Data Information Literacy Toolkit

Interview Worksheet
Faculty

Name of Interviewee:

_____

Department or Center:

_____

Institution:

_____

This worksheet is designed to elicit the information necessary for a librarian to develop a data information literacy program that is based on norms of your research field and on your local lab practices.

This worksheet is meant to be filled out as a part of an interview and your responses to the questions in each module will serve to guide the conversation. The interviewer will then ask you some follow up questions about your responses to gather additional details and to better understand your priorities and needs.

If you have any questions, need more information, or would like clarification about any item listed in this worksheet, please do not hesitate to ask your interviewer.

Module 1 – Data Sets

Please provide a brief description of the data sets that are generated in your lab:

Data Set #1: _____

    Research question or purpose of the data:

    _____

    _____

    Funding sources:

    _____

    Name(s) of the Graduate Student(s) working on this data set:

    _____

    _____

Data Set #2: _____

    Research question or purpose of the data:

    _____

    _____

    Funding sources:

    _____

    Name(s) of the Graduate Student(s) working on this data set:

    _____

    _____

Data Set #3: _____

    Research question or purpose of the data:

    _____

    _____

    Funding sources:

    _____

    Name of the Graduate Student(s) working on this data set:

    _____

    _____

Data Set #4: _____

    Research question or purpose of the data:

    _____

    _____

    Funding sources:

    _____

    Name of the Graduate Student(s) working on this data set:

    _____

    _____

If needed, please use the back of this Worksheet to list additional data sets.

Module 2 – The Lifecycle of the Data Set

For each data set identified in the previous section please complete the following tasks:

1. Please draw a diagram that depicts the data lifecycle of each of the data sets listed in the previous section from beginning to end.

   In your diagram, consider activities such as data collection, processing, analysis and publication, as well as any other activities that significantly add to or change the data set.

2. In each of the diagrams you develop, please identify where and how you interact with the graduate students who are working with the data.

[this page is left blank for additional diagrams of data lifecycles if needed]

Module 3 – Learning about Data

As you know, there are many critical facets to handling, managing, sharing and working with data. From your perspective, what are the most important skills for your graduate students to learn in these areas?

_____

_____

_____

_____

Module 4 - Acquiring External Data

Please indicate how important you believe it is for your students to be knowledgeable in each of the competencies listed below by the time they graduate by circling a response below:

Discovery and Acquisition of Data
Skills may include:
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.

1	2	3	4	5
Not Important	Somewhat Important	Important	Very Important	Essential

I don't know or NA

Additional Questions:

1. Are data sets from outside of your lab used in your research, or the research of your graduate students?

   Yes   No   I'm not sure

2. What are the major data repositories or sources of data in your discipline (if any)?

   _____

   _____

Module 5 – Formats

Please indicate how important you believe it is for your students to be knowledgeable in each of the competencies listed below by the time they graduate by circling a response below:

Databases and Data Formats
Skills may include:
Understands the concept of relational databases, how to query those databases, and becomes familiar with standard data formats and types for their discipline. Understands which formats and data types are appropriate for different research questions.

1	2	3	4	5
Not Important	Somewhat Important	Important	Very Important	Essential

I don't know or NA

Data Conversion and Interoperability
Skills may include:
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.

1	2	3	4	5
Not Important	Somewhat Important	Important	Very Important	Essential

I don't know or NA

Module 6 – Tools

Please indicate how important you believe it is for your students to be knowledgeable in each of the competencies listed below by the time they graduate by circling a response below:

Data Processing and Analysis
Skills may include:
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.

1	**2**	**3**	**4**	**5**
Not Important	Somewhat Important	Important	Very Important	Essential

I don't know or NA

Data Visualization and/or Representation
Skills may include:
Proficiently uses basic visualization tools of discipline. Avoids misleading or ambiguous representations when presenting data in tables, charts, diagrams, etc. Chooses the appropriate type of visualization, such as maps, graphs, animations, or videos, based on their understanding of the reason / purpose for visualizing or displaying data.

1	**2**	**3**	**4**	**5**
Not Important	Somewhat Important	Important	Very Important	Essential

I don't know or NA

Module 7 – Organization and Description of Data

Please indicate how important you believe it is for your students to be knowledgeable in each of the competencies listed below by the time they graduate by circling a response below:

Data Management and Organization
Skills may include:
Understands the lifecycle 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.

1	**2**	**3**	**4**	**5**
Not Important	Somewhat Important	Important	Very Important	Essential

I don't know or NA

Data Quality and Documentation
Skills may include:
Recognizes, documents, and resolves any apparent artifacts, incompletion, or corruption of data sets. Utilizes metadata to facilitate an understanding of potential problems with data sets. Documents data sufficiently enough to enable the reproduction of the research results and the data by others. Tracks data provenance and clearly delineates and denotes versions of a data set.

1	**2**	**3**	**4**	**5**
Not Important	Somewhat Important	Important	Very Important	Essential

I don't know or NA

Additional Question:

1. Do you have a system for version control in place for your data?

    Yes   No   I don't know

Metadata and Data Description
Skills may include:
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.

1	2	3	4	5
Not Important	Somewhat Important	Important	Very Important	Essential

I don't know or NA

Additional Questions:

1. Is the amount of documentation and description that your graduate students currently provide sufficient for you to be able to understand and make use of the data?

   Yes   No   I don't know

2. Is the amount of documentation and description that your graduate students currently provide sufficient for another person outside of your lab to be able to understand and make use of the data?

