(Original Title) Now that you have lots of experimental data, is it worth looking at?

What information should accompany data regarding measurement uncertainty, accuracy, and relevance for integration with other data?

Anne Plant, Susan Gregurick and

A broad breadth of applications and scientific expertise present at breakout. Ranging from health care to military stakeholders; from experimentalists to library science and data science.

Topics of Discussion:

Linking metadata with experimental data, or **MetaData Standards**

- Are some existing stds for metadata (e.g. ISO) includes content, input & format of data. **This is Important across disciplines.**

- Some precedent for collecting provenance data and linking w (meta) data via **DOI or PID** in a separate metadata repository.

- Provenance data needs to be retained during data analysis.

Could this be achieved/available with pipeline analysis software tools? **We should automate provenance and assure original data are preserved.** Some automated archival tools exist and may be adaptable?

Tools for integrating heterogeneous data or..
Workflows and Reproducibility

- **Reproducibility requires the workflow for data analysis to be captured**.
  - Involves sociology in sharing workflows.
  - Publication should capture workflow info. Applies to experimental, analysis, modeling, etc
  - Led into a discussion of use of Virtual machines for collecting operations used in a workflow.
- **Identity management is part of Provenance**.
  - Changing format: analog>digital, tif>jpeg How does impact quality of data?
  - Should we Store all version of data? **Track/Store data usage information** (****this could provide incentivization for undertaking the hard task of metadata)?
  - Append metadata with usage data to help guide the best use for the data.
  - Keeping metadata w the data by links not by imbedding in the datafile. **Persistent identifiers.**

Challenges for future directions

- **Data extraction to do discovery of patterns within large experimental datasets.**
  - **Caution:** getting value out of data (particularly found data). Could be useful for discovering hypothesizes but not always clear that is leads to real value information.

- **How to evaluate the validity of data?**
  - Are there Tools available for data evaluation?

One Use Case Example

Patient centered outcome Research: VA / U CA Health care systems. 20 research networks (contract)
Deliverables: using existing data for retrospective analysis of medical data across institutions.
Funding to bring technical and policy issues together.

This model could/should be pursued with other use cases