Data Curation Profiles Toolkit

11-29-2010

The Data Curation Profiles Toolkit: User Guide

Jake Carlson
Purdue University, jakecar@umich.edu

The Data Curation Profiles Toolkit

User Guide

Author	Jake Carlson
Publisher	Purdue University Libraries / Distributed Data Curation Center
Contact	jrcarlso@purdue.edu
Date of Creation	November 29, 2010
Date of Last Update	November 29, 2010
Version	V 1.0
Acknowledgement	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).
URL	http://www.datacurationprofiles.org
License	Creative Commons Attribution-NonCommercial-ShareAlike 3.0 License.

Table of Contents

Part 1 - Background

What is a Data Curation Profile?
A Data Curation Profile is a tool that can be used to provide a foundational base of information about a particular set of data that may be curated by an academic library or other institution.

Data curation, as defined by the Graduate School of Library and Information Sciences at the University of Illinois at Urbana-Champaign is “the active and on-going management of data through its lifecycle of interest and usefulness to scholarship, science, and education. Data curation activities enable data discovery and retrieval, maintain its quality, add value, and provide for re-use over time, and this new field includes authentication, archiving, management, preservation, retrieval, and representation” - http://www.lis.illinois.edu/academics/programs/ms/datacuration.

A completed Data Curation Profile will contain two types of information about a data set. First, the Profile will contain information about the data set itself, including its current lifecycle, purpose, forms, and perceived value. Second, a Data Curation Profile will contain information regarding a researcher’s needs for the data including how and when the data should be made accessible to others, what documentation and description for the data are needed, and details regarding the need for the preservation of the data.

How was the Data Curation Profile developed?
The Data Curation Profile is the result of a two year research project conducted by the Purdue University Libraries and the Graduate School of Library and Information Science at the University of Illinois at Urbana-Champaign. The central goals of this research were to identify how researchers are managing or curating their data as well as to determine “who is willing to share their data with whom, when and under what conditions”. A total of nineteen faculty participated in the project which consisted of multiple interviews with project personnel about their data and their needs. The interviews were semi-structured in nature to allow faculty participants to be able to direct the discussion about their data and their needs.

The Data Curation Profile tool was developed as a key component of this research. The transcribed interviews were reviewed and coded using a grounded theory approach to capture and compare the information that had been gathered. Although the needs of faculty naturally differed from one another, a number of commonalities surfaced amongst the participants of the study. The Data Curation Profile was developed as a framework to distill and present the information collected from participants in a way that enabled comparisons to be made.

The principal investigator on this project was D. Scott Brandt, Associate Dean of Research at Purdue University. The co-principal investigators were Jake Carlson and Michael Witt from Purdue University, and Melissa Cragin and Carole Palmer from the University of Illinois at Urbana-Champaign.

A more in-depth description about the development of the Data Curation Profile tool is contained in the following article:

- Witt, M., Carlson, J., Brandt, D.S., & Cragin, M.H. (2009) “Developing the Data Curation Profiles” *International Journal of Digital Curation*, 4(3), 93-103. http://www.ijdc.net/index.php/ijdc/article/view/137

Part 2 – Purpose and Use

Principles of the Data Curation Profile
The key element of the Data Curation Profile is that it attempts to represent the perspective of the researcher or research group, rather than those of the data curator, librarian, archivist, etc. The Data Curation Profile template provides a means to distill these needs into a standardized, yet flexible, set of categories. The information needed to generate a Data Curation Profile is primarily gathered through conducting interviews with the lead researcher and/or other research personnel working on a particular research project.

The Data Curation Profile is intended to address the needs of an individual researcher or research group with regards to the "primary" data generated or used for a particular project. The data that will be addressed in the interviews and serve as the focus of the subsequent Data Curation Profile should be determined in the preparation phase of the process of developing a Data Curation Profile (see "Part 4 – How to Develop a Data Curation Profile" in this User's Guide).

Primary data is defined as the data that constitutes the focus or basis of the research and without which the research could not take place. The primary data may include more than one data set or data type. For example, the primary data for a clinical research project may include multiple samples taken from the subject, multiple measurements of the samples, and multiple analyses of the results that may appear in different formats (spreadsheets, images, text, etc.).

