

Provenance, End-User Trust and Reuse: An Empirical Investigation

Devan Ray Donaldson and Kathleen Fear, School of Information, University of Michigan

Research Questions

- How does provenance affect end-users' trust in data?
- How does provenance affect end-users' confidence in data with respect to reuse?

Methodology

- Proteomics and ProteomeCommons.org
- Semi-structured interviews with end-users of scientific data (17 proteomics researchers)

How we define provenance

- We examined each element in each module of the MIAPE standard and selected those that we deemed related to provenance
- These elements include:
 - the date on which the data were initiated
 - the name(s) of the person(s) responsible for the creation of the data
 - information about data transformation techniques used, analysis tools used, and information about data generation, including the location of the raw data, databases queried or specifications of equipment and conditions under which the data were produced

Findings

- Provenance information on its own is sufficient to engender some amount of trust in the data housed in ProteomeCommons.org: trust that the data have the potential to be reused. However, this trust is *provisional*
- The addition of information about **data quality, the author(s), and the dataset itself** helps end-users trust data *even more*
- No subject indicated that any provenance information was unnecessary

Implications of this Research

- Studies of end-users and the environments in which they make decisions about trust and reuse can shed light on factors that impact the role of provenance in facilitating trust and potentially offer a more nuanced view of the interrelationship between users, trust and provenance.

Acknowledgements

- Ann Zimmerman
- Phil Andrews
- Paul Conway
- Margaret Hedstrom
- School of Information, University of Michigan
- Rackham Graduate School, University of Michigan