learners articulate the financial, time and resource costs of managing data in their research plans and research group management, as well as in their research grant applications.*			
		Explain the costs associated with production, hosting, and access of information, whether online or in print, data or publications.	Identify who is likely to care enough to produce specific info and pay to make it available as a way to locate it.	Appreciate the non-financial costs that can be paid to make info available (copyright, etc.)		Create a research plan that factors in time and resources of data management as well as research protocols.	Organize research plan in such a way that data management is factored in at all stages of the research life cycle.	Prioritize data management so that data curation is possible at the end of the research life cycle.		Create a research plan that factors time, money and resources of data management as well as research protocols.	Organize research plan so that data management is factored in at all stages of the research life cycle.	Prioritize data management and leaves time so that data curation is possible at the end of the research life cycle and grant award period.	
	U21	*Learners recognize data management plans are artifacts that describe research projects and practices in their disciplines and look to data management plans to learn about their disciplinary culture.* [21]			G21	*Learners follow templates to learn the parts of a data management plan and practice completing a data management plan.* [22]			S21	*Learners create data management plans tailored to their research proposals that reflect the nuances of their research design and the anticipated data management needs resulting from that research design. Learners update the data management plan as research designs change over time.* [20]			
		Recognize data management plans as a document describing research projects.	Identify research practices described in the DMP common to their discipline.	Discuss disciplinary culture as shown in DMPs.		Use templates such as DMPtool.org to learn the parts of a data management plan.	Assemble a data management plan for their project using the template.	Understand utility of data management plan for organizing and managing a research project.		Construct a data management plan specific to the research design described in the funding request.	Revise data management plan as the research design changes over the life of the project.	Modify templates for custom content as needed to effectively describe needs of the research project being planned.	
Cultures of Practice													Cultures of Practice
	U22	Learners recognize data sets as a part of domain knowledge and learn to find and use data sets along with scholarly journal articles and monographs as appropriate for their discipline. [23]			G22	*Learner recognizes disciplinary practices, values and norms for their discipline as they relate to managing, sharing, curating and preserving data.* [24]			S22	*Learner teaches and models practices, values and norms for their discipline as they relate to managing, sharing, curating and preserving data.* [24]			
		State that data sets are a part of domain knowledge similar to scholarly journal articles and monographs.[19]	Find and retrieve data sets from disciplinary or other repositories.	Discuss the relationship between data sets and other research outputs in their disciplines.		Describe disciplinary practices, values and norms for data in their discipline.	Follow disciplinary practices, values and norms for data in their discipline.	Ask for clarification and further guidance about disciplinary practices, values and norms for data.		Construct educational curricula or interventions relating practices, values and norms for their disciplinary data practices.	Originate learning opportunities for personnel regarding disciplinary data practices, etc.	Influence the behavior and practice of other research personnel to adopt disciplinary data practices, values, and norms.	
	U23	*Learners define the concept of a standard and recognize that standards govern physical objects, processes and methodologies and that metadata is a methodology governed by standards.* [12]			G23	*Learners recognize that metadata is governed by standards and apply a metadata standard to a data set in order to make their data set available to others and, if applicable, to comply with disciplinary norms.* [13]			S23	*Learners use, critique, and participate in the modification of disciplinary metadata standards to improve the efficacy of existing metadata standards or bring into existence metadata standards appropriate to the disciplinary specifications.*			
		Summarize the function of a standard as a guideline ensuring consistent action or process.	Locate standards and examines the contents of a standard.	Compare a metadata standard (governing a process) to a standard governing a physical object.		Describe a data set with standard metadata.	Follow standard while using best practices to create meaningful, human and computer readable metadata.	Initiate updates to metadata at regular intervals throughout the research life cycle.		Critique metadata standards for gaps or problems that complicate metadata production in their discipline.	Initiate participation in standard issuing bodies' revisions of metadata standards to correct gaps.	Solve lingering issues with metadata standards through standard revision processes to contribute to discipline.	

