

# Data Curation Profiles: Commoditize, catalyze or change?

*Is it nobler in the mind to commoditize the slings and arrow of outrageous catalysis; or to take dots against a sea of troubles and by proposing, change them?*

The Data Curation Profiles (DCP) Toolkit in its current form facilitates discussions with researchers about data in a particular project, what they are doing with it, and what they would like to do with it. As an interview instrument, it has been used successfully by librarians and information professionals for data discussions, adapted and repurposed for other data-related explorations, and adopted in library schools as a teaching tool. Help us determine how to expand and extend the use of the DCP Toolkit.

Human Cell Database Systems

The categories in the "Data Stage" column listed in the table below were developed by the authors of the data curation profiles. The data specifically incorporated in the profiles to create profiles are indicated by the small blue icon in the "Format" column.

Data Stage	Output	Typical File Size	Format	Other Notes
Experiments	Primary results of experiments		xls	Can be copied from a series of single experiments identified by a unique identifier. The information along with the data is included in a 10-150kb workbook. This includes a variety of results from scientific experiments.
Dissect	Description of experimental data		xml	Can be downloaded but show pictures and images that are larger than the data. The information along with the data is included in a 10-150kb workbook. This includes a variety of results from scientific experiments.
Processed	Combinatorial files for statistical analysis		xls	Can be downloaded but show pictures and images that are larger than the data. The information along with the data is included in a 10-150kb workbook. This includes a variety of results from scientific experiments.
Analysed	Statistical analysis		xls, xls	Can be downloaded but show pictures and images that are larger than the data. The information along with the data is included in a 10-150kb workbook. This includes a variety of results from scientific experiments.
Summarized	Summary of statistical analysis	Between 20-50 kb	xls	Can be downloaded but show pictures and images that are larger than the data. The information along with the data is included in a 10-150kb workbook. This includes a variety of results from scientific experiments.
Published	Outputs, charts and tables for inclusion in scientific output		xls	Can be downloaded but show pictures and images that are larger than the data. The information along with the data is included in a 10-150kb workbook. This includes a variety of results from scientific experiments.

**Note:** The data curation profiles for the research data are only available in individual files for the data curation profiles. The data curation profiles are only available in individual files for the data curation profiles.

**Target data for sharing:**  
The target data for sharing is the summarized data. The data that has been shared to have a statistically significant effect on the data capacity to select against certain data.

**Value added value of the data:**  
The data curation profiles for the data with the most value to offer researchers are the summarized results. Therefore, the data that are most valuable in creating a meta-analysis of some kind. The data might also be useful.

**Copyright considerations:**  
The data is considered to be data as it does not change or grow after it has been collected. The data is not a primary component of a larger data set. The data is not a primary component of a larger data set. The data is not a primary component of a larger data set.

**Other notes:**  
The data is considered to be data as it does not change or grow after it has been collected. The data is not a primary component of a larger data set. The data is not a primary component of a larger data set.

**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, databases, 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, databases, JPEG 2000 images, a raw data file from a Campbell CR10 data logger, etc.)

Please continue on to the next page



We'd like your help... We have feedback and input from a variety of sources (attendees of workshops, participants of a DCP symposium, users of the Toolkit), but we would like you, savvy iDCC'ers, to help us move a step further toward DCP2.0.

Help us out by "voting" on elements listed here that you think would make the DCP a better instrument and tool. If you could select one element in each column, what would you suggest would be the key to creating DCP2.0? Ask Scott for a dot to make your choice, and tell us why.

## Commoditize?

What would it take to commoditize the DCP? Outsource a new version to a publisher or repository? Just a matter of making an app so everyone/anyone could use it? Standardizing the format even though it was originally designed to be structured but flexible?

## What should the primary purpose of DCP be?

- Learn about researcher challenges
- Create story/big picture about project's data
- Gather specific info about data mgmt/workflow
- Identify gaps/needs in data mgmt or curation
- Match curation services to researcher needs
- Create data curation profiles for others to use
- Determine data curation metadata for deposit

## Catalyze?

What would it take to catalyze the DCP? (Besides a lot of time and money...) After we did the research project that resulted in the DCP, we chose to share it through workshops to show people what a data discussion might look like. Now we want to change stuff...

## What should the format of DCP become?

- Change it a smidge, something WordPress-y
- Not just an online form, something visual
- Just use Qualtrics or SurveyMonkey
- Database backend driven
- All webpage driven—xml and javascript
- Don't change, keep it like it is: all Word & PDF
- Why not do some with Google docs?

## Change?

How would you change the DCP? Can they be a multipurpose tool, should they be focused on a single outcome for one entity, or can there really be one ring to rule them all? Should there be one primary recipient of the outcome of using the DCP?

## Who should the DCP be for?

- Any individual who wants to fill one out
- A researcher seeking data curation ideas
- A researcher seeking very specific responses
- A librarian to facilitate a data discussion
- A data specialist diagnostic for a project
- An info pro who is a collaborator
- Library info science students to learn from

## DCP Toolkit accomplishments & achievements include:

- 14 Data Curation Profiles Toolkit Workshop delivered to 375 librarians—analysis of the workshops and participants' attitudes toward engaging in data discussion was published in *Journal of eScience Librarianship* 2013 <http://escholarship.umassmed.edu/jeslib/vol2/iss2/2/>
- 107 views of the Data Curation Profiles Symposium archive—the 2012 symposium included panels by library professionals who had used the Toolkit and published Profiles, as well as experts providing context for data curation <http://docs.lib.purdue.edu/dcpsymposium/>
- 32 Published Profiles in the Data Curation Profiles Directory—have been downloaded 4000+ times in the Directory's first year and used by librarians, information professionals, and others <http://docs.lib.purdue.edu/dcp/>
- 2013 ACRL STS Innovation Award—recognized the Distributed Data Curation Center for the development of the Data Curation Profiles Toolkit