# 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.

curation profile. Tows shaded in go		nated by the scientist to n	nake publicly avail	able are indicated by the
Data Stage	Output	Typical File Size	Format	Other / Notes Data resulting from a
Esploratory	Preiminary results of experiments		xis.	series of small experiments; often stored in a spreadsheet
Record- Keeping	Description of experiments I data	NA.	Lab notebook - print	The experiment, along with the data, is described and captured in a lab notebook. This includes pent-outs of results from scientific instruments.
Provensed	Consolidated files for statistical products		wh.	Data from experiments that show promise are recepanized into larger speardsheets and prepared for analysis.
Analyzed	Statistical analysis		No. 540	Depending on the nature of the statistical tests needed, the raw data is either analyzed in Excel SAS, or both.
Surrenninad	Results of the statistical analysis	Between 20-50 kb	vis	Data shown to be statistically significant.
Published	Graphs, charts and tables for inclusion in articles / output			The summarzed data is usually organized into a graph and submitted with sourced articles.
Note: The di shaded in gra not provide a	alla specifically designated by Empty cells represent response.	by the scientist to make cases in which information	publicly available on was not collecte	are indicated by the rows d or the scientist could
Target data for The target d statistically s	sharing lata for sharing is the si significant effect on the	memarized data – the o cell's capacity to defer	data that has bee nd against ceidal	e shown to have a
results. How	se of the data I notes that the data wi sever, he notes that if a the raw data might also	n individual is interesta		
digital data e not provide a	considered to be static	ed of multiple tabular fil asset but stated that the	les in Excel form average size of	as been collected. The at. The scientist could his Excel files range
The data ou	touts for this dataset in statics from the Excel s	clude Excel and SAS fi ext SAS files	ies as vel as gr	aphs incorporating

#### 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
000000000000
Create story/big picture about project's data
000000000000
•Gather specific info about data mgmt/workflow
000000000000
•Identify gaps/needs in data mgmt or curation
•Match curation services to researcher needs
000000000000
•Create data curation profiles for others to use
OCCOCOCOCOCOCO-Determine data curation metadata for deposit



. 8	econd 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, QB, etc.)
٠	What format(s) are the data in?
	(For example: MS Excel 2007, MySQL database, JPEG 2000 images, a raw data file from a Camobal CR10 data looser, etc.)

#### 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
000000000000
•Not just an online form, something visual
000000000000
<ul> <li>Just use Qualtrics or SurveyMonkey</li> </ul>
0000000000000
Database backend driven
OOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOOO
000000000000
•Don't change, keep it like it is: all Word & PD
000000000000
•Why not do some with Google docs?
- 200 200 200 200 200 200 200 200 200 20



#### Change?

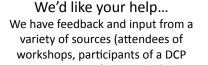
How would you change the DCP? Can they be a multipurpose tool, should they be focused on a single outcome for one entity, or can there really be one ring to rule them all? Should there be one primary recipient of the outcome of using the DCP?

## Who should the DCP be for?

•Any individual who wants to fill one out
000000000000
•A researcher seeking data curation ideas
000000000000
•A researcher seeking very specific responses
000000000000
•A librarian to facilitate a data discussion
000000000000
•A data specialist diagnostic for a project
000000000000
•An info pro who is a collaborator
000000000000
·Library info science students to learn from
- 200, 200, 200, 200, 200, 200, 200, 200

**PURDUE** 

LIBRARIES



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.

## 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







D. Scott Brandt Professor of Library Science Purdue University Libraries techman@purdue.edu datacurationprofiles.org