Request for Information (RFI):
Data Annotation in Biomedical Core Research Facilities
and Related Needs for Community Education and Training
Notice Number: NOT-OD-16-091

I. Executive Summary

The objective of this RFI was to learn about sources of data and metadata standards in biomedical core research facilities, best practices that guide creation of metadata, and the roles of different stakeholders in creating essential metadata.

A total of 15 responses were received, with some of the responses describing practices in multiple cores. *Highly variable comments about practices in data annotation and related training* were reported in the responses. They indicate the complexity of the topic, suggesting that different scientific communities and different core centers are at different stages of developing and implementing methods for data handling. The practices in some cores may serve as models for wider dissemination.

The *most sophisticated strategies for data management are adopted by cores that form partnerships with bioinformatics or biocomputing groups*. However, in many research cores knowledge of data-related issues is lacking; there is no sufficient communication between data scientists, technical staff, and investigators. Within a specific community, different laboratories or cores may adopt their own approaches. In some cases, cores provide metadata only if they are requested by an investigator. Those who recognize this problem express a *need for assistance with implementation of standardized approaches for data creation and procedures for data management, including assignment of metadata.*

Some users report a *need for uniform standards of metadata* while others realize the *challenges associated with adapting uniform standards across different fields and across many technologies*.

Responses to this RFI also point to *needs for agreement upon best practices and methods for data generation, management and analysis, and for training all participants – researchers, technical staff, and data scientists*. Some of the responses indicate that clearer guidance from the NIH would help to foster more consistent and high-quality metadata across cores and institutions. In the absence of such guidance, each core develops and implements its own best practices.

In conclusion, the responses from a variety of stakeholders in core research facilities suggest the NIH can provide guidance in the following aspects of data management:

1. Facilitate training and education of researchers, technical staff and data scientists on data management with respect to basic knowledge of data types, annotation methods, databases, etc., and inform the cores about already existing resources that support such training;
2. Encourage intra-institutional and multi-institutional partnerships among cores and bioinformatics centers to promote standardization and use of best practices for data annotation and management;
3. Support community development of guidelines for metadata standards and best practices in data annotation, and assist with dissemination and implementation of these guidelines;
4. In partnership with other funding agencies, provide guidance on data management (including data annotation) that would foster high-quality data and metadata across scientific communities.