

Creating Data Profiles

NeDICC Workshop on The role of the information professional in relation to RDM, 22 June 2016, CSIR, Pretoria, South Africa

Presented by Johann van Wyk
(University of Pretoria)

Content/Programme

- 9:45 – 11:00
 - What is a Data Curation Profile?
 - How was the Data Curation Profile Developed?
 - Purpose and Use
 - Components of the Toolkit
 - User Guide
- 11:00-11:30 - TEA
- 11:30-13:00
 - Interviewer's Manual
 - Interview Worksheet
- 13:00-14:00 - LUNCH
- 14:00-15:45
 - Data Curation Profile Template
 - Publishing on Data Curation Profiles Directory (Website)

Introduction:

What is a Data Curation Profile?

“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” (Witt et al., 2009:3)

“A completed Data Curation Profile will contain two types of information about a data set:

- Information about the data set itself, including its current lifecycle, purpose, forms and perceived value.
- Information regarding a researcher’s need 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” (Witt et al., 2009: 3)

How was the Data Curation Profile Developed?

- 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
- Goals: 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.
- The Data Curation Profile Tool was developed as a key component of this research



Purpose and Use

- “The Data Curation Profile attempts to represent the perspective of the researcher or research group, rather, than those of the data curator, librarian archivist” (Witt et al., 2009: 4)
- A Data Curation Profile can be used:
 - At individual level
 - to provide a structure for conducting an interview between an information professional and researcher/research group
 - to provide a means for a researcher/research group to thoughtfully consider their needs for their data beyond its immediate use

Purpose and Use (2)

- A Data Curation Profile can be used
- At Institutional level:
 - to serve as foundational document to guide the management and/or curation of a particular data set
 - to share with staff providing data services and others to inform them and ensure that everyone is on the same page
 - to inform the development of data services to be offered by the institution, as well as to identify the types of tools, infrastructure and responsibilities of data services staff.
- At its broadest level: The 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 curate or preserve, and the needs of researchers in doing so.

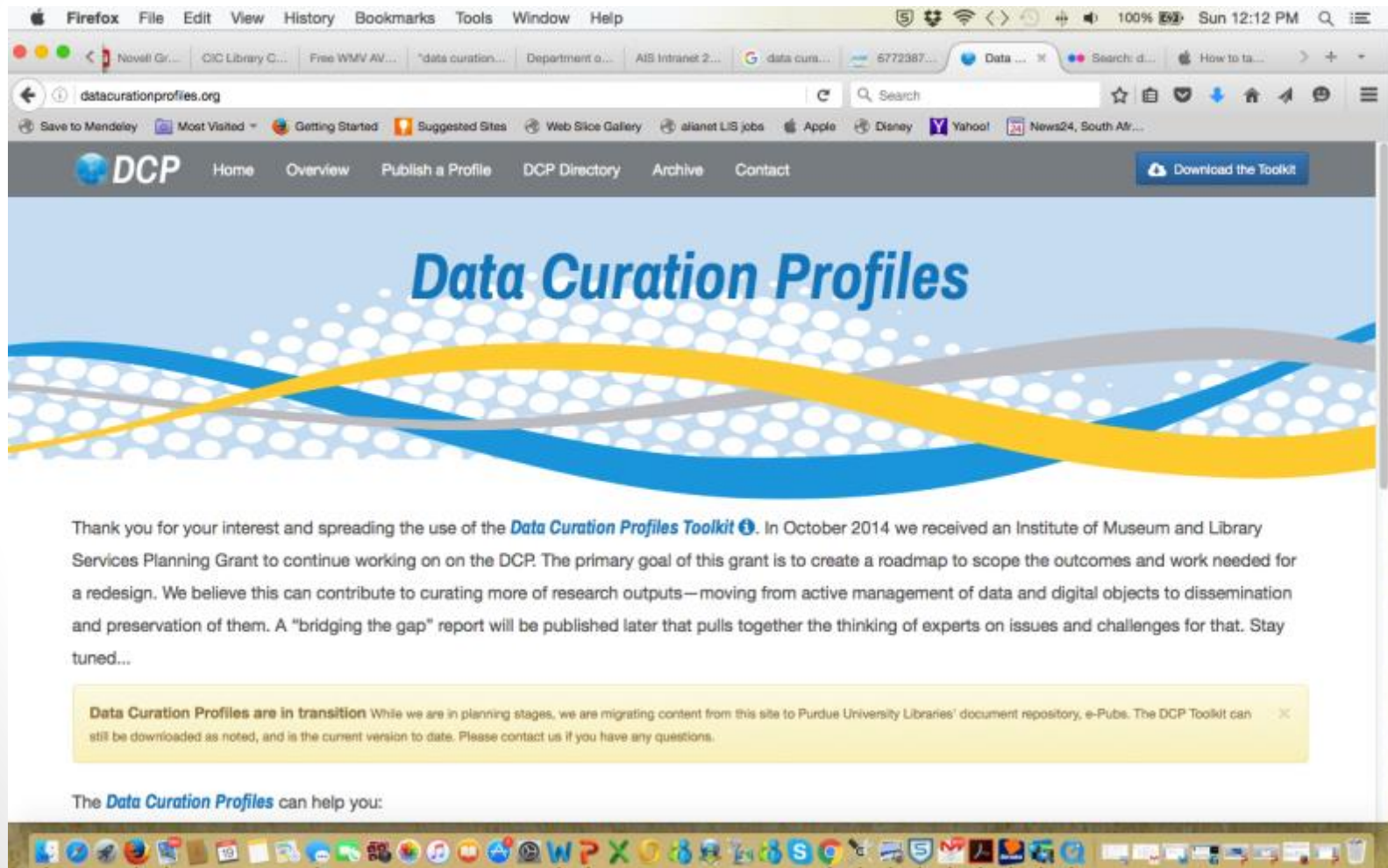
Purpose and Use (3)

