Making Your Data Story Inclusive

Jessica Koos, MLS, MSEd, AHIP & **Laurel Scheinfeld**, MSLIS

Stony Brook University

Introduction:

Increasing attention towards transparency and reproducibility in research has highlighted the importance of data management strategies. Libraries have been assuming a pivotal role in training researchers on best practices in RDM. However, it is essential to consider how RDM tools, systems and practices may impact diverse communities. This poster is an exploration of some of the various challenges to inclusivity in research data management, including accessibility, biased metadata, privacy, and how big data may marginalize certain communities.

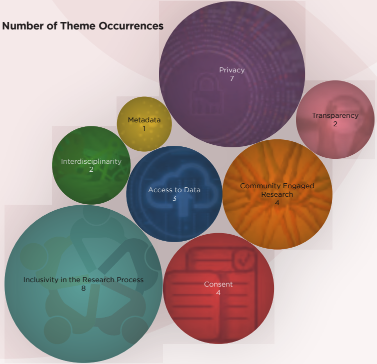
Methods:

A literature review was conducted by both authors. The databases that were searched included PubMed, Web of Science, SOCIndex, LISTA and Academic Search Complete. Google and Google Scholar were searched as well in order to capture grey literature. Search terms such as "ethics," "inclusivity," and "research data management" were included in the search. The date range was limited to 2015-2020 in order to keep the number of results manageable and current. Only articles with a prime focus on ethics and/or inclusivity in regards to data management were included. Each author analyzed the results and coded them based on themes. Both authors reviewed and agreed upon the themes.

Results:

Twenty-one documents that met the above criteria were identified. The document types included commentaries, editorials, research articles, a blog post and a scoping review. The themes and the number of articles in which they appeared are included below, as well as potential solutions from these articles.

Number of Theme Occurrences



Solutions to each of the challenges presented by the themes are as follows:

Privacy
- Strong oversight is needed in small communities where identity breaches may have unintended consequences
- Advocate for government policies that require big data producers to preserve the privacy of subjects

Transparency
- Make individuals aware of how their data is being used
- Indicate who is creating and using the data, in order to provide context

Metadata
- Create and use alternate vocabularies for metadata
- Consult the communities reflected in descriptive metadata to facilitate moving to more appropriate and respectful terms
- Use standardized ontologies to facilitate integration of multiple data sources

Interdisciplinarity
- Build 'acceleration teams' that include not only patients, advocates, community leaders, clinicians and researchers, but also funders, entrepreneurs and policy-makers
- Creation of new, multidisciplinary professionals, called "health information counselors"

Community-engaged research
- Practice community-engaged research whenever possible
- Encourage public involvement and shared ownership in RDM
- Gain trust of minority community members to garner their participation

Access to data
- Convert older educational modules on research data management from static PowerPoint slides to slides in R Markdown and post in GitHub, so they are open and customizable by the broader research data management community
- Share data whenever possible
- Create shareable datasets on the specific health concerns of disadvantaged populations
- Provide training and support to diverse groups on working with data
- Advocate for government policies requiring dataset sharing among large corporations

Inclusivity in data collection, processing and distribution
- Use algorithms and models that are designed to incorporate diversity into big data
- Promote diversity within the data science workforce
- Include socio-behavioral determinants of health in big data
- Develop a model for distributing benefits in a way that includes study participants.
- Include gender/sex information in data reporting
- Ensure that the information used in developing AI is inclusive and that ethnicity, gender and other data is accurately recorded
- Include data from small populations
- Create incentives to include research subjects from diverse populations

Consent
- Enhance informed consent by providing detailed data sharing information, including who will be benefiting from the research
- Involve data ethics committees in commercial research
- Advocate for legislation that allows subjects to control the use of their own data
- Create an upfront license for consenting patients that results in a royalty payment every time their data are included in a research study
- Develop education programs where experienced digital data researchers inform ethics committees about their research, highlighting both the benefits and the risks
- Follow the International Ethical Guidelines for Biomedical Research Involving Human Subjects, which are recognized as universally applicable

Poster designed by **Joel Koos, MFA** - *jkoosart@gmail.com*