Overview of the Responses to RFI NOT-ES-15-002: Making Data Usable--A Framework for Community-Based Data and Metadata Standards Efforts for NIH-relevant Research.

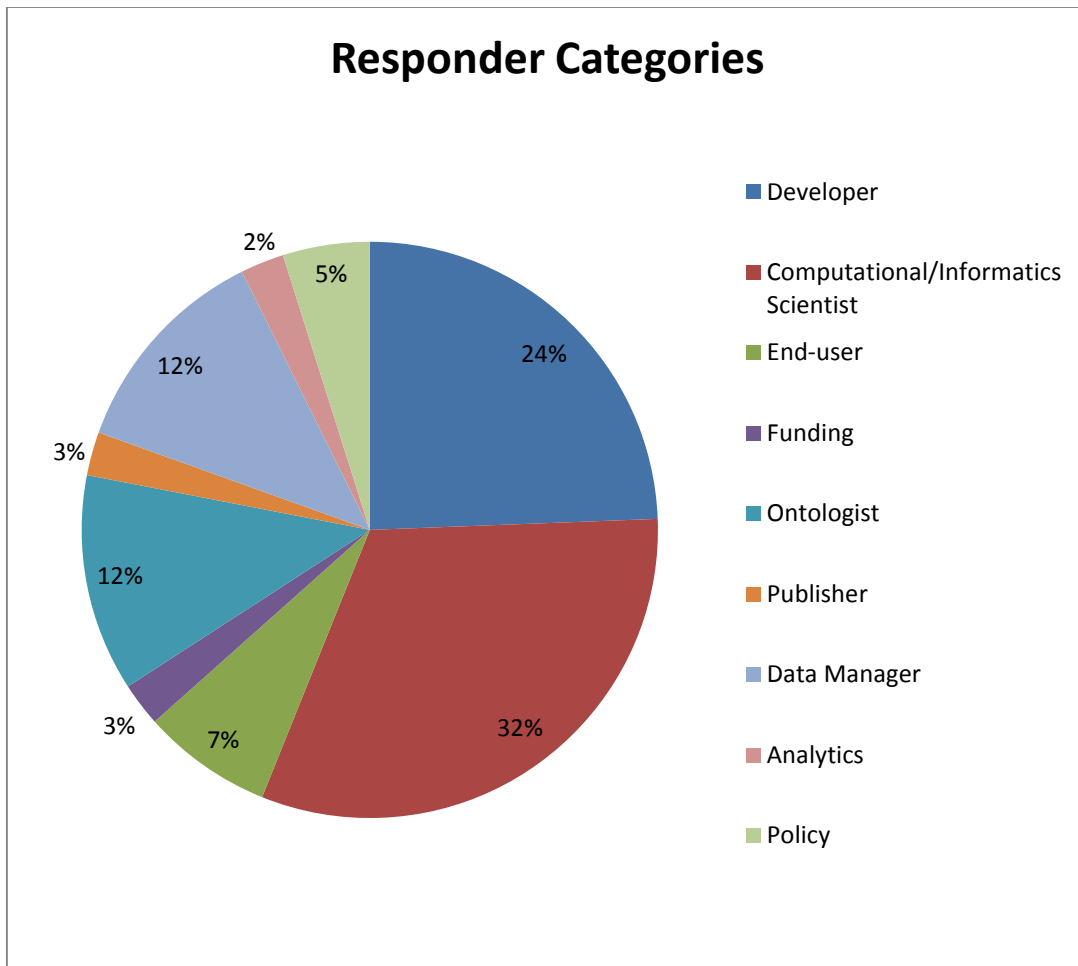### *Themes and implications for BD2K CBS initiatives and activities*

In Request for Information (RFI) NOT-ES-15-002, the NIH invited comments and ideas from interested people to inform on how community standards activities are initiated, developed, disseminated, and sustained, and any role that NIH might play in helping to catalyze such efforts. Some suggested areas of interest related to Community-based data standards were as follows:

- Effective approaches, processes, and activities that could advance the community-based standards landscape (e.g., creating a collaborative workspace or an advising structure toward standards development, extension, or adoption).
- Gaps in community-based data standards of relevance to biomedical research, including real use-cases (e.g., emerging fields and technologies, or research domains with multiple existing data standards that could benefit from additional work, integration and/or reconciliation).
- Lessons learned from existing CBS efforts, particularly examples with field-tested processes and infrastructure or known examples of failures by CBS efforts.
- Common challenges in CBS development (e.g., methods for community engagement or building interoperability with other related standards).
- Considerations for evaluating progress and milestones to assess data standards development and utility.
- Effective approaches for addressing the need to sustain useful standards, and to update existing standards as a field develops.

