

Published on *Data Management Plan Tutorial* (<https://www.e-education.psu.edu/dmpt>)

[Home](#) > [Course Outline](#) > Part 2: Documenting the Data

## Part 2: Documenting the Data

SCENARIO



### Kim Learns about Metadata

Kim has now taken stock of Professor Smart's data, noting in appropriate detail its types and formats, as well as documenting the size of it and estimating its rate of growth. Familiarizing herself in this way with the data, she's now ready to think about how to describe what it is and investigate what the community of researchers working with similar data use to describe it. She's heard of “metadata” but is not quite sure what it means or why she might need it.

In this part, you will understand the value of metadata and metadata standards, including the following:

- Determine what metadata are required (what information about your data is important to document?)
- Document how metadata will be recorded (who's going to do it? where? with what frequency?)
- Review directory and file-naming conventions

### Metadata

This file contains additional information, probably added from the digital camera or scanner used to create or digitize it. If the file has been modified from its original state, some details may not fully reflect the modified image.

Camera manufacturer	Canon
Camera model	Canon EOS 30D
Author	unknown
Exposure time	1/160 sec (0.00625)
F Number	f/5.6
Date and time of data generation	20:21, 20 January 2005
Lens focal length	400 mm
<a href="#">Show extended details</a>	

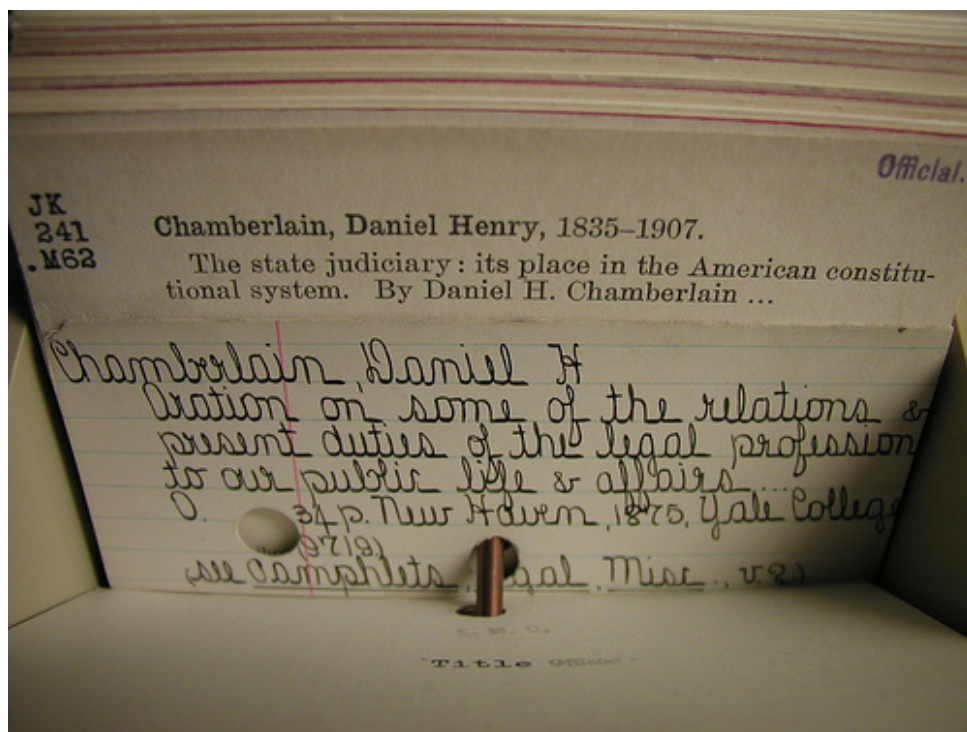
Table showing metadata ("data about data") for a photograph (not pictured).  
Credit: [Wikimedia Commons](#) <sup>[1]</sup>

## 2.1 What is Metadata?

Further documentation and description about your data are needed beyond details about its types and formats. Such description, also known as “metadata,” answers the following questions about your data:

- Who created it?
- What is it?
- When was it created?
- How was it generated?
- Where was it created?
- How may it be used?
- Are there restrictions on it?

Metadata is, literally, “data about data” - meaning, it is information about a resource. Any answers to the above questions would then serve as metadata.



Card catalogs used to be how information about books and journals, or metadata, were organized and displayed.

Credit: [dfulmer on Flickr](#) [2]. CC-BY 2.0.

In the context of a DMP, the resource being described and managed is data. In fact, documenting the types and formats of your data, as described earlier, could also be seen as capturing metadata.