Data Information Literacy Competencies Scaffolding
2016/05/16

Data Preservation	U24	*Learners understand that there is a cost associated with the preservation of all information.* [19]			G24	*Learners articulate the time and resource costs of preserving data in their research plans.*			S24	*Learners articulate the financial, time and resource costs of preserving data in their research plans and research group management, as well as in their research grant applications.*			Data Preservation
		Explain the costs associated with preserving information, whether online or in print, data or publications.	Identify who is likely to care enough to preserve info and pay to make it available as a way to locate it.	Appreciate the non-financial costs that can be paid to preserve info (manpower, tech).		Infer costs of putting preservation practices in place from beginning and update inference across life cycle.	Follow best practices for preservation as articulated for discipline or data type.	Ask for assistance in identifying best practices in preservation.		Infer costs of putting preservation practices in place from beginning and update inference across life cycle.	Follow best practices for preservation and encourages research personnel to do likewise.	Prioritize data preservation and leaves time and resources for preservation at the end of the research grant period.	
	U25	*Learners integrate regular backups and Lots of Copies Keeps Stuff Safe (LOCKSS) to preserve personal files.* [25,26]			G25	*Learners integrate backups, loss prevention, version control and other preservation best practices as appropriate to their research project and discipline.* [27]			S25	*Learners integrate backups, loss prevention, version control, periodic migrations, media-based preservation solution, and other preservation best practices as appropriate to their research project and discipline.*			
		Schedule regular backups for personal files and follow through on performing the backups.	Initiate multiple backups in multiple locations to ensure preservation of personal files.	Appreciate the importance of backups for the mitigation of file loss through theft or accident.		Set up backups, loss prevention, version control, etc in a logical and routine manner integrated within the research workflow.	Mix preservation methods logically in order to create a comprehensive preservation plan.	Relate the function of a preservation technique and describes how the technique adds a layer of added security.		Oversee implementation of backups, loss prevention, migration, etc in a routine manner integrated in workflow.	Adopt preservation best practices and implement logically in order to address preservation comprehensively.	Instruct research personnel on preservation techniques and verify that preservation initatives are being carried out effectively.	
	U26	*Learners recognize that some information has more value than other information and will need more preservation efforts.*			G26	*Learners identify unique research data sets or data sets that can not be easily recreated as targets for preservation techniques and resources.*			S26	*Learners design research projects and write data management plans that articulate which data sets will be unique or not easily reproduced, and therefore in need of long term preservation. Those data sets will be identified early in the project life cycle and resources set aside for their preservation.*			
		Prioritize authored information on a continuum from least important to most important to preserve.	Differientiate between those files requiring long term backup and storage and those files that do not.	Justify the selection of files for long term backup and storage.		Explain which research data sets cannot be easily re-created or reproduced.	Proceed with preservation protocol for specific research data sets.	Describe importance of preservation to sharing, reproducibility, and the scientific record.		Appraise likely value of data at each phase of life cycle for reproducibility and uniqueness at project design phase.	Initiate preservation protocol for data sets identified as valuable and sets aside resources for preservation.	Instruct research personnel on the importance of preservation to scientific reproducibility and the scientific record.	
Data Analysis	U27	*Learners employ analytical methods during applied research opportunities such as undergraduate research fellowships, internships, or undergraduate research experiences.* [28]			G27	*Learners analyze data using the data processing, graphical analysis, and statistical analysis tools that are commonly used or preferred within the discipline.* [29]			S27	*Learners recommend specific data processing, graphical analysis, and statistical analysis tools to research personnel as best practices for the local research lab or discipline as a whole.* [29]			Data Analysis
		Employ analytical methods as directed by research supervisor.	Follow instructions or protocols for analyzing data using specific analytical tools prescribed by research lab.	Discuss use of analytical methods within broader discipline as whole with research supervisor.		Identify analysis tools that are commonly used w/n discipline via exposure to disciplinary research (articles, data sets.)	Design analyses using widely accepted tools. Accept advice from expert users to improve analyses.	Verify design of analyses with research supervisor and other experts in discipline to assure research quality.		Recommend specific analytical tools to research personnel.	Instructs supervisees regarding disciplinary preferences regarding analytical tools.	Modifies analytical protocols as new tools emerge, are appropriate for research and are embraced by discipline.	