   Yes   No   I don't know

Module 8 – Cultural Practice and Ethical Behavior

Please indicate how important you believe it is for your students to be knowledgeable in each of the competencies listed below by the time they graduate by circling a response below:

Cultures of Practice
Skills may include:
Recognizes the practices, values, and norms of his/her chosen field, discipline, or sub-discipline as they relate to managing, sharing, curating, and preserving data. Recognizes relevant data standards of his/her field (metadata, quality, formatting, etc.) and understands how these standards are applied.

1	**2**	**3**	**4**	**5**
Not Important	Somewhat Important	Important	Very Important	Essential

I don't know or NA

Ethics and Attribution
Skills may include:
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.

1	**2**	**3**	**4**	**5**
Not Important	Somewhat Important	Important	Very Important	Essential

I don't know or NA

Additional Questions:

1. Do your students know how to cite a data set according to disciplinary standards?

    Yes   No   I don't know

2. Who owns the data sets that are created in your lab?

    _____

Module 9 – Curation and Preservation

Please indicate how important you believe it is for your students to be knowledgeable in each of the competencies listed below by the time they graduate by circling a response below:

Data Curation and Re-use
Skills may include:
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 lifecycle. Articulates the planning and activities needed to enable data curation, both generally and within his/her local practice. Understands how to cite data as well as how to make his/her data citable.

1	**2**	**3**	**4**	**5**
Not Important	Somewhat Important	Important	Very Important	Essential

I don't know or NA

Additional Question:

1. Have you or your graduate students ever deposited data into a data repository?

     Yes   No   I don't know

Data Preservation
Skills may include:
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 lifecycle. Articulates the potential long term value of his/her data for him/herself or others and is able to determine an appropriate preservation timeframe. Understands the need to develop preservation policies and is able to identify the core elements of such policies.

1	**2**	**3**	**4**	**5**
Not Important	Somewhat Important	Important	Very Important	Essential

I don't know or NA

Data Information Literacy Toolkit

Interviewer's Manual
Faculty

Notes on the Interviewer's Manual

The Interviewer's Manual provides the framework for the interview. It contains text and questions to be read to the participating researcher over the course of the interview.

It is meant to be used in conjunction with the Interview Worksheet. The Interview Worksheet should be given to the interviewee to fill out over the course of the interview. Some of the questions you will ask will be in response to the answers given by the researcher in the Interview Worksheet.

To aid in the readability of this document during the interview the font has been enlarged.

The instructions to the interviewer are colored in red and are in italics.

The explanatory text that is meant to be read to the interviewee is in "quotes".

Begin the Interview

It is suggested that the interviewer read the following text to the interviewee as an introduction:

"This interview is designed to elicit information about the data sets that graduate students in your lab are working with and the knowledge and skills needed by graduate students to manage, share and curate data.

During this interview, I will be asking you to complete the questions in the Interview Worksheet that I have given you, and then I will ask you several follow up questions about your responses to gather additional details and to better understand your data and your priorities and needs in educating your students.

This worksheet is divided into individual modules. Please fill out a module when I ask you to do so and then stop after you have completed the section. I will then ask you some additional questions before we proceed to the next module in the interview.

If you have any questions, need more information, or would like clarification about any item listed in this worksheet, please do not hesitate to ask me.

Do you have any questions for me at this time?"

Module 1 – The Data Set

Read the following text to the interviewee:

"I would now like to have you fill in module 1 of the worksheet: please provide me with a brief description of the data sets that are generated in your lab by answering the questions on the worksheet."

> Data Set is defined here as: The data collected and analyzed for a specific project or problem. It may consist of more than one data type (Excel Spreadsheets and .jpg images for example).

Ask the researcher to provide a name or title for the data set on the first line. This is just to help you distinguish the data sets from each other during the interview.

For the "Research question or purpose of the data" - ask the researcher to provide a brief statement describing the reason behind the development or use of the data set. This will help give an idea of the nature of the data set and may lead to additional questions during the interview. However, do not worry about getting the details of the data set (format, etc.) at this stage in the interview.

For the funding source(s) – there may be more than one – ask the researcher to name the specific program if possible.

Once the researcher has filled in information about all of the data sets, ask the following questions:

- Do any of your funding sources require you to…
    - …Draft a data management plan?
      If "yes", then ask:
        - What is covered in this plan?

    - …Share your data with others in some fashion?
      If "yes", then ask:
        - "How do you plan on sharing this data?

    - …Preserve your data beyond the life of the funding period?
      If "yes", then ask:
        - "How do you plan on preserving this data?

For the graduate students - The graduate students that are currently working on the data sets are all that are needed. Don't worry about getting the names of past graduate students.

Once the researcher has filled in information about all of the data sets, ask the following question:

- Did the graduate student(s) work on this data set from the beginning or did any of them inherit data from a previous student?

If the answer is "Inherit", then ask

- How is/was the transfer of responsibility for this data set accomplished?

If the researcher lists more than one graduate student on a data set, then ask:

- How do these graduate students work together on this data set?

If the researcher has more than 4 data sets then use the back of the Worksheet to capture the same information about them.

Module 2 – The Lifecycle of a Data Set

It is suggested that the interviewer read the following text to the interviewee as an introduction to this module:

"I would now like you to walk me through the lifecycle of each of the data sets you described from start to finish.

Please draw a diagram of the lifecycle and highlight the interaction between yourself and your graduate students with the data. Things to include in your diagrams might be how and where they data are collected, how the data are transformed as they pass through the lifecycle, the processes or tools that are used to describe, transform, manipulate and/or analyze the data, and any other actions that may affect the nature of the data."