Examples of metadata elements that may be used to describe data:

- Title of the project
- Names of researchers involved
- Abstract or summary about the project and the data
- Research topic, or subject of research
- Temporal coverage / information
- Spatial coordinates / Location information
- Instrumentation used
- Access or rights policies/restrictions
- Hardware and/or software used

Metadata also helps organize information about a resource and give it a structure, which enhances search and discovery of data, as well as enables easier access to it.

Your community of interest (research community) may have recommendations as to what kind of metadata to capture. The next section gives some guidance on where to obtain such information.



## Data Description

[Show Map](#) [Google Earth](#) [RIS](#) [BisTrX](#)

**Citation:** Nemirovskaya, IA; Chemyavskii, NG (2011): Concentration of organic compounds in slush, under-ice water, and the snow-ice fast-ice cover of the southeast sector of the Antarctic in April 2008.  
doi:10.1594/PANGAEA.769949,  
*Supplement to: Nemirovskaya, Inna A; Chemyavskii, NG (2011): Studies of hydrocarbons in the waters and snow-ice cover of the southeast sector of the Antarctic. Translated from Okeanologiya, 2011, 51(1), 5-15, Oceanology, 51(1), 4-14, doi:10.1134/S0001437011010139*

**Abstract:** Data are presented on concentration of hydrocarbons (HC) relative to concentrations of suspended matter, lipids, organic carbon, and chlorophyll a in surface waters and snow-ice cover of the East Antarctic coastal areas. It was shown that growth of concentrations of aliphatic HC (AHC) to 30 µg/l in surface waters takes place in frontal zones and under young ice formation. AHC concentration in snow increases with growth of aerosol concentration in the atmosphere. In the lower part of ice, at the boundary with seawater, despite low temperatures, autochthonous processes may provide high AHC concentrations (up to 289 µg/l). Within the snow-ice cover on fast ice, concentration co-variations of all the compounds considered take place.

**Project(s):** [Archive of Ocean Data \(ARCOD\)](#) 🔍

**Coverage:** *Median Latitude: -68.894167 \* Median Longitude: 62.978333 \* South-bound Latitude: -69.810000 \* West-bound Longitude: 10.580000 \* North-bound Latitude: -66.533333 \* East-bound Longitude: 92.900000*

*Date/Time Start: 2008-04-04T00:00:00 \* Date/Time End: 2008-04-12T00:00:00*

**Event(s):** [AF2008-NLS](#) 🔍 \* *Latitude: -69.810000 \* Longitude: 10.580000 \* Date/Time: 2008-03-25T00:00:00 \* Location: Southern Ocean - Atlantic Sector 🔍 \* Campaign: AMK50 🔍 \* Basis: Akademik Mstislav Keldysh 🔍 \* Device: Multiple investigations 🔍*

[DruzhnSt-2008](#) 🔍 \* *Latitude: -69.616667 \* Longitude: 74.216667 \* Date/Time: 2008-04-06T00:00:00 \* Location: Prydz Bay, Collaboration Sea 🔍 \* Campaign: AMK50 🔍 \* Basis: Akademik Mstislav Keldysh 🔍 \* Device: Multiple investigations 🔍*

[MirnySt-2008](#) 🔍 \* *Latitude: -66.533333 \* Longitude: 92.900000 \* Date/Time: 2008-04-12T00:00:00 \* Location: Mirny Station 🔍 \* Campaign: AMK50 🔍 \* Basis: Akademik Mstislav Keldysh 🔍 \* Device: Multiple investigations 🔍*



**License:** Creative Commons Attribution 3.0 Unported

**Size:** 2 datasets



An Example of metadata for an item in [Pangaea: Data Publisher for Earth and Environmental Science](#).  
Credit: [Pangaea](#) [3]

## 2.2 The Importance of Standards in Metadata

In documenting your data with metadata, it is important to adhere to standards - standard vocabularies, standard schemas, etc. Organizing your data using a standards-based approach helps ensure interoperability between systems, which also enhances discovery of, and access to, data. To Ben Goldman, metadata is an essential tool for managing collections, and adherence to common standards is vital to making the data reusable. He shares some of his experiences with trying to decipher old data in the video below.

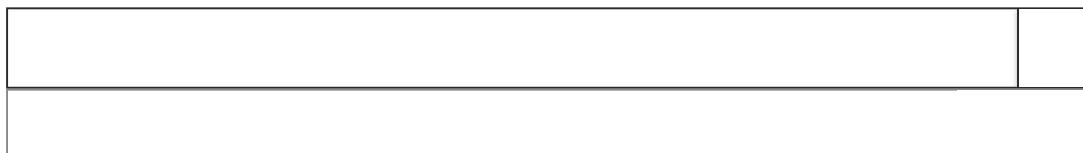
[Link to YouTube video.](#) [4]

One common standard or schema followed in many repositories is the [Dublin Core Metadata Element Set](#) [5], which has 15 metadata fields (e.g., creator, title, subject, rights, description, etc.). Your research community may already have metadata standards that it follows. A useful list of disciplinary standards to consult is at the [Digital Curation Centre \(DCC\)](#) [6] website. Another way to find out more about metadata standards for your research data is to consult the data repositories that your data might be suitable for. Two registries of repositories, in particular, are worthwhile checking out:

- [DataBib](#) [7] is searchable catalog / registry / directory / bibliography of research data repositories."
- [DataCite Repositories List](#) [8] - Working document listing data repositories.

