Data Management Workshop Series

Session 2: How to Create Data That You (and Others) Can Understand

Handout: Data Documentation in Practice

Approach	Problem/Question	Examples
Lab Notebook (or an ELN)	Lots of day-to-day changes and variations in collection methods or conditions?	LabArchives.com, ResearchSpace.com
Data Dictionary or Codebook	Many codes/variables used in your data files?	SPSS , R (Hmisc, Swieve addons)
Structured Documentation File	Collecting lots of data files of the same type?	Create a Google Template
Blog/Wiki Tool	Want to keep track of your research but also engage the community?	UThink, WordPress
Unstructured Documentation File	Just need to document the basics of your process?	Readme.txt
Instrument Log	Have an instrument that produces your data?	Digital cameras
Track Changes or Revision History	Do you create textual data?	MS Word or Google Docs
Version Control	Do you write software code?	GItHub, Subversion, eg TortoiseSVN
Reference to an established protocol	Is there a well known way of creating your data?	Chemistry Protocols

Data Management Workshop Series

Session 2: How to Create Data That You (and Others) Can Understand

Exercise #3: Fill out a Structured Form using the Dublin Core (dc) metadata standard

Record Creation Date: 2013-10-18

Title Local business perspectives of the proposed monorail system in the Metro Area, 2012-13

Titles include the what, where, when, who and scale of your data.

Creator _____

Use complete names eg. "James M. Swanson."

Contributor _____

List any contributing authors of the data.

Subject Terms_____

Separate with ",". Keyword help others locate your data via search.

Coverage:

Spatial _____

Eg. Minnesota, USA.

Temporal _____

Period of time, eg. 2006-2009.

Abstract

_____

Summarize the dataset, describe how it will meet a need.

Date _____

Date the data was created or gathered. YYYY-MM-DD format.

Type _____

This is the nature of the data, such as "Image" "Text" or "Software". Separate multiple types with a ",".

Format _____

The file extension works well here.

Description

_____

Describe how (multiple) files relate to each other (eg. documentation files vs dataset).

Identifier _____

Not all data will have an identifier. This could be an ISSN, DOI, or URL.

Relation _____

Track versions of your data to other versions (eg. Raw, Processed.)

License _____

Applying a Creative Commons CCO license to your data might help facilitate attribution.

Data Management Workshop Series

Session 2: How to Create Data That You (and Others) Can Understand