Potential follow-up questions:

- Who is involved with the data at (each of the) stages?

    o What role(s) do you play with the data at this point?

    o What role(s) do your graduate students play?

- o Who is responsible for administering the data at this stage, and what are their responsibilities?

  (If the people involved and the responsibilities for the data shift between stages, ask how the data are transferred from one person to the next.)

- Why are the data being handled in this way? What purpose does this step have?

- What tools and instruments are being used at this stage?

- How are the data stored and made accessible to those who need them at this stage?

- What happens to the data after the research activity has been completed? (and/or, the papers/reports have been published?)

Keep in mind that you will be getting more detail about the individual data sets from the graduate students when you interview them. Therefore, keep the discussion at a high to mid-level to ensure you have enough time to cover other modules.

Module 3 – Learning About Data

Have the researcher answer the question on the worksheet. Then ask the following questions:

- Why do you think these skills are important?

- How are your graduate students learning these skills currently?

    You may need to emphasize that there is no right answer and that this question is not meant to be judgmental.

    If needed, ask questions about the approaches and methods of instruction used:
    - formally / informally
    - direct hands-on / indirect in the class room

- Are the skills your graduate students are learning needed for disciplinary practice or are they more specific to your particular lab?

Module 4 – Acquiring External Data

Once the researcher has circled a number in the "Data Discovery and Acquisition" competency, ask the following questions:

- How would you describe your graduate students' skills and abilities, on average, in discovering and acquiring data from external sources?

- How do your graduate students develop skills in the discovery and acquisition of data? Where are they learning these skills?

- Are there any skills that are not listed in this competency that you think should be included?

Additional Questions:

If the researcher answers "yes" to Question #1, then ask:

- Have you or your graduate students encountered any challenges in using this data?

- How did you or your graduate students evaluate the quality of this data?

Module 5 – Formats

Once the researcher has circled a number in the "Databases and Data Formats" competency, ask the following questions:

- How would you describe your graduate students' skills and abilities, on average, in formatting data or developing databases?

- How do your graduate students develop skills in formatting data or developing databases? Where are they learning these skills?

- Are there any skills that are not listed in the "Databases and Data Formats" competency that you think should be included?

Once the researcher has circled a number in the "Data Conversion and Interoperability" competency, ask the following questions:

- How would you describe your graduate students' skills and abilities, on average, in converting data or interoperating data with another data set?

- How do your graduate students develop skills in converting data or interoperating data with another data set? Where are they learning these skills?

- Are there any skills that are not listed in the "Data Conversion and Interoperability" competency that you think should be included?

Module 6 – Tools

Once the researcher has circled a number in the "Data Analysis" competency, ask the following questions:

- How would you describe your graduate students' skills and abilities, on average, in using data analysis tools?

- How do your graduate students develop skills in using data analysis tools? Where are they learning these skills?

- Are there any skills that are not listed in the "Data Analysis" competency that you think should be included?

Once the researcher has circled a number in the "Data Visualization and/or Representation" competency, ask the following questions:

- How would you describe your graduate students' skills and abilities, on average, in visualizing or representing data?

- How do your graduate students develop skills in visualizing or representing data? Where are they learning these skills?

- Are there any skills that are not listed in the "Data Visualization and/or Representation" competency that you think should be included?

Module 7 – Organization and Description of Data

Once the researcher has circled a number in the "Data Management and Organization" competency, ask the following questions:

- How would you describe your graduate students' skills and abilities, on average, in managing and organizing data?

- How do your graduate students develop skills in managing and organizing data? Where are they learning these skills?

- Are there any skills that are not listed in the "Data Management and Organization" competency that you think should be included?

Once the researcher has circled a number in the "Data Quality and Documentation" competency, ask the following questions:

- How would you describe your graduate students' skills and abilities, on average, in documenting and assuring the quality of their data?

- How do your graduate students develop skills in documenting and assuring the quality of their data? Where are they learning these skills?

- Are there any skills that are not listed in the "Data Quality and Documentation" competency that you think should be included?

  Provenance is defined here as: documentation of who has custody of the data set and any changes made to the data set (an approximation of the OAIS definition).

Additional Questions:

Version control is defined here as: tracking and controlling changes to a data file, a data set and the associated documentation.

If the researcher answers "yes" to Question #1, then ask:

- Could you please describe the system for version control that you have in place?

If the researcher answers "no" or "I don't know" to Question #1, then ask:

- Is the version control of data an issue for you or your graduate students?

Once the researcher has circled a number in the "Metadata and Data Description" competency, ask the following questions:

- How would you describe your graduate students' skills and abilities, on average, in developing metadata or other descriptions for their data?

- How do your graduate students develop skills in developing metadata or other descriptions for their data? Where are they learning these skills?

- Are there any skills that are not listed in the "Metadata and Data Description" competency that you think should be included?

Additional Questions:

If the researcher answers "no" or "I don't know" to Question #1, then ask:

- What additional organization or description would your graduate students have to provide for you to be able to understand and make use of their data?

If the researcher answers "no" or "I don't know" to Question #2, then ask:

- What additional organization or description would your graduate students have to provide for another person outside of your lab to be able to understand and make use of their data?

Module 8 – Cultural Practice and Ethical Behavior

Once the researcher has circled a number in the "Cultures of Practice" competency, ask the following questions:

- How would you describe your graduate students' knowledge of disciplinary practices and norms in working with and managing data?