Over the course of the interview, the researcher may bring up additional data that has been generated or brought in from outside sources. Ancillary data refers to data that is used for primarily verification or reliability purposes or data that is secondary in nature to the primary data and the research being conducted. Continuing the example from the previous paragraph, a clinical study may include data generated from analyzing urine samples for the presence of a nutritional supplement that was administered to subjects. This is ancillary data as it is needed to verify that the subject received the correct amount of the supplement, but is not used for research purposes directly.

It can sometimes be difficult to determine which data are primary to the project and which are ancillary in nature. In these situations the interviewer may want to ask the researcher being interviewed to define which of the data he or she considers to be "primary" and which are "ancillary".

What can a Data Curation Profile be used for?
At an individual level, the Data Curation Profile:

- Provides a structure for conducting a data interview between an information professional and a researcher or research group.
- Provides a means for a researcher or a research group to thoughtfully consider their needs for their data beyond its immediate use.

At an institutional level (Library, University, etc.), the Data Curation Profile:

- Can serve as a foundational document to guide the management and/or curation of a particular data set.
- Can be shared with staff providing data services and others to inform them and ensure that everyone is on the same page.

- May be used to inform the development of data services to be offered by the institution, as well as to help to identify the types of tools, infrastructure and responsibilities for data services staff.

At its broadest level, the Data Curation Profile:

- May be used by others as a guide in developing data services at their own institutions.
- May be used as objects of research to further a better understanding of data types researchers want or need to share, curate or preserve, and the needs of researchers in doing so.

What a Data Curation Profile is not designed to do.
The Data Curation Profile is not aimed at generating an inventory of data sets. Its purpose is to provide a depth of information about a particular data set rather than a breadth of information or a general awareness of data sets that exist within a particular institution or organization.

Although Data Curation Profiles are meant to provide detailed information, they are not meant to be comprehensive in nature. Instead, Data Curation Profiles are structured to encapsulate information about a data set to be curated and capture many of the basic elements pertaining to curating the data, as well as to form a foundational document for investigating and developing curation infrastructures, services, and policies.

The needs of the researcher and the actions taken by data curators may shift, evolve or require more granular explanations over time. Although a completed Data Curation Profile may be annotated to account for these changes if desired, it is not meant to serve as a means of capturing needs on an ongoing basis or as a mechanism to document actions taken in curating the data. Instead, a Profile represents a snapshot of the data and researcher needs at a particular point in time.

Part 3 - Components of the Data Curation Profile Toolkit

The Data Curation Profile Toolkit is composed of four documents.

1. **The User Guide –**
   The User Guide (this document) provides basic information about the Data Curation Profiles, and directions on how to construct a Data Curation Profile.

2. **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. Some of the questions to be asked will be in response to the answers given by the researcher in the Interview Worksheet (see below).

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

3. **The Interview Worksheet –**
   The Interview Worksheet is to be given to the researcher by the interviewer at the start of the interview (or sent in advance). It is the worksheet that the participating researcher

will fill out over the course of the interview. In addition to capturing important information, the responses provided by the researcher will serve as the basis for further discussion during the interview.

4. **The Data Curation Profile Template –**
   The Data Curation Profile Template describes the structure of the Data Curation Profile. Each section or sub-section within the Data Curation Profile template contains a brief definition of the information that is needed to populate an individual Data Curation Profile for the participating researcher.

Part 4 - How to Develop a Data Curation Profile

Methodology
A Data Curation Profile is developed through 3 stages:

- Stage 1 – Preparation
- Stage 2 – Interviews
- Stage 3 – Constructing the Profile

Stage 1 – Preparation:
The Data Curation Profile tool is best applied with researchers who have been identified as having data and as having a desire or a need to do something with that data. Possible use cases for a Data Curation Profile are listed in the "What can a Data Curation Profile be used for?" section of this User Guide, which can be used to explain the potential value of generating a Data Curation profile to the researcher. Naturally, the interviewer or researcher may generate other possible use cases or additional value propositions for developing a Data Curation Profile.

