Overlooked and Overrated Data Sharing

Tushar B. Kute,
http://tusharkute.com

Data Curation

- Data curation, particularly within academic libraries, has gained appreciable momentum by developing an energetic community dedicated to providing widespread access to well-curated data.

- In one vision of the future, the data required to validate or extend a research study is readily available, and the publication of data itself will bear an equal importance to that of the article publication.

Data Curation

- The data curation community is eager to help catalyze that transformation through services and advocacy.

- Yet in practice, it's not uncommon to encounter scientists who question the cost-benefit ratio of the time and effort involved with curation, publication, and preservation of research data.

- How can something that seems so self-evident to the data curation community be so challenging to implement in the wild?

- Scientists frequently think of article publication as a form of data sharing, and it is critical to acknowledge not only that this concept exists, but also that it has been recapitulated throughout their communities, including funding agencies.

- As of October 2015, NIH's Data Sharing workbook still says "Some studies, such as small laboratory-based projects, make raw data available in publications.

Overlooked Data Sharing: Article Publication

- Likewise, example data management guidance available from NSF and USGS websites reference data sharing via publication.

- In an analysis of 1,260 Data Management Plans (DMPs) submitted for NSF applications at the University of Illinois, Mischo, Schlembach, and O'Donnell found "publication" listed as a data sharing mechanism 44 percent of the time.

tusharkute
.com

- Similarly, a form of data sharing often overlooked in the data curation community is supplemental material provided along with a published journal article, also known as supplemental data, auxiliary information, supporting information, or supplementary content.

- Supplemental material is generally supplied to a publication in free form as an extra file or files that help support the main article.

- A prototypical example is a PDF that may include additional text, methods, analyses, figures, tables, and/or data, but other supplement examples may include file formats incompatible with article format or layout, such as video or code.

- PLOS and Science, as just two examples, allow a myriad of file formats as supplemental files.

Overrated Data Sharing

- In juxtaposition to the unstructured nature of supplemental materials or the limitations of published articles, an untold number of highly structured and sophisticated data resources have also been developed.

- When the topic of domain repositories is broached, successful well-known examples such as Inter-university Consortium for Political and Social Research (ICPSR), GenBank, or the Sloan Digital Sky Survey quickly spring to mind; however, as of October 2015 the Registry of Research Data Repositories, re3data.org, contained 1,363 reviewed repositories with representation across both the humanities and sciences.

Overrated Data Sharing

- The number of resources represented in re3data.org is steadily growing, and it's understood that the registry is not yet comprehensive.

- For example, since 1993 Nucleic Acids Research has published an annual "Database Issue" and maintained an online Molecular Biology Database Collection that currently references 1,549 databases dedicated solely to bioinformatics and molecular biology.

- Thus it's difficult to estimate how many data resources are currently available, but clearly data resources are of keen interest to many research communities.

Thank you

@mitu_skillologies

/mITuSkillologies

@mitu_group

/company/mitu-skillologies

MITUSkillologies

Web Resources
https://mitu.co.in
http://tusharkute.com

contact@mitu.co.in

tushar@tusharkute.com