Meta Matters - Enriching & Exploiting Your Metadata

Hacker, A[1]

[1]University of Oxford

Introduction

Data is nothing without context: if you don't know how, when or why a variable was gathered, it's nigh impossible to draw conclusions from it. This presentation discusses different sorts of metadata and how they can be gathered, stored, and used to enrich data; drawing examples from our biobank.

Objectives and Approach

Each data item has two types of metadata: variable-level and value-level. For example, consider a questionnaire. The variable-level metadata covers each question: exact wording, validation rules for the answers, etc. The value-level metadata covers each individual answer: details of the questioner, date and time of response, and so on. We also have database-level metadata: datasets which list every dataset or every field in the database. While some of this information needs to be gathered alongside the data itself, much can be extracted or imputed from results or documentation. We present some generalizable examples.

Results

Like any other data, metadata is only worth having if you're using it. We will present principles and examples of applications that we have developed for it:

- Data management – Deriving useful variables and tables, and helping to make your data easier to parse, extract, and validate.

- Presentation – Making your data more human-readable by labelling variables and decoding values.

- Documentation – Metadata tables make ideal repositories for granular institutional knowledge about your data: known issues, potential pitfalls, or explanations for missing values.

- Analysis – Identifying which metadata variables are most valuable for analysts, and how best to provide them.

- Automation – Using the metadata to generate code that can automatically produce summary statistics, tables, graphs. . . and more metadata!

Conclusion/Implications

Every dataset comes with some metadata. When examined and built upon, it can deepen understanding of the data within, as well as becoming a powerful resource in its own right.