The data that will serve as the focus of the interviews and the resulting Data Curation Profile should be identified. The criteria for selecting the data to discuss in the interviews and the profile will vary depending on the researcher and his or her situation, and should be negotiated between the interviewer and participating researcher. The interviewer may want to consider data that are in a more mature state rather than in a planning stage, as well as data that best represent the "typical" data that are generated or used by the researcher for use in generating a profile.

Once the data to be profiled has been selected, it is recommended that the interviewer do some preliminary investigations about the work being done by the researcher and his/her use of the data as preparation for conducting the interviews. This could include reviewing information posted on the researcher's website or reading some of the researcher's publications. Identifying the broad needs of the researcher beforehand enables the interview to be modified if desired (see "modifications" below).

The personnel involved in the development of the Data Curation Profile should be identified at this stage. In particular, the responsibilities for conducting the interviews, transcribing the interviews, and drafting the Data Curation Profile itself should be determined early on in the process.

It is recommended the person who conducts the interviews also be the one to create the Data Curation Profile, as he or she will likely have the greatest in-depth understanding of the data and the researcher's needs. However, as this is not always possible, another person familiar with the Data Curation Profile process and structure could draft the Profile using the recording of the interview, the interview worksheet and the other materials gathered from the researcher. In such cases, the interviewer should review the draft of the Profile, make corrections or additions as needed, and sign off on the Profile before it is considered complete.

All personnel should familiarize themselves with the components of the Data Curation Profile Toolkit before proceeding to the next stages.

Institutional Review Board Review and Approval

If you plan on using the Data Curation Profile for research purposes, including presenting or publishing content from the interviews or the Profile, you may need to undergo a review from your institution's Institutional Review Board (IRB). For research purposes, the researcher and anyone else you interview may be considered to be human subjects and therefore need to be made aware of their rights in being interviewed. Even if you are not planning on using your Data Curation Profile for research purposes, if there is any question about whether or not you need to go through the IRB review process, you should ask your institutions' Institutional Review Board for guidance.

If a review by the IRB is required, the application to the IRB must be approved by the IRB before any interviews can be scheduled or conducted. The application and review process can take some time so be sure to plan accordingly. The National Institutes of Health's web page on "Research Involving Human Subjects" provides additional information on this subject – http://grants.nih.gov/grants/policy/hs/index.htm.

Modifications

The Data Curation Profile tool provides for some flexibility in its application and can be modified if needed.

1. The Data Curation Profile is modular in nature. At its core, the Data Curation Profile is a tool to capture information about a particular data set and a means to determine when, with whom, and under what conditions the data will be shared with others. Although it is recommended that **all** sections be included in the process of developing a Data Curation Profile, some pieces of the Profile can be removed or replaced with locally generated sections if desired.

    a. The sections of the Data Curation Profile that are **Required** are:
    Section 2 - Overview of the research
    Section 3 - Data kinds and stages
    Section 5 - Organization and description of data
    Section 7 – Sharing & Access

    b. The sections of the Data Curation Profile that **can be removed** if necessary are:
    Section 1 - Brief summary of data curation needs
    Section 4 - Intellectual property context and information
    Section 6 - Ingest / Transfer
    Section 8 - Discovery
    Section 9 - Tools
    Section 10 - Linking / Interoperability

Section 11 - Measuring Impact
Section 12 - Data Management
Section 13 - Preservation
Section 14 - Personnel

2. The Data Curation Profile contains a base set of questions to be answered during the interview process; however additional questions can be added to the modules within the Profile. Or entire modules could be added if needed, depending on individual needs.

   Additional questions generally should not preclude or replace the original questions listed in each module within the Interview Worksheet and Interview Manual. Be sure that any changes made in the Interview Worksheet, Interview Manual, or Data Curation Profile Template are reflected and accounted for in the other documents.

Time Needed
The time needed to complete the preparation stage will obviously vary depending on the amount of background research required, the extent of the modifications made to the interview documents, and if a review by an IRB is required. Assuming that IRB is not required, the average preparation time should be approximately 1 – 2 hours.

Stage 2 –Interviews
The primary means of gathering the information needed to produce a data curation Profile is through interviewing the participating researcher. Other personnel associated with the data may be interviewed as well in order to add to the richness of the Profile, but it is not a requirement.

Conducting the Interviews
The Data Curation Profile tool kit contains an "Interview Manual" and an "Interview Worksheet". These two documents are meant to be used in conjunction with one another.