- Where do your graduate students acquire a knowledge of disciplinary practices and norms in working with and managing data?

- Are there any elements that are not listed in the "Cultures of Practice" competency that you think should be included?

Once the researcher has circled a number in the "Ethics and Attribution" competency, ask the following questions:

- How would you describe your graduate students' knowledge, on average, of intellectual property and ethical behavior in working with data?

- Where do your graduate students develop their knowledge of intellectual property and ethical behavior in working with data?

    o Be sure any university mandated programs in this area are covered in your discussion.

- Are there any skills that are not listed in the "Ethics and Attribution" competency that you think should be included?

Additional Questions:

As a follow up to Question #1, ask

- What are the standards for citing data, or providing attribution more generally, in your discipline?

As a follow up to Question #2, ask

- Are there any other stakeholders associated with your data sets?

    Stakeholders is defined here as: any groups, organizations, individuals or others that you see as having made an investment in the data or that you would feel the need to consult regarding the data's treatment, release, or use.

Module 9 – Curation and Preservation

Once the researcher has circled a number in the "Data Curation and Re-use" competency, ask the following questions:

Data Curation can be defined as: the activity of managing and promoting *the use of data*, *starting from the point of creation*, to ensure its fitness for contemporary purposes and availability for discovery and re-use.

- How would you describe your graduate students' skills and abilities, on average, in preparing their data for its eventual curation and re-use by others?

- How do your graduate students develop skills in preparing their data for its eventual curation and re-use by others? Where are they learning these skills?

- Are there any skills that are not listed in the "Data Curation and Re-Use" competency that you think should be included?

Additional Question:

- *If the researcher answers "yes" to Question #1, then ask:*
    - o Which repository?
    - o Why was that repository selected?

- *If the researcher answers "no" or "I don't know" to Question #1, then ask:*

    - o Do you intend to at some point? Why or why not?

Once the researcher has circled a number in the "Data Preservation" competency, ask the following questions:

- How would you describe your graduate students' skills and abilities, on average, in preparing their data for its eventual preservation to ensure long-term access?

- How do your graduate students develop skills in preparing their data for its eventual preservation? Where are they learning these skills?

- Are there any skills that are not listed in the "Data Preservation" competency that you think should be included?

Collecting Data Samples and/or Additional Sources of Information

- Ask the interviewee if you could have a sample of the data.

- Ask the interviewee if there are any relevant documents, assignments or publications that would help you better understand the data that he/she would be willing to share with you.

END OF INTERVIEW – THANK THE INTERVIEWEE FOR THEIR TIME

Data Information Literacy Toolkit

Interview Worksheet
Graduate Student

Authors	Jake Carlson, Megan Sapp Nelson, Marianne Stowell Bracke and Sarah Wright
Publisher	Purdue University Libraries / Distributed Data Curation Center
Contact	jakecar@umich.edu
Date of Creation	January 29, 2012
Date of Last Update	April 6, 2015
Version	V 1.0
Acknowledgement	This toolkit was developed with funds from the IMLS award LG-07-11-0232-11
URL	http://datainformationliteracy.org
License	Creative Commons Attribution-NonCommercial-ShareAlike 3.0 License.

Name of Interviewee: _____

Department or Center: _____

Institution: _____

This worksheet is designed to elicit the information necessary to develop a data curation profile about data from a particular research project. In your responses to this worksheet and to the interview questions, please limit your focus to the data associated with the particular research project you have selected.

This worksheet is meant to be filled out as a part of the interview and your responses to the questions in each module will serve to guide the conversation. The interviewer will then ask you some follow up questions about your responses to gather additional details and to better understand your priorities and needs.

If you have any questions, need more information, or would like clarification about any item listed in this worksheet, please do not hesitate to ask your interviewer.

Module 1 – The Data Set

Please provide a brief description of the data:

_____

_____

_____

_____

_____

_____

Module 2 – The Lifecycle of the Data Set

1. Initial Data Stage: _____

- Approximately how many data files exist at this stage?

    _____

- What is the approximate average size of each file at this stage?

    _____

    (Please include the unit of measurement – kb, MB, GB, etc.)

- What format(s) are the data in?

    _____

    (For example: MS Excel 2007, MySQL database, JPEG 2000 images, a raw data file from a Campbell CR10 data logger, etc.)

2. Second Data Stage: _____

- Approximately how many data files exist at this stage?

    _____

- What is the approximate average size of each file at this stage?

    _____

    (Please include the unit of measurement – kb, MB, GB, etc.)

- What format(s) are the data in?

    _____

    (For example: MS Excel 2007, MySQL database, JPEG 2000 images, a raw data file from a Campbell CR10 data logger, etc.)

3. Third Data Stage: _____

- Approximately how many data files exist at this stage?

  _____

- What is the approximate average size of each file at this stage?

  _____

  (Please include the unit of measurement – kb, MB, GB, etc.)

- What format(s) are the data in?

  _____

  (For example: MS Excel 2007, MySQL database, JPEG 2000 images, a raw data file from a Campbell CR10 data logger, etc.)

4. Fourth Data Stage: _____

- Approximately how many data files exist at this stage?

  _____

- What is the approximate average size of each file at this stage?

  _____

  (Please include the unit of measurement – kb, MB, GB, etc.)

- What format(s) are the data in?

  _____

  (For example: MS Excel 2007, MySQL database, JPEG 2000 images, a raw data file from a Campbell CR10 data logger, etc.)

5. Fifth Data Stage: _____

- Approximately how many data files exist at this stage?

    _____

- What is the approximate average size of each file at this stage?

