Data Practices

Data Practices

An empirical view of what people creating, analyzing, and managing data *actually do*.
(or would do)
so that we can improve efficiency and reliability

V1. Data Practices	(how do we know what works?)
V2: What's going on in the lab?	(brace yourself; it ain't pretty)
V3: Data sharing	(no, no, no, no, no. It's *mine*!)
V4: Data Reuse	(if you didn't make it, it is hard to use it)

V1. Data Practices

How do we know what works?

[The *empirical science* of data curation]

Empirical studies

Constructed, Naturalistic.

Data collection, data analysis

The empirical science of data curation

The science of *data analytics* is interdisciplinary
(and partly social science)

but mostly **mathematics**

The science of *data curation* is interdisciplinary
(and partly mathematics)

but mostly **social science**

A fundamental question in the science of data curation is

How can we more efficiently and reliably support the use of data?

This is clearly an *empirical* question

The empirical science of data curation: How to we do it?

How can we more efficiently and reliably support the use of data?

To answer this question we must conduct *empirical* research.

We can divide empirical research in this area into two rough categories

Constructed studies

Naturalistic studies

Constructed studies

These typically follow the classic scientific model. . .

conjecturing a hypothesis
and constructing an experiment or other targeted data collection to elicit confirmation.

Constructed studies are especially useful for
resolving an issue about what is influencing outcomes,
determining how a particular intervention might affect outcomes,
testing a new tool or practice.

Naturalistic studies

These typically collect data about an actual ongoing research situation,

observing what researchers and their staff actually are doing,
and then using that data to develop theories about data curation,
or ideas for changes in practices.
[There may or may not be an hypothesis at the outset.]

Naturalistic studies can be useful for identifying problems and opportunities
or developing general picture of a research and data management practices in a field.

Methods

Data collection methods in empirical studies of research and data curation processes include:

interviews, surveys, transaction log analysis, work product analysis (code, data, workflow) time and motion studies, experiments, simulations, and so on.

Analytical methods in empirical studies of research and data curation processes include

mathematics, qualitative analysis, interpretative methods, often informed by theories and results in computer science, information science, and cognitive science.

References (General)

Ball, A. (2010). Data lifecycles. In Review of the State of the Art of the Digital Curation of Research Data. Project Report. Bath, UK: University of Bath.

Borgman, C.L. (2012). The conundrum of sharing research data. *Journal of the American Society for Information Science & Technology,* 63(6): 1059-1078.

Babeu, A. (2011). “Rome Wasn’t Digitized in a Day”: Building a Cyberinfrastructure for Digital Classics. Washington DC.

Chao, T. C., Cragin, M. H., & Palmer, C. L. (2014). Data Practices and Curation Vocabulary (DPCVocab): An empirically derived framework of scientific data practices and curatorial processes. JASIST.

Hanson, K., Surkis, A., & Yacobucci, K. (2012). Data Sharing and Management Snafu in 3 Short Acts [video].

Hey, A. J., Tansley, S., & Tolle, K. M. (2009). The fourth paradigm: data-intensive scientific discovery.

Cragin, M.H., Palmer, C.L., Carlson, J.R., & Witt, M. (2010). Data sharing, small science, and institutional repositories. *Philosophical Transactions of the Royal Society A*, *368*(1926), 4023-4038.

Pepe, A., Goodman, A., Muench, A., Crosas, M. & Erdmann, C. (2014). How Do Astronomers Share Data? Reliability and Persistence of Datasets Linked in AAS Publications and a Qualitative Study of Data Practices among US Astronomers. *PLoS ONE,* 9(8): e104798.

Research Information Network. (2008). To Share or Not to Share: Publication and Quality Assurance of Research Data Outputs.

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., Manoff, M., et al. (2011). Data sharing by scientists: Practices and perceptions. PloS ONE, 6(6), e21101.

Tenopir C, Dalton ED, Allard S, Frame M, Pjesivac I, Birch B, et al. (2015) Changes in Data Sharing and Data Reuse Practices and Perceptions among Scientists Worldwide.