At the beginning of the interview give the participating researcher the "Interview Worksheet", then open the "Interviewer's Manual" and read the "Introduction to the Interview" statement.

- Alternatively, the interview worksheet may be given to the researcher beforehand to enable him or her to review and prepare for the interview. However, the researcher may have questions or need direction on how to fill out the worksheet. Therefore, even if the worksheet is given to the researcher before the interview, it is strongly recommended that the researcher wait to fill out the worksheet until the actual interview takes place.

After the introductory statement there are several "Background / Demographic Questions" contained in the "Interviewer's Manual" that should be asked of the interviewee.

From this point onwards, both the "Interview Worksheet" and the "Interviewer's Manual" documents are organized by module. At the beginning of each module:

- Read any introductory statement listed in the "Interviewer's Manual" (not all modules have an introductory statement).

- Then have the participating researcher complete the list of questions for the module in the “Interview Worksheet”.

- Review the participating researcher’s responses in the worksheet and ask the additional questions listed in the “Interviewer’s Manual” as appropriate.

- Ask any follow up questions you feel are needed for clarification, to provide more detail or to pursue additional information of interest.

- Be sure that you and the participating researcher have completed all of the questions within a module before moving on to the next module.

Several of the questions listed in the Interviewer’s Manual are follow-up questions and the need to ask them will be contingent on how the participating researcher responds in the Interview Worksheet. For example in “Module 4 – Access”, if the participating researcher circles “yes” on question #1, the interviewer should ask the follow-up question listed in the first bullet point in Section 4 of the Interviewer’s Manual. If the participating researcher circles “no” or “I don’t know”, then there is no need to ask the follow-up question.

It may also be useful to take the User Guide and Data Curation Profile Template documents along to the interview to refer to if needed.

The Need for Two Interviews

The information needed to populate a Data Curation Profile is typically gathered by conducting two separate interviews with the participating researcher, having the researcher complete a worksheet as a part of the interviews, and through collecting and reviewing other relevant information such as a sample of the data or publications that describe how the data was generated, used, etc. Other members of a research group may be interviewed separately as warranted or desired.

Each of the two interviews usually requires an hour’s time to complete at minimum. The interviewer may want to consider allotting additional time for the interviews depending on the willingness and the temperament of the participating researcher.

Conducting two separate interviews is strongly recommended as the amount of information needed to construct a Data Curation Profile cannot be solicited adequately within an hour’s time. Condensing the two interviews into one extended length interview is possible, but not recommended as both interviewer and interviewee may grow fatigued causing the quality of the information gathered to suffer. Conducting two interviews also enables the following advantages:

- Conducting a second interview allows the interviewer to identify responses or areas in which clarification or more details are needed, and then to ask follow up questions to elicit this information as a part of the second interview.

- Through conducting the first interview, the interviewer will gain a direct understanding of the temperament and conversational style of the participating researcher which will better enable the interviewer to prepare for and accommodate the researcher in the second interview.

The second interview should be conducted in the same way as the first interview was, as described in the “conducting the interviews” section above, with the researcher filling out a module within the Interview Worksheet and the interviewer asking questions from the Interview Manual. Follow up questions resulting from a review of the first interview should be incorporated into the Interviewer’s Manual and/or Interviewers Worksheet as appropriate to ensure that they are included in the second interview.

Coverage

It is strongly recommended that the first interview cover the following modules at minimum:

- Module 1 - The Data Set
- Module 2 - The Lifecycle of the Data
- Module 3 - Sharing
- Module 4 - Access

The precise stopping point for the first interview is a decision left to the interviewer and may need to be made during the interview itself, as the nature of each interview will be different based on the temperament and conversational styles of the participating researcher (among other things). In determining how much of the material to cover in the first interview, keep in mind that the remainder of the interview and worksheet will need to be completed by the end of the second interview.

The second interview should be scheduled after information from the first interview has been reviewed and analyzed, but care should be taken not to allow too much time between the first and the second interview.