    _____

    (Please include the unit of measurement – kb, MB, GB, etc.)

- What format(s) are the data in?

    _____

    (For example: MS Excel 2007, MySQL database, JPEG 2000 images, a raw data file from a Campbell CR10 data logger, etc.)

Please list any additional stages on the back of this sheet as needed

Additional Question:

Did you inherit any part of the data set that you have described from another student or researcher?

   Yes   No   I don't know

Module 3 – Data Management

1. How and where do you store your data? (If more than one storage method is used, please list all of them).

    _____

2. Are there formal policies or procedures that you follow in managing your data set?

    Yes   No   I don't know

3. Do you do any of the following as a part of managing your data set?

    a. Use a file naming system to assign names to your data files?

        Always   Sometimes   Occasionally   Never   I don't know

    b. Make back-up copies of your data?

        Always   Sometimes   Occasionally   Never   I don't know

    c. Take security measures to protect your data?

        Always   Sometimes   Occasionally   Never   I don't know

    d. Follow a system to identify and track different versions of your data files as they progress through the data lifecycle?

        Always   Sometimes   Occasionally   Never   I don't know

    e. Provide a means to identify the "official" or "authoritative" version of your data files?

        Always   Sometimes   Occasionally   Never   I don't know

Module 4 – Tools

1. What tools, including software or hardware, are used in generating your data?
   (e.g., "a data logger", "a remote sensor", etc. – please describe.)

   _____

   _____

   _____

2. What tools, including software or hardware, are required to analyze or utilize the data?
   (e.g., "Microsoft Excel 2007", "ArcGIS", "I wrote my own program", etc. please
   describe.)

   _____

   _____

   _____

3. Do you generate any visual representations of your data? (Visual representations
   include charts, graphs, plots, videos, images, etc.)

         Yes   No   I don't know

4. Over the course of your research, have you or do you anticipate having to convert any
   part of your data from one format into another? (For example, converting a Microsoft
   2007 file into a .csv file)

         Yes   No   I don't know

Module 5 - Documentation and Description of Data

1. Please explain briefly how this dataset is documented and how it has been described (e.g. "detailed annotations", "a lab notebook", "a code book", "a data dictionary", "column headings in a spreadsheet", etc.).

    If you have used any standardized forms of documentation, description or metadata please identify the standard(s) in your response.

    _____

    _____

    _____

2. Is the amount of documentation and description that you currently create sufficient for another person with similar expertise **to be able reproduce your data** set on his or her own?

    Yes   No   I don't know

3. Is the amount of documentation and description that you currently create sufficient for another person with similar expertise **to be able to understand and use the data** for their own purposes?

    Yes   No   I don't know

4. Are there standard practices in your discipline for organizing, documenting or describing data that you are aware of?

    Yes   No   I don't know

5. Are you familiar with the concept of metadata?

    Yes   No   I don't know

Module 6 – Acquiring and Using External Data

3. Have you ever made use of data that was generated by a source outside of your lab? This could include downloading data from a data repository, accessing data from a supplemental file, requesting a data set from another researcher, etc.

       Yes   No   I don't know

4. Do you know how to cite a data set?

       Yes   No   I don't know

5. What are the major data repositories or sources of data in your discipline (if any)?

_____

_____

Module 7 – Sharing Data

1. Have you ever deposited any data into a data repository?

    Yes   No   I don't know

2. Are you planning to deposit your data into a data repository in the future?

    Yes   No   I don't know

3. Would you be willing to share your data with someone outside of your lab?

    Yes   No   I don't know

    3a. If you answered "yes" to Question #3, would you place any conditions on sharing this data with someone outside of your lab?

        Yes   No   I don't know

4. In the journals of your field, are data or other supplemental information, accepted for publication?

    Yes   No   I don't know

5. Who would you imagine might be interested in the dataset that you are producing? (For example, other researchers in my field, researchers outside of my field, practicing professionals, policy makers, etc.)

    _____

6. How would you imagine this data being used by the groups / people you listed in the previous question? What value would the data have for these groups / people?

    _____

7. Will you leave a copy of your data with your advisor and/or lab when you graduate?

    Yes   No   I don't know

Module 8 – Intellectual Property & Ethics

1. Who is the owner of the data?

   _____

2. Is this dataset bound by any privacy or confidentiality concerns?

          Yes   No   I don't know

3. Do you feel that you have a good understanding of the university's policies on research data?

          Yes   No   I don't know

4. Have you received any training or instruction on intellectual property issues as they relate to data and/or data ethics generally?

          Yes   No   I don't know

Module 9 – Data Preservation

1. What are the most important parts of your data set to preserve (manage and maintain over time)?

   _____

   _____

   _____

   _____

2. How long would your data set be useful or have value for you or others if it were to be preserved?

   ___ My dataset does not need to be preserved.

   ___ Less than 3 years.

   ___ 3 years or more but less than 5 years.

   ___ 5 years or more but less than 10 years.

   ___ 10 years or more but less than 20 years.

   ___ 20 years or more but less than 50 years.

   ___ 50 years or more but less than 100 years.

   ___ Indefinitely.

   ___ I don't know

3. Are you or anyone else in your lab taking steps to preserve this data set?

   Yes   No   I don't know

Module 10 – Educational Needs

Please indicate how important you believe it is for you to be knowledgeable in each of the competencies listed below by the time you graduate by circling a response below:

Discovery and Acquisition of Data
Skills may include:
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.

