Data Management Plans for the Deep Carbon Observatory: What, Why and How?

Xiaogang (Marshall) Ma (max7@rpi.edu), Yu Chen (cheny18@rpi.edu), Han Wang (wangh17@rpi.edu), Patrick West (westp@rpi.edu), John Erickson (erickj4@rpi.edu), Peter Fox (foxp@rpi.edu)

Tetherless World Constellation, Rensselaer Polytechnic Institute, 110 8th St., Troy, NY 12180 United States

Motivation

Leading up to the decadal lifetime of the Deep Carbon Observatory (DCO), DCO-Data Science (DCO-DS) will assemble the **Deep Earth Computer**; creating a fundamental change in the conduct of Carbon-related research, resting upon a 21st century data science platform, and a series of aggregate data holdings that have never existed before. Data science combines aspects of informatics, data management, library science, computer science and physical science using supporting cyberinfrastructure and information technology. For DCO-DS, an essential work to do is the Data Management Plan for research projects within the DCO community. The Data Management Plan will be a primary driving force that promotes the data science platform of DCO.

Key: Well described | Somewhat described | Not described; Required | Suggested | Omitted

Research Funders	National Science Foundation (NSF)	NSF Basic Research to Enable Agricultural Development (BREAD)	NSF Division of Earth Sciences (EAR)	NSF Division of Ocean Sciences	NSF Integrated Ocean Drilling Program	NSF Ocean Acidification Reaserach	NSF Office of Polar Programs	NSF Engineering Directorate	NSF Social Behavioral and Economic Sciences	National Institutes of Health (NIH)	NIH - Genome-Wide Association Studies (GWAS)	NIH - National Human Genome Research Institute	United States Department of Agriculture (USDA)	United States Department of Energy (DOE)	DOE Atmospheric Radiation Measurement Program (ARM)	United States Department of Education (DoEd)	United States Environmental Protection Agency (EPA)	United States Agency for International Development (USAID)	National Aeronautics and Space Administration (NASA) - Heliophysics	National Aeronautics and Space Administration (NASA) – Earth Sciences	Office of Naval Research (ONR)	Office of Naval Research Policy for In Situ Ocean Data (ONR)	National Oceanographic and Atmospheric Administration (NOAA) Climate Observations and Monitoring (COM)	National Oceanographic and Atmospheric Administration (NOAA) Coastal Ocean Program (COP)	American Heart Association	Sloan Foundation
Open Access to Publications	S	S	R	O	O	O	O	O	O	R	R	R	S	O	O	O	O	O	O	R	O	O	O	R	O	O
Publication Repository Specified	O	O	R	O	R	O	O	O	O	R	R	R	O	O	O	O	O	O	O	O	O	O	O	R	O	O
Publication Repository Supported	O	O	O	O	O	O	O	O	O	R	R	R	O	O	S	O	O	O	O	O	O	O	O	O	O	O
Organizational Data Policy	R	R	R	R	R	R	R	R	R	R	R	R	R	S	R	R	O	R	R	R	O	R	O	R	O	O
Data Plan in Proposals	R	R	R	R	R	R	R	R	R	R	R	R	R	O	R	O	O	O	R	R	O	O	R	O	O	O
Data Timeframe	S	R	R	R	R	R	R	R	R	R	R	R	S	O	R	O	O	O	R	R	O	R	O	R	O	O
Data Access	R	R	R	R	R	R	R	R	R	R	R	R	R	O	R	O	O	R	R	R	O	O	R	R	O	O
Data Embargo	S	R	S	O	S	O	O	S	O	S	S	S	S	S	S	O	O	O	O	O	O	O	O	O	O	O
Data Preservation	S	O	R	R	O	R	O	R	O	R	R	R	S	O	R	S	O	R	R	R	O	R	R	R	O	O
Data Standards	S	R	O	O	O	R	R	R	R	O	O	O	S	O	R	O	O	O	R	O	O	R	O	O	O	O
Metadata Standards	S	O	O	S	O	R	R	R	O	R	O	R	S	O	R	O	O	O	R	O	O	O	O	O	O	O
Compliance	O	O	R	R	R	O	R	R	O	R	O	R	O	O	O	O	O	O	O	R	O	O	O	O	O	O
Data Center Specified	O	R	R	R	O	O	R	O	R	R	R	R	O	O	R	O	O	R	R	R	O	R	R	R	O	O
Data Center Supported	O	O	O	S	O	O	O	O	O	O	R	O	O	O	R	O	O	O	R	R	O	O	R	R	O	O
Funding	O	R	O	O	O	R	S	O	O	R	R	R	O	O	R	O	O	O	O	R	O	O	R	O	O	O
Scope	O	R	R	R	R	R	R	O	O	R	O	R	O	O	R	O	O	O	R	R	O	R	R	R	O	O
Guidance	O	O	O	O	O	S	O	S	S	O	R	R	O	O	O	O	O	O	O	O	O	O	O	O	O	O
Policy Date	2010		2002	1988, 1994, 2003	2009		1998			2003	2007, 2008	2008	2008			2008		2010	2009	1990s, 2000s	2010	1999	2011	2002		

Fig. 1 Data policies of major research funders in USA [2].

What is a Data Management Plan (DMP)

A DMP is a formal document that outlines what you will do with the data products during and after you complete your research [1]. Data management exists in various forms within almost all research projects. In recent years, many funding agencies have published formal requirements for DMPs (e.g., Fig. 1). These requirements cover issues in creating, organizing, managing and sharing data outputs throughout a research project. Through DCO-DS the DMP and related methods and tools will be promoted within the DCO community. We also hope the outputs of the DCO-DS works will contribute to the data policies of the AP Sloan Foundation.

