Big Data to Knowledge
Scientific Data Council
Sustainability Working Group

Request for Information: Metrics to Assess Value of Biomedical Digital Repositories (NOT-OD-16-133)

Executive Summary

Biomedical digital data repositories manage and store biomedical content, and provide the research community with access to these data. The increasing volume of biomedical data and the requirements for data sharing have contributed to the growing number, scale, and complexity of digital repositories. With this context in mind, the National Institutes of Health (NIH) issued a Request for Information (RFI) on metrics to assess the value of biomedical digital data repositories. The RFI was released on August 12, 2016 and was closed on October 17, 2016. The goals of this RFI were to solicit input from stakeholders on existing metrics used to assess digital repositories, metrics that are most meaningful and desired, and metrics that are most appropriate given the diversity of biomedical research.

The RFI produced ninety-eight responses from the research community representing over ten different professional domains including academia, government, private industry, publishers, and non-profits. Respondents were from over ten countries including the US, UK, Japan, Australia, and Germany.

The metrics suggested included both quantitative and qualitative metrics. The RFI respondents recommended that when possible, quantitative metrics should be used in sets of two or more to provide a multifaceted assessment of value. Respondents also suggested the use of qualitative metrics such as descriptions of users' overall satisfaction with a given repository.

A number of respondents expressed the benefit of using both quantitative and qualitative metrics in an ecological perspective. That is, given the diverse range of data types and scientific disciplines, evaluating each repository in a context-specific manner. Furthermore, it was suggested that metrics may be more useful when accompanied by a description of their intended use and limitations. Another valuable insight was that metrics can be more meaningful when assessments are preformed iteratively, at intermediate intervals, not just at end points. To this end, many respondents recommended that metrics should also consider repository lifecycle stages (e.g. emerging, mature, retiring).

Stakeholders communicated their recommendations for the use of multiple robust metrics that could assess the breadth and depth of biomedical digital data repositories such as: percent coverage of the user base; guidelines for data and repository citations; use of community accepted metadata standards; an appropriate level of data curation; availability of user support from expert staff; existence of terms of use documentation; a

plan for sustainable funding; and the use of user surveys or case studies to collect qualitative feedback.

The consensus from respondents was that there is a large diversity of digital repositories in terms of their functionality, data objects, and user community. Thus, it would be beneficial to use multiple metrics in combination to assess repository value while accommodating this diversity. Lastly, several respondents suggested that existing resources from international organizations could serve as an excellent foundation for the NIH's efforts toward digital data repository assessment.