Additional Considerations for Conducting the Interviews

The interviews should be recorded. Although the “Interview Worksheet” will capture a lot of the basic information needed to construct a Data Curation Profile, the questions asked by the interviewer and the subsequent discussion will provide the details and the context needed to develop a Profile. Recording the interview may require some extra time and effort. However, it is extremely difficult to conduct an interview and take notes simultaneously, and written notes, even if taken by another person attending the interview, will not fully capture the details of the interview. Therefore, important information will almost certainly be lost if the interview is not recorded.

Be sure to inform the researcher that this interview will be recorded and secure his/her permission to make the recording before the interview takes place.

It is recommended that the audio recordings of interviews be transcribed. In addition to making the information contained in the interviews easier to work with, the transcription process can help personnel become more knowledgeable about the information contained in the interviews. The standard estimated time for transcribing is approximately one (1) hour of transcription time for every 10 - 15 minutes of audio recording.

An alternative to transcribing the interview fully would be to create an index of the audio file of the interview. In indexing the interview, you would indicate the time where a new topic of discussion was introduced in the audio recording of the interview and take notes on the key issues raised and points made by the interviewee. An example of an indexed interview is below:

5:10 - Module 2 explanation (interviewer)

6:00 - Background information on the research
Test apparatus at UIUC. Hydraulic actuator that drives the damper

6:18 - Data acquisition stage - obtaining measurements from the damper. The displacements, velocities, voltages that are applied as well as the resulting outputs - mainly the force. Data are stored on a sig lab box as a vna file - which is used by a specific data acquisition system.

Indexing the audio file should take less time to develop in comparison to a full transcription and still provide enough information to generate a profile, provided that the notes are taken at a sufficient level of detail. Indicating when a discussion took place also allows you to easily identify a relevant portion of the audio file should questions arise or more information is needed.

Recommendations - Interviews

Interviewing faculty and other research staff can be a challenging endeavor and may take some practice. The following are some recommendations for the interviewer:

- If you are new to conducting interviews, you may want to practice beforehand - Interviewing a colleague about the work that they do, or a recent project they completed is one possible way to gain experience with interviewing people.

- Familiarize yourself with the Data Curation Profiles interview framework and questions before conducting the interview –
  Knowledge of the questions beforehand and being at ease with the structural framework will help the interview go smoothly.

- Keep participating researcher on task and focused –
  Remember that the purpose of the Data Curation Profile is to provide information on a particular data set, and capture the researcher’s needs for that data set (and only that data set). One of the challenges that the interviewer may encounter is the participating researcher going off tangents. For example, the researcher may start referring to other data sets that he/she works within the discussion, or the researcher may veer off towards discussing the aspects of the research that do not pertain to the data.

  Some deviation in the discussion is perfectly acceptable and even desirable as it may lead to interesting information and insights that would not have otherwise come to light. However the interviewer may need to rein in the discussion if it appears that it is going too far afield. This can be done through redirection such as: “Is what you have said also true for the data we are discussing for this interview?” or “Going back to the data we are discussing for this interview…” Another strategy could be to suggest a topic be “parked” for discussion later, at the end of the interview or another meeting.

- Sit close to the researcher –
  Many of the follow up questions in the Interviewer's Guide will come in response to the answers given by the participating researcher in the Interview Worksheet. Therefore, the interviewer will need to sit in close proximity to the researcher so that he/she can see what the researcher has written in the Interview Worksheet.

- Ask follow up questions -
  The questions in the "Interview Worksheet" and the "Interviewer's Manual" are meant to serve as the base set of questions to be asked. As the circumstances and needs of each participating researcher will be unique, these base questions will likely be insufficient in and of themselves to capture a full understanding of the data set and the needs of the participating researcher. The interviewer may want to ask follow up questions that are not listed in the Worksheet or the Manual for this purpose.

- Ask for clarification or definitions if needed –
  During the interview the participating researcher may discuss technical or specialized information that is unfamiliar to you. If this should happen, it is okay to ask the researcher to define or clarify a term or concept, or to provide more information. The participating researcher should understand that you are not an expert in his or her field.

- Manage your time effectively -
  The interviewer will need to maintain a balance between gathering information from the participating researcher and keeping the interview moving forward. During the interview, the interviewer will need to monitor how much time has elapsed and take care to inform the participating researcher of the time as well. A strategy for this might be to divide the time available by the number of topics and try to stick to that many minutes for discussion, realizing that if you go over 5 minutes on one topic you may need to rush a bit through another topic.