	U28	*Learners employ workflow management tools during applied research opportunities such as undergraduate research fellowships, internships, or undergraduate research experiences.* [28]			G28	*Learners fully document all workflows including parameters and analysis to encourage reproducibility of research.* [29]			S28	*Learners recommend specific work flows and workflow management tools to research personnel as best practices for the local research lab or discipline as a whole.* [29]			
		Employ workflow management tools as directed by research supervisor.	Follow instructions or protocols for producing documentation and using workflow management tools.	Discuss use of workflow management tools within broader discipline as a whole with research supervisor.		Document workflows formally via workflow software or informally via workflow diagrams for reference.	Follow best practices such as documenting code and constructing end-to-end scripts.[29]	Verify that documentation accurately reflects all aspects of analysis to facilitate reproducibility and sharing.		Generate workflows, either formally through software or informally via diagrams for all projects in research project.	Initiates conversations regarding appropriate work flows and tools for specific research projects.	Verifies that workflows have been adequately documented to ensure reproducibility and future sharing, if desired.	
Data Visualization	U29	*Learners employ visualization tools during applied research opportunities such as undergraduate research fellowships, internships, or undergraduate research experiences.* [28]			G29	*Learners visualize data using tools that are appropriate to their research project and that are commonly used or preferred with their discipline.*			S29	*Learners recommend specific data visualization tools to research personnel as best practices for the local research lab or discipline as a whole.*			Data Visualization
		Employ visualization tools as directed by a research supervisor.	Follow instructions or protocols for visualizing data that are specific to research group.	Justify decision to present quantitative information visually[17]		Identify visualization tools commonly used w/n discipline via exposure to disciplinary research (articles, data, etc.)	Construct visualizations using widely accepted tools. Accepts advice from expert users to improve visualizations.	Verify visualization with research supervisor and other experts in discipline to ensure clear communication.		Recommend specific visualization tools to research personnel.	Instruct supervisees regarding disciplinary preferences regarding visualization tools.	Question visualization parameters selected by supervisees in order to clarify and enhance communication in end product.	
	U30	*Learners recognize ambiguity and misleading presentation of data in visualizations authored by others and critique how the ambiguity is engendered.* [30,31]			G30	*Learners select the most appropriate visualization type for their data and project and choose to remove ambiguity and misleading appearance from visualizations of their work.* [30,32]			S30	*Learners critique visualizations produced by research personnel, seeking to remove ambiguous or misleading visualizations in order to maximize communication through the visualization.* [30,32,33]			
		Recognize ambiguity and misleading presentation of data.[31]	Describe the techniques/choices that the visualization author made to create ambiguity.[31]	Propose alternative presentations of data that would be less ambiguous or biased.[31]		Select most appropriate visualization for their data[30]	Follow best practices of visualization design[32]	Justify the choices that they made in designing their visualization based upon the parameters of their data.[32]		Critique visualizations produced by research personnel.[31, 33]	Detect ambiguous or misleading visualization and determine the source of the problem.[31,32,33]	Clarify the problematic nature of the visualization with supervisee in order to enhance clarity and communication in publication.	
	U31	*Learners list a variety of types of visualizations, note those that are frequently used within their discipline, and recognize them by format or appearance.* [36]			G31	*Learners articulate the advantages and disadvantages for a variety of data visualization types.* [35]			S31	*Learners recommend specific data visualization types to research personnel based upon criteria specific to the data being represented.* [35,36]			
		List variety of types of visualizations.[36] State visualizations frequently used in their discipline.[32]	Detect specific visualization types by format or appearance.[36]	Understand that types of visualizations display different content and/or contextual meaning.[36]		Describe the different strengths and weaknesses of data visualization types for their own data.[31,32,35,36]	Alter selection of data visualization type based upon content or contextual meaning.[31, 35]	Discuss the strengths/ weaknesses of the data visualization created in terms of content/meaning displayed.		Recommend specific data visualization types based upon articulated criteria.[35]	Identify criteria based upon best practices of visualization design and specific research communication goals.[34,35]	Help supervisees to understand best practices of visualization design through instruction and mentoring.	