Why we need DMPs for DCO

There are several reasons for us to promote and implement DMPs in DCO projects [cf. 3].

- **Efficiency** in project works: focus on research rather than worry about data
- **Understandability** of data: with detailed documentation the data can be understood by you and others in the future
- **Accessibility** of data: data in a repository will be available in a long time
- **Visibility** of your research: reuse and citation of data by others and relevance of your research
- Facilitate **new discoveries**: Open data and open scientific discussion lead to new discoveries

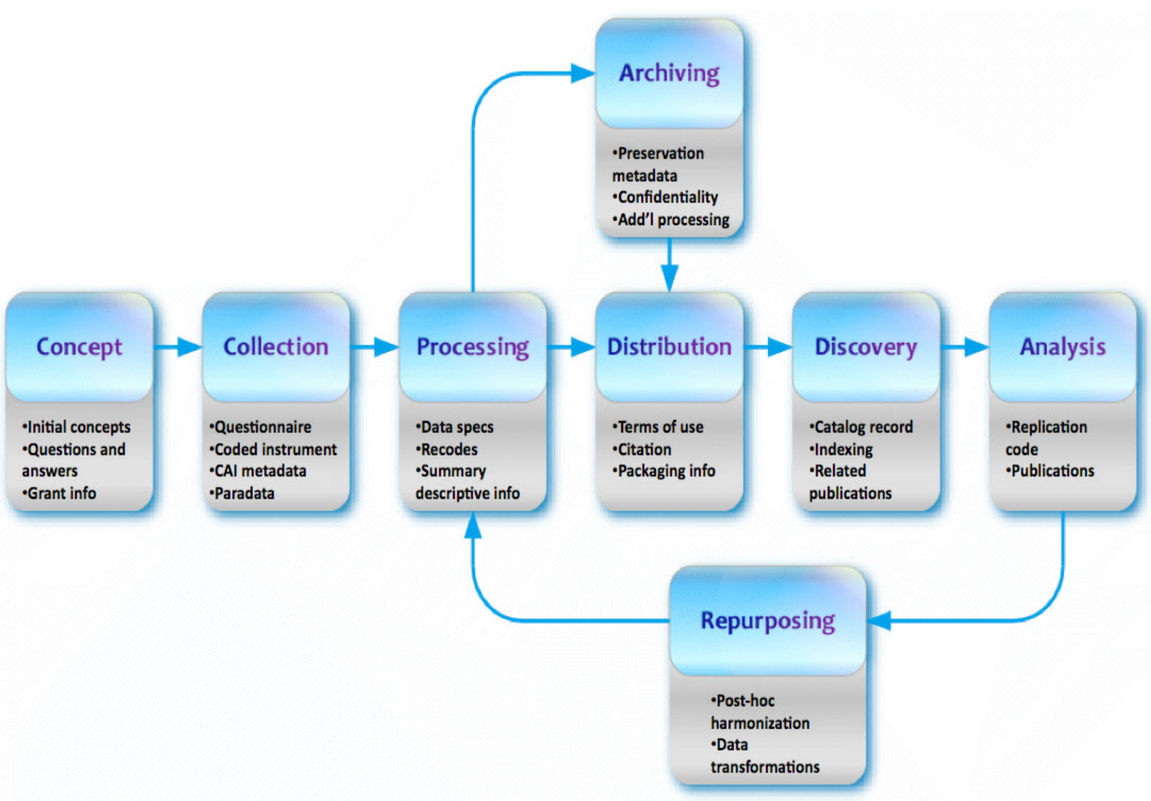

Fig. 2 The Data Documentation Initiative (DDI) data life cycle depiction [4].

How to let DMPs work for the DCO Community

There is a data life cycle (Fig. 2) in scientific activities such as those conducted in the form of research projects. To let the DMP work for the DCO community, the research data life cycle needs and requirements analysis for existing DCO projects will continue. Those analyses will be conducted by interviews structured around three broad phases of the data life cycle (acquisition, curation and preservation – since these typically have different people/roles involved) and placed in the context of best practices of DMPs for connected portions of the data life cycle. The goal of these interviews and analyses is to determine what the similar and different data management approaches/principles are for the DCO and its Directorates to effectively fulfill their short-term and long-term objectives.

DCO-DS and DMPs for DCO

As a positive consequence, new DCO projects will be informed by the current range of data science and DMP options, and be asked to respond to these topics in their proposals. We will add to online tools such as the IEDA DMP Tool [5] and the CDL DMPTool [6] that have DMPs for NSF programs and a few other foundations. The DCO-DS Data Management Team is adding a template for DCO purposes, and will work with prospective DCO projects to include plans in their proposals and retrofit plans for existing projects. DCO-DS is also developing a data portal that provides data preservation services and assigns identifiers and links to data of DCO projects that are stored in other repositories.

Visit DCO homepage

Get this poster

Sponsors:
Alfred P. Sloan Foundation

References:

[1] http://www.cdlib.org/services/uc3/dmp/index.html
[2] Dietrich, D., Adamus, T., Miner, A., Steinhart, G., 2012. De-Mystifying the Data Management Requirements of Research Funders. Issues in Science & Technology Librarianship (70), http://dx.doi.org/10.5062/F44M92G2.
[3] http://libraries.mit.edu/guides/subjects/data-management/why.html
[4] http://www.ddialliance.org/what
[5] http://www.iedadata.org/compliance/plan
[6] https://dmp.cdlib.org