Very Important:

A profile represents a snapshot of the data and researcher needs at a particular point in time

Data Curation Profiles Website

<http://datacurationprofiles.org/>



The screenshot shows a Firefox browser window displaying the Data Curation Profiles website. The browser's address bar shows the URL datacurationprofiles.org. The website's navigation menu includes links for Home, Overview, Publish a Profile, DCP Directory, Archive, and Contact, along with a "Download the Toolkit" button. The main heading reads "Data Curation Profiles" in a large, blue, serif font. Below the heading is a decorative graphic with wavy lines in blue, yellow, and grey over a background of white dots. A paragraph of text follows, thanking visitors for their interest and mentioning a grant received in October 2014. A yellow callout box contains a notice about the website's transition to a new repository. At the bottom, a section titled "The Data Curation Profiles can help you:" is partially visible.

Thank you for your interest and spreading the use of the [Data Curation Profiles Toolkit](#). In October 2014 we received an Institute of Museum and Library Services Planning Grant to continue working on on the DCP. The primary goal of this grant is to create a roadmap to scope the outcomes and work needed for a redesign. We believe this can contribute to curating more of research outputs—moving from active management of data and digital objects to dissemination and preservation of them. A “bridging the gap” report will be published later that pulls together the thinking of experts on issues and challenges for that. Stay tuned...

Data Curation Profiles are in transition While we are in planning stages, we are migrating content from this site to Purdue University Libraries’ document repository, e-Pubs. The DCP Toolkit can still be downloaded as noted, and is the current version to date. Please contact us if you have any questions.

The [Data Curation Profiles](#) can help you:

Components of the Data Curation Profiles Toolkit

- User Guide
- Interviewer's Manual
- Interview Worksheet
- The Profile Template
- The Profile

User Guide

The User Guide provides basic information about the Data Curation Profiles, and directions on how to construct a Data Curation Profile

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

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.

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.