You might also consult the metadata librarian at your institution's library to find out more about what standards to apply to your data. Ultimately, an exploration of these standards can provide insight into what kind of information about your data is important to document.

Dr. Alfred Traverse, Curator of the Penn State Herbarium, has extensive experience in managing large of collections of specimens with varying levels of metadata. In the video below, he describes two specimens from the Herbarium's collection, how they were preserved, and their recorded metadata. It is a good illustration of how metadata standards change over time. Standards evolve much as data collection practices do.

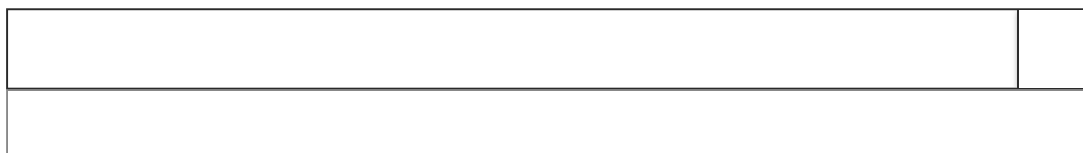


[Link to YouTube video.](#) <sup>[9]</sup>

## 2.3 Recording the Metadata

Additional information you may wish to include in a DMP is a brief statement on who will be responsible for recording metadata and having oversight of it, as well as how metadata will be recorded (e.g., in a spreadsheet, or through some other means - perhaps automated).

At the Herbarium, Dr. Traverse copes daily with challenges related to file names and limitations on how the metadata for the specimens was and is recorded. In the following video, he describes some of these challenges as well as how advancements in science can also impact metadata.

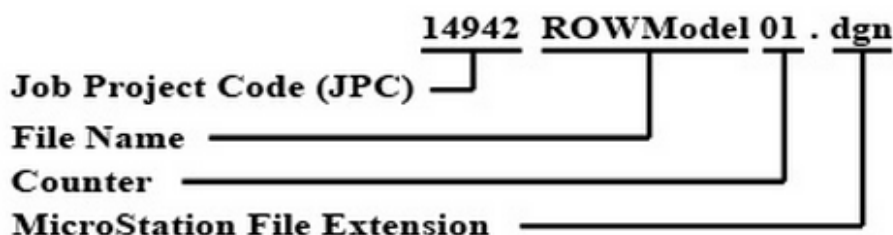


[Link to YouTube video.](#) <sup>[10]</sup>

One aspect of recording or documenting metadata is the creation of file names - particularly since data is often kept in file directory systems. Think, for example, of the files you maintain in a hard drive, where files are kept in folders, which are part of a hierarchical structure, or directory, in that drive.

### CDOT File Naming Structure

#### Existing Right-Of-Way project deliverable sheet



An Example of file naming guidance.



Credit: [Colorado Department of Transportation](#) <sup>[11]</sup>

Although file-naming conventions do not have to be discussed in a DMP, it is a key part of data management planning in your local context. When naming your files, think about using descriptive labels that have meaning to others besides yourself, especially if these are data files to be shared with a team. Besides being descriptive, file names should follow a pattern. For example, when including a date and/or year, make sure it follows a pattern like YYYYMMDD. Think about coming up with models for file names - e.g.: **project-name\_content-of-file\_date.file-format**. File names should also reflect the version of the file. Versioning helps track the history and progress of data collection and is crucial to collaboration when more than one person needs access to a data file at any given time.

