Working with / linking to specialized data repositories: how do identify the ‘right’ one?

Enriching the description of the experiments to enhance query capability: do community standards help?

Susanna-Assunta Sansone

Associate Director, Principal Investigator
Should / could Dryad:

- Capture more details of the experimental workflow…
- Make the annotation explicit and discoverable…
- Allowing more query options to filter/retrieve data?
Growing number of reporting standards

- allow data to flow from one system to another
- use the same word and refer to the same ‘thing’
- report the same core, essential information
Growing number of reporting standards

To track provenance of the information and ensure richness of data and experimental metadata descriptions, to maximize reusability
But how much do we know about these standards

Which ones are mature enough for me to use or recommend?

What are the criteria to evaluate status and value?

CURATORS & DEVELOPERS

I use high throughput sequencing technologies, which ones are applicable to me?

TERMINOLOGIES

CHECKLISTS

DATABASES

FORMATS

POLICIES

HOW CAN I GET INVOLVED TO PROPOSE EXTENSIONS OR MODIFICATIONS?

RESEARCHERS

Which tools and databases implement which standards?

FUNDS & JOURNAL EDITORS

I work on plants, are these just for biomedical applications?
A coherent, curated and searchable registry of standards for describing and reporting experiments in life science, environmental, biomedical and biotechnological domains.
• A coherent, curated and searchable registry of standards for describing and reporting experiments in life science, environmental, biomedical and biotechnological domains

• Progressively associate standards to data policies and databases

• Develop assessment criteria for usability and popularity of standards

• Help stakeholders to make informed decisions on e.g. what standards or databases to use or recommend
Criteria to be used in evaluating standards for adoption:

- Existence of a formal specification, with:
  - good level of documentation, with scope and use cases
  - ease of implementation
  - human and machine readability

- Broad adoption and implementation, outside the initial group by:
  - community databases (hence existence of standards-annotated datasets)
  - software (e.g. for reporting, editing, curating, submitting to databases)

- Active user community, also providing:
  - support
  - responsiveness to community requests
  - examples

- Interoperability with and extensibility to other standards, ranging from:
  - compatibility with other standards
  - flexibility to cover new domains
  - conversion and mapping, if applicable

- Openness

Jessica D. Tenenbaum
Duke Translational Medicine Institute
Melissa Haendel
OHSU Library
Simon Lin
Marshfield Clinic Research Foundation
part of the NIH Clinical and Translational Science Award (CTSA) program
Core attributes to describe databases and assist in evaluating scope and relevance as well as access to data:

- Database name
- Main resource URL
- Contact information
- Date resource established (year)
- Conditions of use (free, or type of license)
- Scope: data types captured, curation policy
- Standards implemented: checklists, terminologies, formats
- Taxonomic coverage
- Data accessibility/output options
- Data release frequency
- Versioning period and access to historical files
- Documentation available
- User support options
- Data submission policy
- Relevant publications
- Tools available