The Interview Worksheet also serves as a source of information to populate the Data Curation Profile Template

Data Curation Profiles Toolkit

- User Guide

- Part 1 – Background
- Part 2 – Purpose and Use
- Part 3 – Components of the Data Curation Profile Toolkit
- Part 4 – How to develop a Data Curation Profile
 - Stage 1 – Preparation
 - Stage 2 – Interviews
 - Stage 3 – Constructing the Data Curation Profile
- Acknowledgements
- Checklist of Activities

Part 4 – Stage 1: Preparation

- Identify the data that will serve as the focus of the interviews
- Criteria for selecting the data to discuss in the interviews will vary depending on the researcher and his/her situation, and should be negotiated between the interviewer and participating researcher. This should preferably be data that best represent the typical data that are generated by the researcher.
- Once data has been selected, do some preliminary investigations about the work being done by the researcher

Part 4 – Stage 1: Preparation (2)

- Determine who takes the responsibilities for:
 - conducting the interviews,
 - transcribing the the interviews, and
 - drafting the Data Curation Profile
- Ethical Clearance
 - If you plan to use the Data Curation Profile for research purposes , including presenting or publishing the content from the interviews or the Profile, you would need to get ethical approval through your institution's ethical committee(s)

Part 4 – Stage 1: Preparation (3)

- **Modifications**

- The Data Curation Profile tool provides some flexibility in its application and can be modified if needed.
- The Profile is modular in nature.
- The designers of the Profile recommend that all sections be included, but some pieces of the profile can be removed or replaced with locally generated sections if desired.

- **Time needed**

- The time needed to complete the preparation stage will vary depending on the the amount of background research needed, the extent of modifications made to the interview documents, and if a review by an Ethics Committee is needed. Approximately 1-2 hours would be needed to prepare.

Part 4 – Stage 2: Interviews

- Conducting the Interviews:
 - The Profile Toolkit contains an Interview Manual and Interview Worksheet that should be used in conjunction with each other.
 - At the start of the interview give the participating researcher the Interview Worksheet, then open the Interview Manual and read the Introduction to the Interview Worksheet statement.
 - Alternatively, give the researcher the Interview Worksheet beforehand to complete and prepare for the interview. However, the researcher may have questions or need direction to fill out the worksheet. It is therefore better to do this during the interview.

Part 4 – Stage 2: Interviews (2)

- After the introductory statement there are several Background / Demographic Questions contained in the Interviewer's Manual that should be asked
- From this point onwards , both the Interview Worksheet and Interviewer's Manual are organised by Module.
- At the beginning of each Module:
 - Read an introductory statement listed in the Interviewer's Manual (Please note not all modules have an introductory statement)
 - Then have the researcher complete the list of questions for the module in the Interview Worksheet.
 - Review the researcher's responses in the worksheet and ask the additional questions listed in the Interviewer's Manual as appropriate
 - Ask follow up questions if needed for clarification
 - Make sure that all the questions have been answered before moving to the next module

Part 4 – Stage 2: Interviews (3)

- The Need for Two Interviews
 - The designers of the Data Curation Profiles recommend that two interviews be held with the researcher, because of the amount of information needed to complete these, and to prevent both interviewer and researcher from growing fatigued.
 - Each interview should take about 1 hour minimum.
 - The first interview should at least cover the following modules:
 - Module 1 – The Data Set
 - Module 2 – The Lifecycle of the Data
 - Module 3 – Sharing
 - Module 4 - Access
- The Second Interview:
 - This interview should be conducted in the same manner as the first interview.
 - Follow-up questions resulting from a review of the first interview should be incorporated in this interview

Part 4 – Stage 3: Constructing the Profile

The final stage stage in the process is to take the information gathered from the interviews and to complete the sections and subsections of the Data Curation Profile template

Interviewer's Manual & Interview Worksheet

Interview: Modules

Elements/Module	Theme
Background / Demographic Questions	
Module 1	The Data Set
Module 2	The Lifecycle of the Data Set
Module 3	Sharing
Module 4	Access
Module 5	Transfer of Data / Ingest into a Repository
Module 6	Organisation and Description of Data
Module 7	Discovery
Module 8	Intellectual Property
Module 9	Tools
Module 10	Linking /Interoperability
Module 11	Measuring Impact
Module 12	Data Management
Module 13	Data Preservation