- Bring an extra audio recorder and extra batteries –
  Mistakes and accidents happen, so it is very important to be prepared. Check your equipment beforehand to be sure that it is in good working order and that you have enough space on the recorder to capture the full interview. Batteries can expire at inopportune times so bring an extra set with you and check to be sure the recorder is still operating periodically during the interview. Having two recorders taping the interview can prevent disaster if one of them stops working.

Storage and Handling of Materials
It is strongly recommended that the audio recording of an interview be transferred from the recording device and into a secure environment as soon as possible after the interview has been conducted. Making back-up copies of digital recordings or other materials and storing them in separate locations from the originals is also recommended.

If you were required to go through a review by your institution's Institutional Review Board (IRB) to develop a Data Curation Profile, be sure that your storage and handling practices comply with the terms and conditions stated in your approved IRB review. At minimum, IRB compliance typically includes ensuring that all materials, whether in physical or digital formats, are de-identified and stored in a secure environment.

Time Needed
Once again the time needed to complete the interviews stage will vary depending on the duration and complexity of the information discussed in the interviews.

Each interview on average will last approximately 1 – 1.5 hours (not including any time spent preparing for the interviews).

The standard estimated time for transcribing is approximately one (1) hour of transcription time for every 10 - 15 minutes of audio recording. This translates to approximately 4 – 9 hours of transcription time on average for each interview.

As an alternative to transcription, indexing the interviews as described in the "Additional Considerations for Conducting the Interviews" section of this User Guide, should take less time to perform. It is estimated that indexing an interview would take approximately 3 – 7 hours, again depending on the length and complexity of the interview.

Stage 3 – Constructing the Data Curation Profile
The final stage in the process is to take the information gathered from the interviews and from any additional sources of information that were gathered and use them to fill in the sections and sub-sections of the Data Curation Profile template.

The sections within the Data Curation Profile template are generally analogous to the modules within the "Interview Worksheet" and the "Interviewer's Manual" to facilitate the process of creating an individual Profile. The table below demonstrates the relationships between interview modules and sections of a Data Curation Profile.

Interview Module		Data Curation Profile Section
Background / Demographic Questions (Interviewer's Manual)		Section 2 - Overview of the Research
Module 1 – The Data Set		Section 3 - Data Kinds and Stages
Module 2 – The Lifecycle of the Data Set		
Module 3 – Sharing		Section 7 – Sharing & Access
Module 4 – Access		
Module 5 – Transfer of Data / Ingest into a Repository		Section 6 - Ingest / Transfer
Module 6 – Organization and Description of Data		Section 5 - Organization and Description of Data (incl. metadata)
Module 7 – Discovery		Section 8 - Discovery
Module 8 – Intellectual property		Section 4 - Intellectual property context and information

Module 9 – Tools		Section 9 - Tools
Module 10 – Linking / Interoperability		Section 10 – Linking / Interoperability
Module 11 – Measuring Impact		Section 11 - Measuring Impact
Module 12 – Data Management		Section 12 – Data Management
Module 13 – Data Preservation		Section 13 - Preservation

Keep in mind, however, that the participating researcher may have discussed relevant aspects of the data and their needs for the data outside of a particular module. For example, the participating researcher may have brought up an issue relevant to the discovery of his/her data while answering questions within the "access" module. Therefore it is important to review the interviews as a whole for information to populate the sections of a Data Curation Profile rather than just relying on information from a particular module.

A Data Curation Profile is meant to incorporate and convey the voice of the researcher. Therefore, although the person constructing the Profile will naturally be the one to craft the language used in populating the sections and sub-sections, care should be taken to provide an authentic representation of the researcher's voice in the Data Curation Profile.

Do not drop sections or sub-sections from the Data Curation Profile (unless of course the decision to exclude a particular module was made before the interviews were conducted). It is likely that there will be some gaps in the information gathered that will lead to some of the sections or sub-sections in the Data Curation Profile containing little or no content. If this happens, do not delete the section or sub-section from the Profile or leave it blank. Instead, indicate that not enough information was available from the information collected to make an entry by writing "Not discussed by the researcher" or a similar statement.

