

11-29-2010

# The Data Curation Profiles Toolkit: Interviewer's Manual

Jake Carlson

*Purdue University*, [jakecar@umich.edu](mailto:jakecar@umich.edu)

Follow this and additional works at: <http://docs.lib.purdue.edu/dcptoolkit>

 Part of the [Library and Information Science Commons](#)

---

## Recommended Citation

Carlson, Jake, "The Data Curation Profiles Toolkit: Interviewer's Manual" (2010). *Data Curation Profiles Toolkit*. Paper 2.  
<http://dx.doi.org/10.5703/1288284315651>




This work is licensed under a [Creative Commons Attribution-Noncommercial-Share Alike 4.0 License](#).

This document has been made available through Purdue e-Pubs, a service of the Purdue University Libraries. Please contact [epubs@purdue.edu](mailto:epubs@purdue.edu) for additional information.



# The Data Curation Profiles Toolkit

## Interviewer's Manual

<b>Author</b>	Jake Carlson
<b>Publisher</b>	Purdue University Libraries / Distributed Data Curation Center
<b>Contact</b>	<a href="mailto:jcarlso@purdue.edu">jcarlso@purdue.edu</a>
<b>Date of Creation</b>	November 29, 2010
<b>Date of Last Update</b>	November 29, 2010
<b>Version</b>	V 1.0
<b>Acknowledgement</b>	Based on research funded by the IMLS (LG-06-07-0032-07) "Investigating Data Curation Profiles across Research Domains" by D.S. Brandt, J. Carlson, M. Witt (Purdue University Libraries), M. Cragin, C. Palmer (GSLIS University of Illinois Urbana-Champaign).
<b>URL</b>	<a href="http://www.datacurationprofiles.org">http://www.datacurationprofiles.org</a>
<b>License</b>	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 to identify areas of need pertaining to the handling and curation of this particular data. 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 question 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:*

“I would now like to have you fill in module 1 of the worksheet: 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 and how would you characterize 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, how would you characterize this stage and its associated activities?”

*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.*

*For each data stage mentioned be sure to gather the following information (if interviewee does not bring it up him/herself):*

- What happens to the data in each 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.)?

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

## Module 3 – Sharing

*For question #1, read the following text to the interviewee:*

“I would like you to indicate what data from the stages in your data’s lifecycle you would be willing to share and with whom by placing a checkmark in the appropriate boxes. Please check as many of the boxes as apply.”

*It may help to write down the “title” that the interviewee gave the data stage in “module 2 – the lifecycle of the data” in these boxes. For example, the interviewee may have written in “data normalization” for the Second Data Stage, so you may want to write “normalized” in the Second Data Stage cell in this table.*

*Once the interviewee has finished checking boxes, confirm your understanding of the data that the interviewee would be willing to share and with whom.*

*For question #2, once the interviewee has provided a response, ask the following question:*

- Do you currently share your data with any one?
  - If “yes” then ask:
    - Who do you share your data with?
    - How do you share your data?
    - Do you place any conditions on sharing your data?  
(For example, do you ask that the recipient keep the data confidential)?



*Once the interviewee has finished completing questions #3, 4, and 5 - discuss his or her responses, ask for more details if needed.*

*Before moving on to the next module, ask the following question:*

- What publications have resulted from this data?

Be sure to get the name of the journal(s) and approximately when the publication(s) appeared in the journal(s).

## Module 4 – Access

(Note: for the purposes of this interview, “repository” is broadly defined and may be institutional, discipline, regional, publisher, etc.)

*Have the interviewee answer questions #1, 2, and 3 on repositories on the worksheet. Then ask him/her the following questions:*

*If the Interviewee answered “yes” to Question #1 then ask:*

- What made you decide to deposit your data into this particular repository?

*If the Interviewee answered “yes” to Question #2 then ask:*

- Are there any services in particular that you would want the repository to provide?

*If the Interviewee answered “no” to Question #2 then ask:*

- What concerns do you have about submitting your data to a public repository?

Question #3a refers to the stages in the Data Lifecycle that the researcher identified in Module 2 – The Lifecycle of the Data Set

*If the Interviewee provided an answer to Question #3b then ask:*

- Could you tell me why you selected (the timeframe selected)?

*Have the interviewee answer question #4 on priorities for repositories on the worksheet. Then ask the interviewee to talk a little about his/her responses:*

- What makes these services a (low/med/high, etc.) priority for you?