Data Curation Profiles Template

Relationship between Interview Modules and Data Curation Profile

Interview Module	Data Curation Profile Section
Background / Demographic Questions	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 and Access
Module 4 - Access	
Module 5 – Transfer of Data /Ingest into a Repository	Section 6 – Ingest / Transfer
Module 6 – Organisation and Description of Data	Section 5 – Organisation and Description of Data (including metadata)
Module 7 – Discovery	Section 8 – Discovery

Relationship between Interview Modules and Data Curation Profile (2)

Interview Module	Data Curation Profile Section
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

Publishing in the Data Curation Profiles Directory

<http://datacurationprofiles.org/>

Data Curation Profiles

Thank you for your interest and spreading the use of the [Data Curation Profiles Toolkit](#). In October 2014 we received an Institute of Museum and Library Services Planning Grant to continue working on on the DCP. The primary goal of this grant is to create a roadmap to scope the outcomes and work needed for a redesign. We believe this can contribute to curating more of research outputs—moving from active management of data and digital objects to dissemination and preservation of them. A “bridging the gap” report will be published later that pulls together the thinking of experts on issues and challenges for that. Stay tuned...

Data Curation Profiles are in transition While we are in planning stages, we are migrating content from this site to Purdue University Libraries' document repository, e-Pubs. The DCP Toolkit can still be downloaded as noted, and is the current version to date. Please contact us if you have any questions.

The [Data Curation Profiles](#) can help you:

Data Curation Profiles Directory

Home About **My Account**

[Home](#) > [Data Curation Profiles](#) > [DCP](#)

Editors

Jake Carlson, University of Michigan
D. Scott Brandt, Purdue University

Follow

A **Data Curation Profile** is a resource for Library and Information Science professionals, Archivists, IT professionals, Data Managers, and others who want information about the specific data generated and used in research areas and sub-disciplines that may be published, shared, and preserved for re-use. Data Curation Profiles capture requirements for a specific data set generated by a single scientist or lab, based on their reported their needs and preferences for the data. More information about the Data Curation Profiles Project is available at <http://datacurationprofiles.org>

The **Data Curation Profiles Directory** provides a stable, citable home for completed Data Curation Profiles. Each completed profile is subjected to review before being formatted, published with a DOI, and distributed through the information supply chain. More information about this process can be found in the "About" section of this site.

[Publication Home](#)
[About this Publication](#)
[Aims & Scope](#)
[Editorial Board](#)
[Guidelines for Authors](#)

[Upload Profile](#)

[Most Popular Profiles](#)

[Receive Custom Email Notices or RSS](#)

Select a year:

All Years

Enter search terms:

in this publication

[Advanced Search](#)

Data Curation Profiles Directory

Home About My Account

Home > [Data Curation Profiles](#) > [DCP](#)

Login

Email address:

Password:

Remember me

Login

[Forget your password?](#)

New Email Address? Please log in and choose [Edit Profile](#) on your [My Account](#) page to update your contact information or customize your password.

Create new account

You will need to create an account to complete your request. It's fast and free.

Sign up

Publication Home

About this Publication

Aims & Scope

Editorial Board

Guidelines for Authors

Upload Profile

Most Popular Profiles

Receive Custom Email Notices or [RSS](#)

Select a year:

All Years

Enter search terms:

in this publication

[Home](#) > [Data Curation Profiles](#) > [DCP](#)

Please complete this form to create your user account. Use proper casing for names and institution as these fields may be displayed for publication purposes. Bepress neither sells nor rents contact information to third parties.

You will receive an email confirmation shortly.