Additional sections and sub-sections can be added to individual Data Curation Profiles as warranted.

After the Data Curation Profile has been drafted you may want to consider asking the researcher(s) who were interviewed to review the draft and provide feedback.

Examples of completed Data Curation Profiles are available at the project's website: http://www.datacurationprofiles.org

Time Needed

The amount of time needed to construct a Data Curation Profile will naturally vary according to the amount of information collected, the complexity of the data and the researcher's associated needs for the data, and other factors. Generally it should take approximately 5 – 10 hours to complete a Data Curation Profile.

Sharing your Data Curation Profiles

As the Data Curation Profiles you construct will likely be of interest and of value to others beyond your home institution, you are strongly encouraged to add your Profile to the collection of Data Curation Profiles that are publicly available at http://datacurationprofiles.org.

There are many reasons to share your Data Curation Profile with others, including:

- Sharing will help the library/information sciences field better understand the issues and needs in working with research data. Data curation has been identified as an important area for librarians and other information professionals to pursue, however there are few mechanisms available to inform where and how librarians could get involved at a practical level. Sharing your Profile with others provides real-world examples at the ground level that can help inform the directions of the library field, and shape the development of data services at individual libraries.

- Sharing is a means for you to gain recognition for your work in developing a Data Curation Profile. Sharing your Profile through http://datacurationprofiles.org provides the following advantages:
    - Your Profile will be indexed by major search engines such as Google and by academic search services as well, making your Profile more easily discovered.
    - Your Profile will be assigned a unique URL, enabling you and others to cite the Profile.
    - You will be provided with usage statistics for your Profile on a monthly basis, allowing you to demonstrate the impact of your work.
    - You will retain copyright over your work

- Sharing your Profile increases the visibility of your institution in an emerging area of importance in the library and information fields.

Any restrictions on use of the Data Curation Profile should be clearly stated in the “Notes” section of the introductory table at the beginning of the Profile.

In addition, please be sure that you contact the participating researcher and secure his/her permission before sharing the Data Curation Profile with others.

Acknowledgements

The Purdue University Libraries would like to thank the Institute of Museum and Library Services (IMLS) for their generous support for the research project that produced the Data Curation Profiles (IMLS LG-06-07-0032-07), and for the support of the Data Curation Profile Toolkit workshops that will be offered in 2011 (IMLS RE-06-10-010-10).

We would also like to thank the following people for reviewing earlier drafts of the Data Curation Profile toolkit and providing their feedback and suggestions:

Dianne Dietrich – Cornell University
Barrie Hayes – University of North Carolina – Chapel Hill
Patricia Hswe – Penn State University
Ardys Kozbial – University of California – San Diego
David Minor – University of California – San Diego
Susan Parham – Georgia Tech University

Marisa Ramirez – California Polytechnic State University
Gail Steinhart – Cornell University
Brian Westra – University of Oregon

Finally we would like to thank the following team of people at the University of California San Diego (UCSD) and Merced (UC Merced) for beta testing the Data Curation Profile Toolkit:

Ardys Kozbial - UCSD Libraries, Technology Outreach Librarian
David Minor - SDSC, Head of Data Curation
Mary Linn Bergstrom - UCSD Libraries, Head, Science and Engineering Library
Declan Fleming - UCSD Libraries, Head of IT Dept.
Emily Lin - UC Merced, Head of Digital Assets
Sue McGuinness - UCSD Libraries, Pharmacy Librarian
Brad Westbrook - UCSD Libraries, Metadata Librarian

Data Curation Profile – Checklist of Activities

Stage 1 – Preparation

__ Identify a researcher who has data and a desire or a need to do something with that data.

__ Have a conversation with the researcher about the Data Curation Profile and its potential uses. Secure the researcher's participation in developing a Data Curation Profile.

__ In conjunction with the researcher, select the data to be discussed in the interviews and profiled.

__ Determine who will be involved in conducting the interviews and generating the Data Curation Profile, and confirm their involvement.

__ Check with your institution's IRB office to see if an IRB review is required.

__ Contact the researcher to:

- ___ arrange a meeting place and time to conduct the first interview.
- ___ ask for any material that could inform the interview. This could include published papers that describe the data set, a sample of the data set itself, a code book for the data, etc.

