Appendix 1 – Data Profile Template for the Department of Earth and Environmental Sciences

Part 1: Data Scope

This first part of the data profile template is created for collecting basic information about the research data that each member in the Department of Earth and Environmental Sciences is working with. The goal of the profile is to understand the characteristics of your research data (e.g., data types, formats, approximate sizes, etc.), providing the fundamental information for developing a departmental data management strategy and infrastructure.

The basic structure and some elements of this data profile template follows the example of the Data Curation Profile Toolkit [http://datacurationprofiles.org/] developed by Purdue University Libraries' Distributed Data Curation Center. While the Data Curation Profile Toolkit aims at facilitating research data's final preservation and archiving, our data profile template focuses on the ongoing data management process.

The content of this data profile template has also been influenced by information provided on the Rutgers University Libraries' Research Data Portal website[http://rucore.libraries.rutgers.edu/research/about], including data management advice, especially about data models, data lifecycle, and metadata.

Your Name:	
Your e-mail:	
Date of submission:	
Faculty advisor:	

Research area:	[ ] Atmospheric Chemistry [ ] Environmental Geochemistry [ ] Marine Microbe-Metal Interactions [ ] Near-Surface Geophysics [ ] Other: _____
Funding sources & requirements related to research data:	1. _____ 2. _____
Special notes:	

1. **Overview of research:**
    a. **Project title:**

    _____
    _____

    b. **Brief description of project, including timeline involved:**

    _____
    _____
    _____
    _____

2. **Data model based on research question, hypothesis, measurement and operationalization:**
    a. **Research question(s):**

    _____
    _____

_____

_____

b. **Hypothesis:**

_____

_____

_____

_____

c. **Conceptual diagram relating the different variables and measurements:**

d. **Operationalization of your research variables:**

Experimental Design		Variable	Measured by	Frequency of data collection	Temporal Scale	Spatial Scale
	Independent variable #1:					
	Dependent variables:					

Experimental Design		Variable	Measured by	Frequency of data collection	Temporal Scale	Spatial Scale
	Independent variable #2:					
	Dependent variables:					

3. **Information about your data:**

Data stage	Data output	# of files	Typical size of each file	Data format	Collection instrument	Analysis method/tool
Primary Data						
Supplementary data (e.g., from external sources)						

4. **Data storage, backup, and documentation:**

Data stage/product	Versioning frequency	Storage media	Backup strategy (frequencies & locations)	Data documentation (information recorded about data sets)

5. **Other notes about the research data:**

_____
_____
_____
_____
_____

6. **What challenges have you faced in managing your research data in the past?**

_____
_____
_____
_____

7. **What challenges do you anticipate for the future?**

_____

_____

_____

_____

Appendix 2 – Data Management Assignment in Preparation for Retreat

1) Bring a laptop, if you have one.

2) Bring along your research data files. We will discuss them during our workshop.

3) Prepare a three-minute presentation answering the following:
- How are you currently dealing with your research data? Describe your workflow,[1] including data products created at each step of the way. Consider using a diagram to help your audience visualize it better.
- What strategies do you use to organize your research data so that you can efficiently find it again later? (e.g., file structure and filing naming)
- What documentation (e.g., metadata,[2] data dictionary,[3] information about how the data was processed/analyzed) is included along with your research data? In what format is this information presented? What metadata standards, if any, are being followed?
- If you create your own analysis code (e.g., MATLAB script), how do you ensure it is something that someone else will understand to the point that s/he will be able to repeat your analysis?

If you have any questions, please feel free to contact Minglu Wang, Data Services Librarian (minglu@rutgers.edu) / Bonnie Fong, Physical Sciences Librarian (bonnie.fong@rutgers.edu).

---

[1] Workflow refers to the process you follow in collecting, cleaning, analyzing, and visualizing your data. Different types of new data may be created throughout the course of a project – e.g., visualizations, plots, statistical outputs, a new dataset created from the integration of multiple datasets, etc.

[2] Metadata is data that provides descriptive information (e.g., content, context, quality, structure, and accessibility) about a data product. It enables others to search for and use the data product. Metadata format is standardized structure and consistent content for metadata, usually in machine readable extensible markup language (XML) that can be represented in other human readable formats (e.g., HTML, PDF, etc.). Metadata standards are requirements for metadata documentation and are intended ensure the correct use and interpretation of data by its owners and users. Different scientific communities use different sets of metadata standards; common examples are EML (Ecological Metadata language), FGDC (Federal Geographic Data Committee) standard, and ISO 19115 (International Organization for Standardization Geographic information metadata).

[3] A data dictionary provides a detailed description for each element or variable in your dataset and data model. It is used to document important and useful information such as a descriptive name, the data type, allowed values, units, and text description. A data dictionary provides a concise guide to understanding and using the data.