Ethics, including citation of data	U32	*Learner identifies data as the intellectual property of a person or entity similar to articles, books, or music.*			G32	*Learners recognize that data is intellectual property governed by local policies and government laws and regulations and that correspondingly, their behavior toward the data is sometimes influenced by larger trends at the university and in society.* [36]			S32	*Learners recognize that they are ultimately accountable for the data collected in their research group and can be held liable to university policy and state and federal laws and regulations.* [36]			Ethics, including citation of data
		Identify data as intellectual property.[19]	Describe intellectual property of data in terms of articles, books, and other familiar IP.[19]	Conform with fair use and ethical use of intellectual property, including data. [19]		Discuss the content of policies, laws and regulations applicable to their data.	Describe how those policies, laws and regulations can change how the learner interacts with their own data.	Comply with all relevant policies, laws and regulations.		Relate relevant policies, laws and regulations to all members of research group.	Initiate compliance checks for all data sets produced within the research group.	Conform to all relevant policies, laws and regulations.	
	U33	*Learners cite data as well as articles, books, or any other resource used in compiling a new scholarly work.* [19, 37]			G33	*Learners post data sets with recommended citations, including Digital Object Identifier, to encourage reuse.*			S33	*Learners track metrics for cited data sets and include those metrics within promotion and tenure documents. Learners recommend the inclusion of data sets as scholarly works worth consideration in promotion and tenure processes.*			
		Write citations for data as well as books or other resources while compiling scholarly works.	Differentiate the citation style for a citation of a book versus that of a data source.[38]	Label the parts of the citation for data sets as well as book or article citations.[19]		Prepare a recommended citation for all data sets shared publically.[37,38]	Register a file for a DOI via uploading the data set to a repository.[38]	Verify impact of the data set through statistics collected via the repository or an independent organization.[38]		Track metrics for cited data sets through statistics collected via repository or independent organization.[38]	Assemble portfolios, dossiers and curriculum vitae with impact statistics featured.	Influence departments and colleges to include impact metrics for data sets as one measure of impact towards promotion and tenure.	
	U34	*Learners state that data including information about individual persons are subject to greater care and scrutiny for research and data management.*				*Learners honor the confidentiality and privacy of their research participants by getting their research approved by Institutional Review Boards (as appropriate) and anonymizing data sets by removing personal identifiers (both direct and indirect) as appropriate.* [39]			S34	*Learners select their storage solutions to match the level of confidentiality and privacy required for their data sets.Learners acknowledge that not all storage solutions are appropriate for human subject data and discuss options with campus IT personnel.*			
		Identify personal data as different for management and access purposes than non-personally identifying data.	Differentiate between personally identifying and non-personally identifying data.	Discuss the need for privacy when personal data is collected using FERPA or HIPAA data as an exemplar.		Develop an IRB protocol for research involving human subject. Take required human subject research training.[41]	Follow best practices for anonymizing data sets.[41]	Conform to institutional, disciplinary, and state and federal laws and regulations governing data.		Construct a storage solution based upon confidentiality and privacy needs as well as available resources.	Adapt storage solutions as requirements for confidentiality and privacy change.	Ask for assistance to identify best possible tech solutions for private and/or confidential data sets.	