Respondent Analysis

The RFI produced 30 responses from a diverse group including standards developers, ontologists, data managers, computational and basic science researchers, publishers, and people involved in funding and policy. Some responses were from individuals and others were from individuals responding on behalf of an organization.

Responder Categories

Developer — 24%
Computational/Informatics Scientist — 32%
End-user — 7%
Funding — 3%
Ontologist — 12%
Publisher — 3%
Data Manager — 12%
Analytics — 2%
Policy — 5%

Response Highlights

Initial analysis of the responses revealed several important themes among them, which are listed below:

- Despite widespread use of virtual communication tools, many respondents emphasized the need for periodic face to face meetings to build consensus and move the process forward. Most bemoaned the lack of available funding for travel and in-person meeting logistics.
- Multiple comments pointed out the ever-expanding universe of data standards, with duplication and overlap the norm rather than the exception. This was tied directly to difficulties in discovering existing standards, with accurate annotation and updating—based on the RFI comments, a robust standards inventory such as being envisioned by BD2K would be much-welcomed.
- There is a general tension between the need to build standards close to the data, with focused use cases that address specific biologic questions while also ensuring that the

data standards interoperate with other existing standards and that their application to related domains is appropriately considered. Some respondents focused their comments on processes for building interoperability and provided good examples of that.

- A common pain point identified was the transition from initial development of a data standard to long-term support for maintenance, updating etc. A few respondents noted that involvement of SDOs in long-term management would be beneficial.
- Inconsistent use of best practices for maintaining versioning and linking data sets/analyses to data standard versions was noted.
- A large number of respondents spoke to the unpaid 'volunteerism' that is the norm for much of data standards development; this raises multiple problems—some valuable contributors simply cannot devote unpaid time, others do so using 'stolen time' from existing grants or activities. In general, this means that standards development happens more slowly than would be ideal and that the individuals with the best expertise may not be those engaged in the process.
- A social/cultural challenge is the lack of recognition of standards development work in evaluating productivity and the need to put in place ways to cite data standards and enable linkage/attribution to the standards developers.
- Many endorsed the need for transparent, collaborative frameworks for developing data standards and open access to the products. Several mentioned the importance of ancillary products—e.g., APIs, clear user guides, educational programs.
- A few respondents noted the lack of a formal framework for evaluating the utility of data standards, with the default evaluation being the extent of uptake by the community.
- There was widespread recognition of the multi-stakeholder nature of data standards development and the need to identify and engage multiple communities—e.g., domain experts, software engineers, bioinformaticians, tool developers, end users.
- Several respondents encouraged NIH to be more proactive in encouraging/requiring documentation about data standards in grant applications. Others mentioned the important role of publishers in encouraging inclusion of information about data standards in publications.
- Many respondents provided examples that highlighted the complex evolution of data standards, from initial conception to maturation and widespread use. Complexity was the norm, e.g., with transitions in involvement of societies, domain experts and management structure over the life cycle. The comments reinforced the landscape of data standards as a complex web, with an initial standard often spawning multiple related data standards.
- An initial pain point described by a few respondents was the lack of standards development experience by initial domain experts who often band together to initiate the development process. A resource that provides consulting to such groups and/or that lays out best practices would be welcomed.

In summary, the RFI received a good number of responses (30) from a range of individuals and groups that represent standards developers and users. Common themes among the respondents included how important face-to-face meetings are for moving the standards development process along, despite all the virtual communication tools, and how hard it is to find funding for those in-person meetings. Many noted that duplication and overlap in standards was the norm not the exception, in part due to difficulties in discovering information about existing standards work. There were also concerns about long term sustainability and maintenance of standards. Many who responded noted the critical ingredients of transparency, collaboration, and engaging diverse communities. The problems that arise associated with unpaid (and often unrecognized) volunteerism were cited multiple times. Finally, gaps in how the usefulness of standards is evaluated, was noted. Overall, there were no big surprises, and the responses matched well with what has been heard in the past from individuals engaged in standards development.