1	2	3	4	5
Not Important	Somewhat Important	Important	Very Important	Essential

I don't know or NA

Databases and Data Formats
Skills may include:
Understands the concept of relational databases, how to query those databases, and becomes familiar with standard data formats and types for their discipline. Understands which formats and data types are appropriate for different research questions.

1	2	3	4	5
Not Important	Somewhat Important	Important	Very Important	Essential

I don't know or NA

Data Conversion and Interoperability
Skills may include:
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.

1	2	3	4	5
Not Important	Somewhat Important	Important	Very Important	Essential

I don't know or NA

Data Processing and Analysis

Skills may include:

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.

1	2	3	4	5
Not Important	Somewhat Important	Important	Very Important	Essential

I don't know or NA

Data Visualization and/or Representation

Skills may include:

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

1	2	3	4	5
Not Important	Somewhat Important	Important	Very Important	Essential

I don't know or NA

Data Management and Organization

Skills may include:

Understands the lifecycle 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.

1	2	3	4	5
Not Important	Somewhat Important	Important	Very Important	Essential

I don't know or NA

Data Quality and Documentation
Skills may include:
Recognizes, documents, and resolves any apparent artifacts, incompletion, or corruption of data sets. Utilizes metadata to facilitate an understanding of potential problems with data sets. Documents data sufficiently enough to enable the reproduction of the research results and the data by others. Tracks data provenance and clearly delineates and denotes versions of a data set.

1	2	3	4	5
Not Important	Somewhat Important	Important	Very Important	Essential

I don't know or NA

Metadata and Data Description
Skills may include:
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.

1	2	3	4	5
Not Important	Somewhat Important	Important	Very Important	Essential

I don't know or NA

Cultures of Practice
Skills may include:
Recognizes the practices, values, and norms of his/her chosen field, discipline, or sub-discipline as they relate to managing, sharing, curating, and preserving data. Recognizes relevant data standards of his/her field (metadata, quality, formatting, etc.) and understands how these standards are applied.

1	2	3	4	5
Not Important	Somewhat Important	Important	Very Important	Essential

I don't know or NA

Ethics and Attribution

Skills may include:

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.

1	2	3	4	5
Not Important	Somewhat Important	Important	Very Important	Essential

I don't know or NA

Data Curation and Re-use

Skills may include:

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 lifecycle. Articulates the planning and activities needed to enable data curation, both generally and within his/her local practice. Understands how to cite data as well as how to make his/her data citable.

1	2	3	4	5
Not Important	Somewhat Important	Important	Very Important	Essential

I don't know or NA

Data Preservation

Skills may include:

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 lifecycle. Articulates the potential long term value of his/her data for him/herself or others and is able to determine an appropriate preservation timeframe. Understands the need to develop preservation policies and is able to identify the core elements of such policies.

1	2	3	4	5
Not Important	Somewhat Important	Important	Very Important	Essential

I don't know or NA

Data Information Literacy Toolkit

Interviewers Manual
Graduate Student

Authors	Jake Carlson, Megan Sapp Nelson, Marianne Stowell Bracke and Sarah Wright
Publisher	Purdue University Libraries / Distributed Data Curation Center
Contact	jakecar@umich.edu
Date of Creation	January 29, 2012
Date of Last Update	April 6, 2015
Version	V 1.0
Acknowledgement	This toolkit was developed with funds from the IMLS award LG-07-11-0232-11
URL	http://datainformationliteracy.org
License	Creative Commons Attribution-NonCommercial-ShareAlike 3.0 License.

Notes on the Interviewer's Manual

The Interviewer's Manual provides the framework for the interview. It contains text and questions to be read to the participating researcher over the course of the interview.

It is meant to be used in conjunction with the Interview Worksheet. The Interview Worksheet should be given to the interviewee to fill out over the course of the interview. Some of the questions you will ask will be in response to the answers given by the researcher in the Interview Worksheet.

To aid in the readability of this document during the interview the font has been enlarged.

The instructions to the interviewer are colored in red and are in italics.

The explanatory text that is meant to be read to the interviewee is in "quotes".

Begin the Interview

It is suggested that the interviewer read the following text to the interviewee as an introduction:

"This interview is designed to elicit information about a particular data set (or data sets) that you have identified from your research, what you are currently doing to manage this data and to identify and capture your needs in sharing, curating or preserving this data. Based on the information you provide, I will construct a Data Curation Profile of this data.

During this interview, I will be asking you a series of questions to learn more about your data and the skills needed to handle and manage this particular data set. Specifically, I will be asking you to complete the questions in the Interview Worksheet that I have given you, and then I will ask you several follow up questions about your responses to gather additional details and to better understand your priorities and needs.

This worksheet is divided into individual modules. Please fill out a module when I ask you to do so and then stop after you have completed the section. I will then ask you some additional questions before we proceed to the next module in the interview.

In your responses to this worksheet and to the interview questions, please limit your focus to the particular data that you have selected as your focus.

If you have any questions, need more information, or would like clarification about any item listed in this worksheet, please do not hesitate to ask me.

Do you have any questions for me at this time?"

Background Question

Ask the following questions to the interviewee:

Before I have you begin Module 1 of the worksheet, could you please provide me with a brief overview of the research project associated with the data that we will be discussing in this interview?

Module 1 – The Data Set

Read the following text to the interviewee:

"Please provide me with a brief description of the data set"

> *Review what the Interviewee has written to be sure that you understand it. Ask for clarification and additional information as needed.*

> *Don't worry about getting all of the details at this point. More details will be gathered in the next modules.*

Module 2 – The Lifecycle of the Data Set

It is suggested that the interviewer read the following text to the interviewee as an introduction to this module:

"I would like you to walk me through the research project you just described from the point of view of the data. In particular, I would like to know what happens to the data at each stage of the research process.