Other file-naming tips:

- Do not use spaces in file names
- Do not use uncommon characters (e.g., #, ?, &, !, etc.)
- Stay within a limit of 32 characters (the more concise yet descriptive, the better)

▼ SIG	Today, 5:29 PM	--	Folder
acrl_board_action_form_20110216.doc	Feb 16, 2011 2:12 PM	217 KB	Micro...ument
acrl_board_action_form.doc	Dec 9, 2010 8:33 PM	74 KB	Micro...ument
acrl_ig_emails.docx	Nov 28, 2010 12:30 PM	168 KB	Micro...ument
acrl_ig_response_emails.docx	Nov 28, 2010 6:51 PM	98 KB	Micro...ument
▼ Board_Action_Forms	Today, 5:52 PM	--	Folder
dcig_final_version.doc	Dec 2, 2010 9:31 PM	90 KB	Micro...ument
dcig_v20101105.doc	Dec 9, 2010 8:40 PM	66 KB	Micro...ument
dcig_v20101122.doc	Dec 9, 2010 8:41 PM	70 KB	Micro...ument
dcig_v20101129.doc	Nov 29, 2010 12:09 AM	111 KB	Micro...ument
ig_petition_20100902.pdf	Sep 2, 2010 5:01 PM	74 KB	Adobe...ument
ig_petition_dc_20101020.pdf	Nov 5, 2010 3:41 PM	119 KB	Adobe...ument
ig_petition_dc_20101105.pdf	Dec 9, 2010 8:13 PM	119 KB	Adobe...ument
ig_petition_dc_final.pdf	Dec 2, 2010 9:30 PM	119 KB	Adobe...ument
ig_petition_ir_v2.pdf	Sep 7, 2010 4:28 PM	86 KB	Adobe...ument
ig_petition.pdf	Aug 30, 2010 5:43 PM	82 KB	Adobe...ument
mr_dcig_v20101129.doc	Dec 9, 2010 8:03 PM	102 KB	Micro...ument
ph_acrl_board-action-form_ig_charge.doc	Apr 4, 2011 1:29 PM	74 KB	Micro...ument
ph_dcig_v20101201.doc	Dec 1, 2010 3:38 PM	106 KB	Micro...ument
ebss.docx	Dec 9, 2010 8:22 PM	164 KB	Micro...ument

An Example of a directory with files showing file names following an agreed-upon convention.

Credit: Patricia Hswe

## 2.4 Summary

In this part, you learned the meaning of "metadata" and its important role in the documentation of data. In the DMP, in addition to noting types and formats the data will take, consider providing examples of the information you will capture about your data (e.g., information about location, instrumentation, rights and access, etc.). Consult resources such as the Digital Curation Centre's [disciplinary standards list](#) <sup>[12]</sup> to find out what standards for describing data your research community follows.

One common metadata schema that is discipline agnostic is the Dublin Core metadata element set. It's strongly advised to consult a metadata librarian, who can assist you in data description. A metadata librarian can also provide guidance on approaches to recording metadata, such as how to name files. A detail such as file naming convention is not necessary to include in a DMP, but you and your project team should decide as early as possible what conventions the team will follow for naming files.

Another resource worthwhile checking out are data repository indexes, such as [DataBib](#) [7], which you can use to find a repository suitable for your data when your project is ready to deposit them. Seeking out an appropriate repository ahead of time and learning about its requirements can inform how you describe your data.



## Check Your Understanding

You are collecting survey data and have a field in your data set for geographic location. Some of your responses include both city and state; others include only the country; some did not respond at all. You've developed a codebook to describe how you've chosen to record this and other data from your survey. Is it appropriate to include the codebook in your metadata?

- (a) Yes
- (b) No



Click for answer.

**ANSWER:** (a) Yes. Codebooks are considered part of the metadata, since they document and describe aspects of the data you have collected. Without a codebook, users of your data, such as other researchers, will not be able to understand your data readily, or at all, making reuse of the data and the ability to build on it highly unlikely. So, if you have developed a codebook as part of documenting your research data, then by all means include it as part of the metadata for your data.

**Source URL:** <https://www.e-education.psu.edu/dmpt/node/659>

### Links

- [1] [http://commons.wikimedia.org/wiki/File%3AMetadata\\_screenshot.png](http://commons.wikimedia.org/wiki/File%3AMetadata_screenshot.png)
- [2] <http://www.flickr.com/photos/annarbor/4350624388/in/photostream/>
- [3] <http://www.pangaea.de/>
- [4] <http://www.youtube.com/watch?v=GgtrmMyF0DE>
- [5] <http://dublincore.org/documents/dces/>
- [6] <http://www.dcc.ac.uk/resources/briefing-papers/standards-watch-papers/what-are-metadata-standards>
- [7] <http://databib.org/>
- [8] <http://www.datacite.org/repolist>
- [9] <http://www.youtube.com/watch?v=mCTea-r3z2o>
- [10] <http://www.youtube.com/watch?v=OCi6Bwq3QCU>
- [11] [http://www.coloradodot.info/content/business/designsupport/CADDmanual/chapter\\_04\\_-\\_file\\_naming\\_structure.htm](http://www.coloradodot.info/content/business/designsupport/CADDmanual/chapter_04_-_file_naming_structure.htm)
- [12] <http://www.dcc.ac.uk/resources/metadata-standards>