__ Prepare for the interview by:

- __ reading or reviewing the materials provided by the research about the data set.
- __ familiarizing yourself with the interview materials and the Data Curation Profile template.
- __ reviewing completed Data Curation Profiles on the http://datacurationprofiles.org website, especially those that are similar to the discipline/sub-discipline of the researcher you will be interviewing.

__ Determine if any modifications or additions need to be made to the interview. Be sure to make the necessary changes to both the "Interview Worksheet" and "Interviewer's Manual", as well as to the Data Curation Profile template documents.

Stage 2 – Interviews

__ On the day of the interview:

- __ Check that your audio recorder(s) are in working order.
- __ Be sure that you have fresh batteries and that you bring them to the interview.

__ Be sure that you have printed out copies of the "Interview Worksheet" and the "Interviewer's Manual" and that you bring them to the interview. If you have made any modifications or additions to these documents be sure that you print out and take the correct versions to the interview.

__ You may want to consider bringing a watch or some other timepiece to the interview.

__ Conduct the 1st interview.

__ Have the researcher read and sign a consent form (if required by the IRB)

__ Give the researcher the "Interview Worksheet" and read the introductory statement on the first page of the "Interviewer's Manual" to the researcher.

__ Begin the Interview

__ Read any introductory statement listed in the "Interviewer's Manual" (not all modules have an introductory statement).

__ Then have the participating researcher complete the list of questions for the module in the "Interview Worksheet".

__ Review the participating researcher's responses in the worksheet and ask the additional questions listed in the "Interviewer's Manual" as appropriate.

__ Ask any follow up questions you feel are needed for clarification, to provide more detail or to pursue additional information of interest.

__ Be sure that you and the participating researcher have completed all of the questions within a module before moving on to the next module.

__ Repeat these steps until the interview is completed or until the time allotted for the interview has expired

__ Transfer the audio recording of the 1st interview off of the recorder and into a secure environment (in accordance with IRB protocol if applicable).

__ Making a back-up copy of the audio file and storing it in a separate location is strongly recommended (in accordance with IRB protocol if applicable).

.

__ Store the signed consent form (if required by the IRB), the completed "Interview Worksheet", and any other materials related to the interview in a secure environment (as defined by the IRB, if necessary).

__ Transcribe the 1st interview, or index the interview by listening to it and taking notes if you decide not to have it transcribed.

__ Analyze the materials you have collected, identify any areas that are unclear or incomplete, and determine what questions you will need to ask in the second interview.

__ Schedule the 2nd interview.

__ Conduct the 2nd interview.

__ Give the researcher the "Interview Worksheet" if it was not completed in the 1st interview.

__ Begin the Interview

__ Read any introductory statement listed in the "Interviewer's Manual" (not all modules have an introductory statement).

__ Then have the participating researcher complete the list of questions for the module in the "Interview Worksheet".

__ Review the participating researcher's responses in the worksheet and ask the additional questions listed in the "Interviewer's Manual" as appropriate.

__ Ask any follow up questions you feel are needed for clarification, to provide more detail or to pursue additional information of interest.

__ Be sure that you and the participating researcher have completed all of the questions within a module before moving on to the next module.

__ Repeat these steps until the interview is completed or until the time allotted for the interview has expired

__ Transfer the audio recording of the 2nd interview off of the recorder and into a secure environment (in accordance with IRB protocol if required).

__ Making a back-up copy of the audio file and storing it in a separate location is strongly recommended (in accordance with IRB protocol if required).

Stage 3 – Constructing the Profile

__ Using the information contained in the two interviews and the other materials gathered from the researcher, construct a draft of the Data Curation Profile.

__ If someone other than the interviewer is constructing the Data Curation Profile, then share the draft of the Profile with the interviewer for his/her review and feedback.

__ If feasible and appropriate, share the draft of the Data Curation Profile with the researcher (and anyone else interviewed) for his/her review and feedback.

__ Develop the final draft of the Data Curation Profile

__ Post your Data Curation Profile to the Data Curation Profiles website to share your work with others. (and please share your experiences in creating and using the Profile with others through the website as well)