For instance, if we were to re-frame the research process as a series of stages consisting of activities around the data, what would those activities be and how would you characterize them?

Some potential stages might include the generation of raw data, the processing or refinement of the data, the data analysis, and the final data product.

Starting with the first stage, what does the data "look like" at this stage, and how do you interact with it?"

Have the interviewee describe a data stage first and then complete the corresponding information listed on the worksheet. Repeat this process for each data stage.

<u>For each data stage mentioned</u> you may want to ask additional questions to gather the following information:

- What happens to the data at this stage?
- Why does this happen (what purpose does it serve)?
- How does it happen (what tools and instruments are used in particular)?
- Who is involved (lab techs, grad students, etc.)?
- What happens to the data to get it to the next stage?
- Is there any need for external or ancillary data at this stage? (i.e. weather databases, external software code, geographic information systems data not produced by sensors in the lab, etc.)

(The Interviewee may bring up these issues him/herself but some follow up questions on these points may be needed.)

If the interviewee answered "yes" to Question #6, then ask:

- What challenges or difficulties, if any, did you encounter with inheriting this data set?

Module 3 – Data Management

Have the interviewee answer the questions below right after he/she completes each corresponding question in the worksheet.

Question #1 - discuss the response as needed.

> *Be sure the response that you get includes the storage media and an identification of any data management tools used by the interviewee.*

Question #2 - If the interviewee answers "yes" then ask the following:

- Please describe these formal policies or procedures, and specify where and how they are applied.

- How did you learn about these procedures or policies, and how to apply them in your work? Were you given any formal instruction in these policies or procedures, or did you learn them on your own?

- Where are the formal procedures kept? May we have a copy of them?

Question #3 -

If the interviewee answered "yes" to any of the questions listed on the worksheet, then follow up by asking for more details on what is done by the student and why.

3a. What file naming system do you use and why? Please describe it to me.

3b. How often do you make back-up copies of your data set? Where are the back-ups stored?

3c. What security measures do you take to protect your data?

- Security measures could include:
  - o digital (such as password protection)
  - o physical (such as housing servers at an off-site location)

3d. How do you identify and track the different versions of your data files?

- Tracking could include:
    - Horizontal tracking – different versions of a particular file, or
    - Vertical tracking – moving back and forth across the lifecycle stages to identify the sources of data / data provenance.

3e. How do you identify the "official" or "authoritative" version of your data files?

Once the Interviewee has answered all of the questions on the worksheet then ask the following:

- Are there any other actions or steps that you take in managing your data that are not listed here?

- How and when did you learn to manage your data? What role did your advisor or peers play in teaching you your data management skills? (in general, not just for this specific project / data set).

Module 4 – Tools

Have the interviewee answer the questions below right after he/she completes each corresponding question in the worksheet.

After reviewing the interviewee's answers for Question #1, ask the following question (if needed):

- How is this tool used in working with the data?

After reviewing the interviewee's answers for Question #2, ask the following question:

- Would this tool be needed for others to make use of your data?

  If the answer to this question is "yes", then ask the following:

  o Are there alternative tools available that could be used instead?

*Question #3 – if the answer is "yes" or "I don't know",
then ask the following questions:*

- What is the purpose of the visual representation(s) of your data?

- How do you generate a visual representation(s) of your data? How and where did you learn to do this?

- How do you manage these visual representations and/or relate them to your data set?

*Question #4 – if the answer is "yes", then ask the
following question:*

- When did (will) you have to convert your data, and for what purpose?

- Was (will) your data affected at all by the conversion process, and if so how?
    (For instance: Did you lose any information or attributes of your data in the conversion process?)

Follow up Questions to ask:

- How did you learn to use these tools?

- Have you documented how you have used these tools in working with your data? Would someone else need to know this information in order to understand or make use of your data?

Module 5 - Documentation and Description of Data

Have the interviewee answer the questions below right after he/she completes each corresponding question in the worksheet.

Question #1 - discuss the response as needed. Be sure you understand the response before moving on.

Question #2 – If the answer is "no" or "I don't know", then ask:

- What additional documentation or description would be needed for someone to be able to reproduce your data set on his or her own?

Question #3 – If the answer is "no" or "I don't know", then ask:

- What additional documentation or description would be needed for someone to be able to make use of your data for their own purposes?

Question #4 – If the answer is "yes", then ask:

- o Could you please describe these standard practices to me? What purpose do these standard practices serve?

Question #5 – If the answer is "yes", then ask:

- What does "metadata" mean to you and what purpose it serve? (in your own words)

Once the Interviewee has answered all of the questions on the worksheet then ask the following:

- How and when did you learn to document and describe your data? What role did your advisor or peers play in teaching you your data documentation and description skills?

Module 6 – Acquisition and Use of External Data

Have the interviewee answer the questions below right after he/she completes each corresponding question in the worksheet.

Question #1 – if the answer is "yes", then ask the following questions:

- Where or from whom did you acquire this data? How did you discover it?

- How did you make use of this data?

- How did you evaluate the quality of this data? What criteria did you employ and why?

- What documentation accompanied this data?

    o Was this documentation sufficient for you to make use of the data? Why or why not?

- Did you have to convert the data or process it in any way before you were able to use it?

    o If so, what conversion or processing did you need to do?