## Module 5 - Transfer of Data / Ingest into a Repository

*Once the interviewee has finished completing question #1, discuss his or her responses. If the researcher needs some prompting, give him or her some examples of actions that might need to be taken before the data could/should be ingested or transferred. These examples might include:*

- *a need to review the data to check for errors or ensure quality*
- *a need to “scrub” the data to modify or remove any variables or information that could be used to identify human subjects.*
- *a need to ensure compliance with funding agency protocols or IRB requirements*
- *a need to obtain sign off from stakeholders*
- *a need to gather and/or review relevant documentation and materials that support the use, curation, etc. of the data*

*Once the interviewee has finished completing this question, discuss his or her response. Be sure that you understand the response(s) before moving on to the next question.*

*Have the interviewee answer question #2 on priorities for ingest into a repository on the worksheet. Then ask the interviewee to talk a little about his/her responses:*

- *What makes these services a (low/med/high, etc.) priority for you?*

## Module 6 – Organization and Description of Data

*Have the interviewee answer questions 1 and 2 on the worksheet about how the data are organized. Review what the Interviewee has written for question 1 to be sure that you understand it. Ask for clarification and additional information as needed.*

*Then ask him/her the following questions:*

If the Interviewee answered “no” to Question #2 then ask:

- What additional description or organization would be needed for someone else with similar expertise to understand and use your data?
- Is there any additional contextual information or documentation would be required for others to understand and make use of your data?

*Have the interviewee answer question #3 on priorities for organization and description of the data set. Then ask the interviewee to talk a little about his/her responses:*

- What makes these services a (low/med/high, etc.) priority for you?

## Module 7 – Discovery

*Have the interviewee answer questions about priorities for the discovery of the data set on the worksheet. Then ask the interviewee to talk a little about his/her responses:*

- What makes these services a (low/med/high, etc.) priority for you?

*Review what the Interviewee has written for question 2 to be sure that you understand it. Ask for clarification and additional information as needed.*

*You may want to suggest some additional possible means of discovery if the researcher needs prompting: through a citation in a publication, library catalog, blog / RSS feed, etc.*

## Module 8 – Intellectual Property

*Have the interviewee answer questions about intellectual property issues pertaining to the data set on the worksheet.*

*After reviewing the interviewee's answers for Questions #1 & #2, ask the following question:*

Are there any other stakeholders associated with this data?

Stakeholder can be defined 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 release and use.

*Review the interviewee's answers for Questions #3 & #4:*

- *If the interviewee responded "yes" to any of parts of Question #3, then ask the following:*
  - Could you provide me with the details of the requirements made by the funding agency?
  
- *If the interviewee responded "yes" to Question #4, then be sure that the privacy or confidentiality concerns are explained well enough for you to understand them. Also ask the following:*
  - How might these concerns associated with your data be addressed?

## Module 9 – Tools

*Have the interviewee answer question #1 on the worksheet and then discuss it with him or her. Ask the following questions, if they are not addressed by the researcher:*

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

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

*Have the interviewee answer question 2 on the worksheet relating to priorities on tools. Then ask the interviewee to talk a little about his/her responses:*

- What makes these services a (low/med/high, etc.) priority for you?

## Module 10 – Linking / Interoperability

*For question #1:*

- If the interviewee responded “yes” to Question #1 on the worksheet, then ask the following question:
  - How do you prepare your data for submission to the journal or place of publication?

*Have the interviewee answer question 2 on the worksheet relating to priorities on services relating to interoperability. Then ask the interviewee to talk a little about his/her responses:*

- What makes these services a (low/med/high, etc.) priority for you?



## Module 11 – Measuring Impact

*Have the interviewee answer question 1 on the worksheet relating to priorities on services relating to measuring impact. Then ask the interviewee to talk a little about his/her responses:*

- What makes these services a (low/med/high, etc.) priority for you?

*If the interviewee assigns a high or a medium priority to “The ability to gather information about the people who have accessed or made use of this data”, then ask the following question:*

- What kind of information about the people who are accessing or using your data would you like to be able to collect?

*Have the interviewee answer question #2 on the worksheet and then discuss it with him or her. Be sure that the interviewee explains his/her needs well enough for you to understand them.*

## Module 12 – Data Management

*After the interviewee answers question #2 on the worksheet, ask him/her the following question:*

- If you were to submit your data set to a repository, how often would you want your data to be backed-up?

*After the interviewee answers question #3 on the worksheet, ask him/her the following question:*

- If you were to submit your data set to a repository or transfer it to us, what security measures would you want enacted to protect your data?

*Have the interviewee answer question 4 on the worksheet relating to priorities on services relating to data management. Then ask the interviewee to talk a little about his/her responses:*

- What makes these services a (low/med/high, etc.) priority for you?

## Module 13 – 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 contains 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?

*Have the interviewee answer question 2 on the worksheet relating to the length of time the data should be preserved. Then ask the interviewee to talk a little about his/her response:*

- Why did you select this length of time?

*Have the interviewee answer question 3 on the worksheet relating to priorities on services relating to preservation. Then ask the interviewee to talk a little about his/her responses:*

- What makes these services a (low/med/high, etc.) priority for you?

## **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**