Required fields are marked with an *:

Email address *

First/Given Name *

Middle Initial

Last/Family Name *

Suffix

Institutional Affiliation **Not on list? Please enter name in full.**

Password *

Re-enter Password *

- [Publication Home](#)
- [About this Publication](#)
- [Aims & Scope](#)
- [Editorial Board](#)
- [Guidelines for Authors](#)

[Upload Profile](#)

[Most Popular Profiles](#)

[Receive Custom Email Notices or RSS](#)

Select a year:

Enter search terms:

[Advanced Search](#)

ISSN : 2326-6651

PURDUE
UNIVERSITY
LIBRARIES



Purdue e-Pubs

[Home](#) [About](#) [FAQ](#) [My Account](#)

Enter search terms:

Search

in this repository

[Advanced Search](#)

Notify me via email or [RSS](#)

Links for Authors

[Submit Research](#)

[Author Addendum](#)

Links

[Purdue Libraries](#)

[Purdue University Press Journals](#)

Browse

[Collections](#)

[Disciplines](#)

[Authors](#)

My Account

[Log Out](#)

BAREND J VAN WYK
<JOHANN.VANWYK@UP.AC.ZA>

ACCOUNT SETTINGS

Edit Profile	Change password, affiliation, or contact information
Research Alerts	Change e-mail notifications and manage e-mail preferences

[Home](#) > [Data Curation Profiles](#) > [DCP](#)

Follow the instructions below to submit a manuscript to the Data Curation Profiles Directory. Shortly afterward, you will receive an email confirming your submission.

The submission process consists of the following steps:

1. Read and accept the Submission Agreement
2. Provide information about yourself
3. Provide information about any authors
4. Upload your submission and related items

Before you begin, please be sure you have the following items:

- **Submission Title** (Follow format "Lead Author_Date of Profile Completion")
- An **abstract** (separate from the article body)
- **Submission in one of the following formats: Microsoft Word, or RTF.**

Articles must be submitted without a title page, abstract, or page numbers. These will be provided by the system.

No part of the submission is final until all steps have been completed and you click the final **Submit** button. The review process begins as soon as Data Curation Profiles Directory directory receives a readable submission, along with the abstract and submission title. You may revise any of these elements later by clicking the submission's title on your **My Account** page.

[Continue](#)



[Publication Home](#)
[About this Publication](#)
[Aims & Scope](#)
[Editorial Board](#)
[Guidelines for Authors](#)

[Upload Profile](#)

[Most Popular Profiles](#)

[Receive Custom Email Notices or RSS](#)

Select a year:

Enter search terms:

[Advanced Search](#)

ISSN: 2326-6651

PURDUE
UNIVERSITY
LIBRARIES

1 Submission Agreement

Article Submission Agreement for Data Curation Profiles Directory

Warranties & Ownership Overview

Please review the following submission agreement and indicate your agreement below.

I understand that the submission will be reviewed using the criteria described in the [Policies](#) section and that, based on that review, it may be accepted, accepted on condition of revision, or declined for publication.

If the submission is accepted for publication, I hereby grant to Purdue University a non-exclusive perpetual royalty-free license to use, duplicate, and distribute it in whole or in part through the Data Curation Profile Directory and/or the [datacurationprofiles.org](#) website.

I further grant to Purdue University the right to transfer the submission to any format or medium now known or later developed for preservation and access in accordance with this agreement.

This agreement does not represent a transfer of copyright to Purdue University and the author reserves all other rights.

I represent and warrant to Purdue University that the submission is my original work and does not, to the best of my knowledge, infringe or violate any rights of others nor does the deposit violate any applicable laws.

I further represent and warrant that I have the authority and/or have obtained all necessary rights (including the express permission of the interviewee) to permit Purdue University to use, duplicate, and distribute the submission and that any third-party content is clearly identified and acknowledged within the Work.

By clicking **Accept** you agree to grant this non-exclusive licence and acknowledge that you have read and agreed to the terms of this agreement and all related [publication policies](#).

Please click **Accept** if you accept the policies and terms of this agreement, and continue with the next step in the submission process.

- Aims & Scope
- Editorial Board
- Guidelines for Authors
- Upload Profile
- Most Popular Profiles
- Receive Custom Email Notices or [RSS](#)

Select a year:

Enter search terms:

[Advanced Search](#)

ISSN: 2326-6651



2 Provide Information about Yourself

about you...

Please make sure that your contact information is complete and correct, as the editors of the journal may need to contact you regarding your submission.

Please be sure to enter your name and affiliation exactly as they should appear on a published article (including appropriate capitalization and punctuation).

Required fields are marked with an *.

