Understanding Data and Their Environment

Metadata and Provenance

EXERCISES FOR SESSIONS 1 AND 2

Nuno Pinto and Mark Elliot

Exercise 1 Creating a Metadata Schema

Design a metadata schema for the group's dataset so an analysist using the dataset can fully understand it. Attribute descriptions, data types, optionality, links to sources (among other attributes) need to be identified and specified.

Students can work on a spreadsheet that will then be uploaded and shared during the class.

Name this file "UDATE_CDT_[SURVEYNAME]_FINAL_METADATA".

Access https://www.dropbox.com/request/7Rv2Gr06iKJ0OSWMYy32 to upload the files.

Exercise 2: Using Metadata Schema for Creating an Indicator

Consider the design of a composite indicator using data coming from the two or more of the four datasets; think about what would be the attributes you would need to construct the indicator and how the available metadata would help or highlight limitations to it.

You need to take into account the data collected, the metadata explaining the data and how these datasets can be linked to support the computation of your indicator.

Produce a 1-2 slide presentation for 2:30 day 1

Exercise 3: Data Provenance

Produce a basic provenance graph for the index that you created in exercise 2.

- probably in a Word or similar document (using drawing tools).

- Closing Q&A at 16:10 on day 1

- Final version of document to be uploaded by 9:00 day 2 (for feedback).