- Did you encounter any challenges in making use of the data?

Question #1 – if the answer is "no" or "I don't know", then ask the following questions:

- Did you ever consider making use of data generated outside of your lab in your research?

- Can you imagine a situation in which you might want or need to use data that was generated from someone outside of your lab?

    o If so, please describe this situation.

- How would you evaluate the quality of a data set generated by someone outside of your lab? What criteria would you employ and why?

Question #2 - if the answer is "yes", then ask the following questions:

- Could you please describe the components of a data citation?

- Does your description come from a particular citation format or standard?

Question #2 - if the answer is "no" or "I don't know", then ask the following question:

- How important do you think it is to cite data sets in your discipline?

Question #3 – if the Interviewee DOES list at least one data repository, then ask the following questions:

- Have you used this repository (-ies) in any way? (use includes searching as well as downloading)

    o If so, what was your experience in using this repository (-ies)?

Question #3 – if the Interviewee DOES NOT list any data repositories, then ask the following question:

    o Do you believe that your research community would benefit from having a public data repository? Why or why not?

Module 7 – Sharing Data

Have the interviewee answer the questions below right after he/she completes each corresponding question in the worksheet.

Question #1 – if the answer is "yes", then ask the following question:

- What was the submission process? What steps did you take to prepare the data for submission?

- Were there any terms and conditions that you had to agree to in order to submit your data to this repository? If so, what were they?

Question #2 – if the answer is "yes", then ask the following questions:

o Which repository would you consider submitting your data to?

o When would you submit your data to this repository?

o What will you need to do to prepare your data for submission to this repository?

Question #2 – if the answer is "no" or "I don't know", then ask the following question:

- o Why not?

Question #3 – once the Interviewee has circled their response, ask the following question:

- Why did you select this response?

- How would you make the decision about whether or not to share your data? What criteria (if any) would you use?

- Are there established norms or practices of sharing data in your research community?

Question #3a – if the answer is "yes" or "I don't know", then ask the following question:

- What conditions would you place on sharing your data?

Question #4 – if the answer is "yes", then ask the following questions:

- What are the names of these journals or other venues of publication?

- Is submitting your data to the journal(s) expected or just accepted?

- How will you prepare your data for submission to these journals or other venues?

Question #5 - discuss the response as needed. Be sure you understand the response before moving on.

Question #6 - discuss the response as needed. Be sure you understand the response before moving on.

Question #7 – if the answer is "yes" or "I don't know", then ask the following question:

- Do you intend to take a copy of the data with you when you graduate?

- When and how will you deliver your data to your advisor?

- In addition to your data, what information will your advisor and/or lab need to have in order to be able to understand and make use of your data once you have graduated?

- How will you provide this information to your advisor or lab?

Question #7 – if the answer is "no", then ask the following question:

- Why not?

Module 8 – Intellectual Property & Ethics

Have the interviewee answer the questions below right after he/she completes each corresponding question in the worksheet.

After reviewing the interviewee's answers for Question #1, ask the following question:

- Have you and your advisor ever discussed ownership of the data that you are working on?
    - If so, what was discussed?

Question #2 – if the answer is "yes" or "I don't know", then ask the following question:

- Could you please describe the sensitive or confidential nature of the data set and how this affects how you manage or administer this data?

Question #3 – if the answer is "yes" or "I don't know", then ask the following question:

- What is your understanding of university policy on research data?

- Have you taken the training required for all PI's here at the university?

Review the interviewee's answer for Question #4, then ask the following questions (if needed):

- Where and from whom did you receive this training or instruction?

- What was covered and what did you learn?

Module 9 – Data Preservation

Have the interviewee answer question 1 on the worksheet. Then ask the interviewee to talk a little about his/her response.

Be sure that you understand the interviewee's response before continuing on to the next question. You may want to ask some follow up questions for clarification purposes such as:

- Which of the lifecycle stages that you identified for your data contain the parts of the data that should be preserved?

- Is the data that you are willing to share with others, the data that needs to be preserved? Are there additional parts or components of the data that need to be preserved beyond the data that you are willing to share?

Question #1 – Follow Up Question

- Is there any additional information about your data set or that relates to your data set that needs to be preserved?

Question #2 – Follow up Questions

- Why did you select this length of time?

Question #3 – if the answer is "yes", then ask the following question:

- What steps are being taken to preserve this data, when and by whom?

- In your opinion, are the steps taken to preserve your data sufficient? Are there any additional steps that you think need to be taken?

Question #3 – if the answer is "no" or "I don't know, then ask the following question:

- What steps would be needed to preserve this data set?

- Why aren't these steps being taken? What prevents you or your lab from taking these steps?

Module 10 – Educational Needs

Have the interviewee respond to the categories on the worksheet by circling his/her response.

Respond to any questions the interviewee may have about the categories and explanatory text, but there's no need to ask follow up questions of the interviewee at this point (beyond the "additional questions" below).

Additional Questions:

- What do you wish you had known or been taught about data management, sharing or preservation before starting this project?

- What was the most useful data training you received (not only during this project but overall in your education so far)?

- We've talked a lot about the systems you use for managing, sharing, and preserving data for your project. How well do you feel these systems are working for you? What additional training or tools do you wish you had?

Collecting Data Samples and/or Additional Sources of Information

- Ask the interviewee if you could have a sample of the data.
- Ask the interviewee if there are any relevant documents or publications that would help you better understand the data that he/she would be willing to share with you.

END OF INTERVIEW – THANK THE INTERVIEWEE FOR THEIR TIME