Apr 6th, 1:00 PM - 2:00 PM

Description and annotation of biomedical experimental data sets: work in progress

Jen Ferguson
Harvard School of Public Health

Description & annotation of biomedical experimental data sets: work in progress

Jen Ferguson • Harvard School of Public Health

1 Objective

Biomedical data deposition is on the rise as more scientists make their experimental data openly available. However, lack of context and metadata can create obstacles to the understanding and reuse of these data sets.

Researchers from the Harvard School of Public Health Bioinformatics Core Group have attempted to address this issue by assembling a team of curators to annotate and contextualize this data. The curation team includes life sciences graduate students and postdoctoral associates, plus one science librarian (JF).

2 Methods

ISATools, an open source software suite, is used to annotate experimental descriptions and data sets. Curators draw upon information from PubMed papers and associated data sets (GEO files) to create **ISATab records.**

Curators annotate and describe both raw and derived data files for each investigation, as well as supplying metadata for the investigation as a whole.

Once annotated, the records are validated using **ISAvalidator** and sent to an internal data management system.

3 Results

As of March 2011, HSPH has collected over 50 annotated studies comprising 900+ assays in their internal data management system.

The ultimate goal:

- create records that clearly tie curated, metadata-enriched data sets to published works, and
- make this information openly available in public repositories.

4 Conclusions

Work on this curation project has yielded useful insights for both scientists and librarians.

Similarities:

- What information is relevant for curation?
- How much metadata is enough/too much?

Differences:

- Areas of expertise
- Perspectives and concepts of scale

There are valuable roles for librarians to play in the curation process.

Some of the most worthwhile contributions that librarians can offer may occur prior to the actual curation process, in areas such as training and software development.

References

1. Field, D., et al. "'Omics Data Sharing." *Science* 326.5950 (2009): 234-36. Print.
2. ISA Team. "Isa - About". January 24 2011. http://isatab.sourceforge.net/index.html.
3. Rocca-Serra, P., et al. "Isa Software Suite: Supporting Standards-Compliant Experimental Annotation and Enabling Curation at the Community Level." *Bioinformatics* 26.18 (2010): 2354-56. Print.
4. Rocca-Serra, Phillippe, Eamonn Maguire, and Susanna-Assunta Sanson. "The Isa Infrastructure: standards & Software for Annotating, Managing and Sharing Life Science Investigation." Ed. Oxford, University of. 2010. Print.