First/Given Name *

Middle Initial

Last/Family Name *

Suffix (e.g. "Jr.")

Institutional Affiliation * (e.g. "University of California, Berkeley")

Phone Number

Fax Number

Postal Address (line 1)

Postal Address (line 2)

Postal Address (line 3)

- About this Publication
- Aims & Scope
- Editorial Board
- Guidelines for Authors

- Upload Profile
- Most Popular Profiles
- Receive Custom Email Notices or RSS

Select a year:

All Years

Enter search terms:

in this publication

Advanced Search

ISSN: 2326-6651





Home About My Account

Home > Data Curation Profiles > DCP

3 Provide Information about Any Co-Authors

Please click Add author to enter each co-author. This will be the final author list that appears on your publication. When you have finished, press Continue.

Add Authors

If no email address is entered the author will not be notified about the status of the submission or be able to access or change any information regarding this submission using these web tools.

Author's email address (optional): Add author

When the list below is accurate and complete, click Continue.

Name/Email	Institution	Order
Barend J. Van Wyk <johann.vanwyk@up.ac.za>	University of Pretoria	1st author

Continue

- Publication Home
- About this Publication
- Aims & Scope
- Editorial Board
- Guidelines for Authors

- Upload Profile
- Most Popular Profiles
- Receive Custom Email Notices or RSS

Select a year: All Years Browse

Enter search terms: Search

in this publication

Advanced Search

ISSN: 2326-6651





Home > Data Curation Profiles > DCP

4 Upload Your Manuscript and Related Items

authors...
(click the back button to change)

Name/Email	Institution	Order
Barend J. Van Wyk <johann.vanwyk@up.ac.za>	University of Pretoria	1st author

about your Submission...

Submission Title

Please use the name of the lead author followed by date of completion, for example "Palmer_08212012"

Enter your submission title:

NCES Discipline

Please visit the [NCES classification list](#), find the appropriate disciplinary code and enter it in the text box below.

- Publication Home
- About this Publication
- Aims & Scope
- Editorial Board
- Guidelines for Authors

- Upload Profile
- Most Popular Profiles
- Receive Custom Email Notices or RSS

Select a year:
All Years

Enter search terms:

In this publication

Advanced Search

ISSN: 2326-6551



Sub-discipline

This should be a descriptive term of the author's choosing that best describes the research/dataset being profiled.

Publication Date

Month	Day	Year
..	..	

Abstract

The abstract for a data curation profile should briefly describe the data set being profiled and include a concise statement as to the data needs articulated by the interviewees. This information is typically found in 'Section 2.1 - Research Area Focus' and in 'Section 1 - Brief Summary of Data Curation Needs'. The information taken from these sections may need to be rewritten so that they fit together as an abstract.

Type or paste your abstract:

B I | | | | | | | | | | | | | | | | | |



DOI

Full Text of Submission

Please upload your primary submission file. If your file is a Microsoft Word or RTF Document, the system will automatically generate a PDF for you.

- Upload file from your computer
- Import file from remote site

Message to the Editor

This will not appear on the final publication. It will just be available to the editors.

Browse... No file selected.

Additional Files

Please check this if you'd like to add additional files

Submit: Completing your upload may take some time; please only click on the submit button once.

Thank You

List of Sources

- Carlson, J. 2010. The Data Curation Profiles Toolkit: User Guide (Data Curation Profiles Toolkit, Paper 1).
<http://dx.doi.org/10.5703/1288284315650>
- Carlson, J. 2010. The Data Curation Profiles Toolkit: Interviewer's Manual. (Data Curation Profiles Toolkit, Paper 2).
<http://dx.doi.org/10.5703/1288284315651>
- Carlson, J. 2010. The Data Curation Profiles Toolkit: Interview Worksheet. (Data Curation Profiles Toolkit, Paper 3).
<http://dx.doi.org/10.5703/1288284315652>
- Carlson, J. 2010. The Data Curation Profiles Toolkit: the Profile Template. (Data Curation Profiles Toolkit, Paper 4).
<http://dx.doi.org/10.5703/1288284315653>