

DATA QUALITY AND USABILITY*

Alex Voss

This work is produced by OpenStax-CN X and licensed under the
Creative Commons Attribution License 3.0[†]

Abstract

Discusses issues with data quality and usability in e-Research.

Most kinds of automated data manipulation and analysis require data to be of good quality, regular, well-defined and well-described. Very often, though, data in the Arts and Humanities, the Social Sciences and in Medicine (e.g., hospital records) is highly irregular, lacks adequate metadata and is of varying quality. Consequently, automated processing cannot be applied without further effort, workarounds or methodological compromises. For example, one researcher said:

"[our project] kind of died a death because the data which was available wasn't good enough to use any of the tools that social scientists [use] to look at the data, to manipulate it because the nature of the data is that it is fuzzy, it is not scientific data. [...] that is on hold until we can get better data."

Data that is made available for research is often anonymised, for example, by removing, restricting or aggregating variables, which makes it less useful for research. A social researcher remarked:

"in my view some of the survey data is unnecessarily reduced in its detail, sometimes I can fully understand why [...], sometimes I don't think it's necessary." (researcher)

*Version 1.2: Jun 9, 2009 4:09 am +0000

[†]<http://creativecommons.org/licenses/by/3.0/>