Metadata

CORINNA GRIES

Purpose

- Understand your own data in the future
- Communicate with collaborators
- Publish data

Understand your own data

- File system
  - Raw data
  - Data manipulations (data and code)
  - Results (data and a readme file)
- File name
  - Descriptive
  - Version
  - E.g. Sparkling2014wtemp_20160630.csv

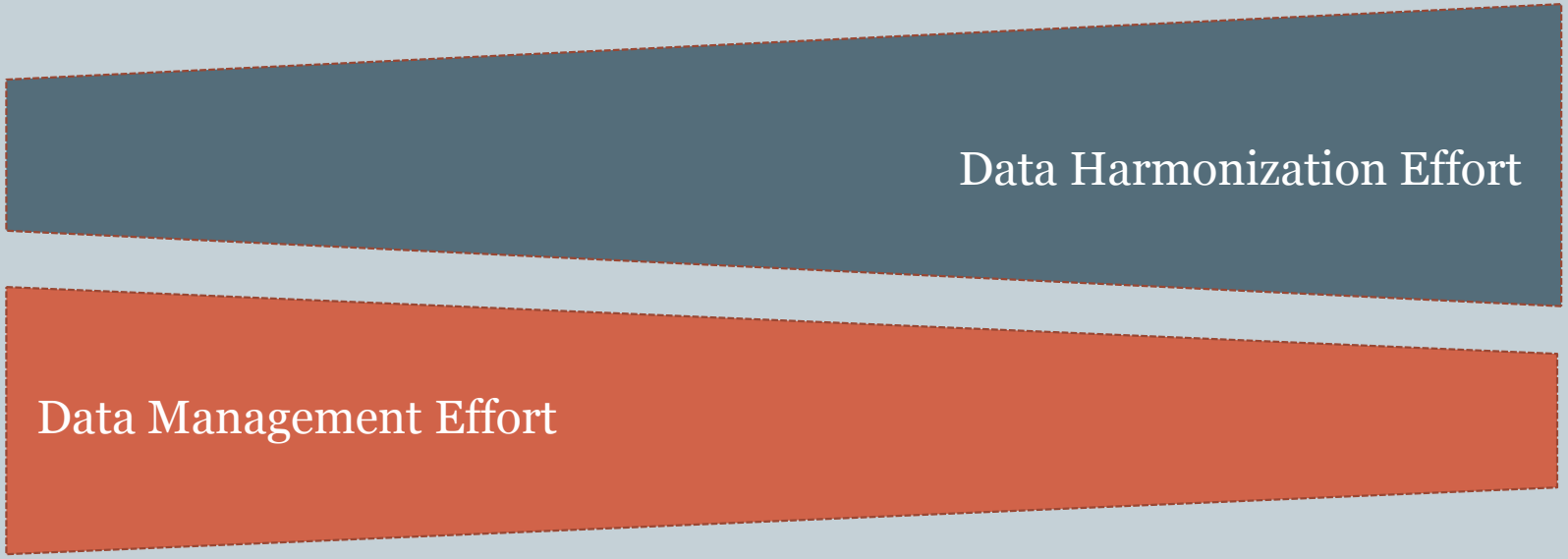
Communicate with collaborators

- Similar to above, but more extensive documentation
- Agreement on versioning system
- File system:
  - Keep originals in a place where they can't be overwritten
  - Manage file versions

Publish Data

- **WHO**
- **WHAT**
- **WHEN**
- **WHERE**
- **HOW**
- **WHY**

- Add data and metadata to well defined format (data model)
  - Pre-harmonized
  - Standardized naming of variables, units
  - Restricted to certain data types
- Document data with detailed metadata document
  - Flexibility to describe highly variable data
  - Little standardization
  - Mostly human readable



CUAHSI ODM

NetCDF
HDF

FGDC

EML
ISO

Title

- Titles are critical in helping readers find your data
- A complete title includes: What, Where, When, Who, and Scale

  - Height of Dominant Woody Plants Inside and Outside Exclosures Located in FP1

  - Daily weather data from Sagavanirktok River DOT site, in the northern foothills of the Brooks Range, Alaska, May-July 2011-2013.

Keywords

- Not already in the title or abstract
- General and specific
  - Organize on project repository (Website)
  - Consider the scope of the repository
  - Use a controlled vocabulary
    - LTER http://vocab.lternet.edu/vocab/vocab/index.php
    - GCMD http://gcmd.nasa.gov/learn/keyword_list.html
    - SWEET https://sweet.jpl.nasa.gov/
    - USGS biocomplexity thesaurus http://www.usgs.gov/core_science_systems/csas/biocomplexity_thesaurus/

People

- Consider an ID system (ORCID http://orcid.org/)

- Creator(s) – like authors of a paper
- Technical assistance – everyone else
  - Field personnel
  - Lab personnel
- Contact – somebody who can be contacted with questions regarding the data

Abstract, Purpose, Methods etc.

- Do not use jargon
- Define technical terms and acronyms:
  - CA, LA, GPS, GIS : what do these mean?
- Clearly state data limitations
  - E.g., data set omissions, completeness of data
  - Express considerations for appropriate re-use of the data
- The more detailed the methods the better
  - Sampling design
  - Instrumentation, lab protocols

Data Product

- Document well where they are coming from
- Document data manipulation
  - Quality control routines
- Archive of immutable data sets
- Data processing approaches
  - Harmonizing Scripts - GitHub
  - Statistical methods and packages (version)
  - Models (version), Variables