	U35	*Learners locate the ethical statements for their disciplinary bodies and investigate them for the presence of information about data management or handling.* [40]			G35	*Learners personally subscribe to the ethical statements of their disciplinary bodies, including those statements covering the care and handling of data in an ethical manner.*			S35	*Learners interpret the ethical statements of disciplinary bodies as those statements interact with the institutional policies, state and local regulations, and requests by publishers for data. Learners explain these interplays to research personnel in their research groups.*			
		Locate ethical statements for their disciplinary bodies.	Detect information on data management or handling.	Conform with disciplinary ethical guidelines for data management or handling.		Support ethical statement of disciplinary body.[40,42]	Proceed according to best practices or ethical guidelines outlined by disciplinary body.[40]	Discuss understanding of obligations with research personnel and supervisors.		Compare ethical statements with policies, laws and regulations.	Note if there are any conflicts or tensions between the expectations of the documents.	Influence others in research group to carefully consider all ethical requirements while in pursuit of research.[39]	

Reference Number	Author	Publication Title	Publication Source	Permanent Identifier/URL	Publication Volume/Issue	Publication Date
1	Campbell, Robert J.	Database Design: What HIM Professionals Need to Know	Perspectives in Health Information Management	http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2047327/	1:06	August 2004
2	Zilinski, Lisa, Sapp Nelson, Megan, and Amy Van Epps	Developing Professional Skills in STEM Students: Data Information Literacy	Issues in Science and Technology Librarianship	10.5062/F42V2D2Z	77	Summer 2014
3	Stephenson, Elizabeth Caravello, Patti Schifter	Incorporating data literacy into undergraduate information literacy programs in the social sciences: A pilot project	Reference Services Review	http://dx.doi.org/10.1108/00907320710838354	35:4	2007
4	Sadamih Project, Oxford University Computing Services	Information practices for researchers - good practice guides: Databases and tools for structured data	Oxford University Computing Services	10.5281/zenodo.28331		July 26, 2013
5	Rolando, Lizzy Carlson, Jake Hswe, Patricia Wells Parham, Susan Westra, Brian, Whitmire, Amanda L.	Data Management Plan as research tool	ASIST Bulletin	DOI: 10.1002/bult.2015.1720410510	41:5	June/July 2015
6	Rinto, Erin	Developing and Applying an Information Literacy Rubric to Student Annotated Bibliographies	Evidence Based Library and Information Practice	10.18438/b8559f	8:3	2013
7	Blakeslee, Sarah	The CRAAP test	LOEX Quarterly	http://commons.emich.edu/loexquarterly/vol31/iss3/4	31:3	2004
8	DataONE	Lesson 5: Data Quality Control and Assurance		https://www.dataone.org/education-modules		2012
9	DataONE	Lesson 7: Metadata		https://www.dataone.org/education-modules		2012
10	Kolker et al.	Toward More Transparent and Reproducible Omics Studies Through a Common Metadata Checklist and Data Publications	OMICS: a Journal of Integrative Biology	10.1089/pmi.2013.0149	18:01	2014
11	Edwards, Paul N. et al.	Science Friction: Data, metadata, and collaboration	Social Studies of Science	**10.1177/0306312711413314**	41:5	2011
12	Erdmann, Charlotte	Standards for New Educators: Guide to ABET Outcomes and Standards Availability in Libraries	ASEE 2010 Annual Conference & Exposition	https://peer.asee.org/16224		2010
13	DataONE	Lesson 8: How to Write Quality Metadata		https://www.dataone.org/education-modules		2012
14	Chou, Guo-Li Anderson, O. Roger	A Study of Undergraduate Physics Students' Understanding of Heat Conduction Based on Mental Model Theory and Ontology-Process Analysis	Science Education	DOI: 10.1002/sce.20385	94:5	2009
15	Libarkin, Julie C. Kurdziel, Josepha P.	Ontology and the Teaching of Earth System Science	Journal of Geoscience Education		54:3	2006
16	Chen, Chih-Ming	Ontology-based concept map for planning a personalized learning path	British Journal of Educational Technology	DOI: 10.1111/j.1467-8535.2008.00892.x	40:6	2009
17	Calzada Prado, Javier Marzal, Miguel Angel	Incorporating Data Literacy into Information Literacy Programs: Core Competencies and Contents	Libri	DOI: 10.1515/libri-2013-0010	63:2	2013

18	Altman, Micah	State of the Art Informatics for Research Reproducibility, Reliability, and Reuse: Or How I Learned to Stop Worrying and Love Data Management	Micah Altman's Blog	https://drmaltman.wordpress.com/2014/03/27/state-of-the-art-informatics-for-research-reproducibility-reliability-and-reuse-or-how-i-learned-to-stop-worrying-and-love-data-management/		2014	
19	American Library Association	Information Literacy Competency Standards for Higher Education		http://www.ala.org/acrl/sites/ala.org.acrl/files/content/standards/standards.pdf		2000	
20	Maglia, Anne	DMPs and Public Access: Agency and Data Service Experiences				2016	
21	Brandt, D. Scott	Data Management for Undergraduate Researchers: Data Management Plans	Purdue University LibGuides	http://guides.lib.purdue.edu/c.php?g=353013&p=2378291		2016	
22	DataONE	Lesson 3: Data Management Planning		https://www.dataone.org/education-modules		2012	
23	Maybee, Clarence Carlson, Jake Slebodnik, Maribeth Chapman, Bert	"It's in the Syllabus":Identifying Information Literacy and Data Information Literacy Opportunities Using a Grounded Theory Approach	Journal of Academic Librarianshi	doi:10.1016/j.acalib.2015.05.009	41:4	2015	
24	Carlson, Jacob Fosmire, Michael Miller, C.C. Sapp Nelson, Megan	Determining Data Information Literacy Needs: A Study of Students and Research Faculty	portal: Libraries and the Academ	https://muse.jhu.edu/	11:2	2011	
25	Stanford University	Preservation Principles		http://www.lockss.org/about/principles/		2016	
26	Valentino, Maura Boock, Michael	Data Management for Graduate Students: A Case Study at Oregon State University	Practical Academic Librarianship	https://ir.library.oregonstate.edu/xmlui/bitstream/handle/1957/57831/ValentinoMauraLibraryDataManagementGraduate.pdf?sequence=4	5:02	2015	
27	Carlson, Jake Bracke, Marianne S.	Agriculture/Graduate Students/Carlson & Bracke/Purdue University/2014	Data Information Literacy Case Studies	http://docs.lib.purdue.edu/dilcs/vol1/iss3/1/	1:03	2014	
28	Kreisberg, Adam Frank, Rebecca D. Faniel, Ixchel D. Yakel, Elizabeth	The Role of Data Reuse in the Apprenticeship Process	ASIST 2013	http://www.asis.org/asist2013/proceedings/submissions/papers/49paper.pdf	50	2013	
29	DataONE	Lesson 10: Analysis and Workflows		https://www.dataone.org/education-modules		2012	
30	Tufte, Edward	The Visual Display of Quantitative Information	Cheshire, CT: Graphics Press			2001	
31	Womack, Ryan	Data Vizualization and Information Literacy	IASSIST Quarterly	http://www.iassistdata.org/sites/default/files/iq/iqvol381_womack.pdf	38:1	2014	
32	Kelleher, Christa Wagener, Thorsten	Ten Guidelines for Effective Data Visualization in Scientific Publications	Environmental Modelling & Software	doi:10.1016/j.envsoft.2010.12.006	26:6	2011	
33	Evergreen, Stephanie Emery, Anne K.	Data Visualization Checklist		http://stephanieevergreen.com/wp-content/uploads/2014/05/DataVizChecklist_May2014.pdf		2014	
34	Tufte, Edward	Envisioning Information	Cheshire, CT: Graphics Press			1990	
35	Lengler, Ralph Eppler, Martin J.	Periodic Table of Visualization Methods	Visual-literacy.org	http://www.visual-literacy.org/periodic_table/periodic_table.html		2007	
36	Briney, Kristin Goben, Abigail Zilinski, Lisa	Do You Have An Institutional Data Policy? A Review of the Current Landscape of Library Data Services and Institutional Data Policies	Journal of Librarianship and Scholarly Communication	10.7710/2162-3309.1232	3:2	2015	

37	Martone, M. (ed.)	Data Citation Synthesis Group: Joint Declaration of Data Citation Principles	San Diego, CA: FORCE11	https://www.force11.org/group/joint-declaration-data-citation-principles-final		2014	
38	DataCite	Cite Your Data	datacite.org	https://www.datacite.org/services/cite-your-data.html			
39	Coates, Heather	Ensuring Research Integrity: The role of data management in current crises	College & Research Libraries News	http://crln.acrl.org/content/75/11/598.full	75:11	2014	
40	Resources for Research Ethics Education	Data Management	UCSD	http://research-ethics.net/topics/data-management/#discussion		2016	
41	Purwin, Tom et al.	The Forum Guide to Education Statistics	National Forum on Education Statistics	http://nces.ed.gov/pubs2010/2010801.pdf		2010	
42	Whitbeck, Elizabeth	The Responsible Collection, Retention, Sharing, and Interpretation of Data	Online Ethics Center	http://www.onlineethics.org/Resources/TeachingTools/Modules/19237/moddata.aspx		2006	