# Report on Frameworks for Community-Based Standards Effort Workshop

**Preamble**
This report is meant to convey the larger points that emerged over the course of an NIH workshop held on September 25 and 26, 2013.  For details of the comments and discussions from which these points issue, the reader is referred to the archived video of the entire workshop (Day 1 and Day 2).

**Purpose**
The purpose of this workshop was to learn - from many different perspectives - what has and what has not worked in community-based standards efforts.  More than a dozen specific issues that pertain to the entire standards life cycle (efforts that include formulating, conducting, and maintaining standards) were addressed and discussed from these perspectives.

**Introduction to BD2K and to the Workshop**
The NIH Big Data to Knowledge initiative (BD2K) aims to:  1) advance the science and technology of biomedical big data, 2) enhance and develop the workforce for biomedical big data, and 3) facilitate the broad use of biomedical research data.

BD2K will facilitate the broad use of data by making data available, discoverable and useable.  Data are most useable when they can work with other data, software tools, and data resources such as data repositories.  It is through the use of data-related standards, including those for metadata, that allow data to so interoperate.

As biomedical research data are increasingly digital, and as the volume and diversity of data increases interconnecting data, tools and resources has become integral to the science, itself, and the value of these interconnections are becoming clear to the research communities using them.  This digital transition is often accompanied by the recognition for the need of data-related standards by members of that community.  Since the adoption and use of standards requires effort, those standards that come from the community for the community are often most readily accepted.  Support for community-based standards efforts[1] represents a leverage point with the promise of broad impact.

While NIH has supported community-based standards efforts, these have been done as one-off activities, as routine ways to support these non-traditional but increasingly important efforts are not available.  In contrast, NIH has many routine ways to support any of a large number of different types of research and training activities (e.g., research project grants, Small Business Innovation Research grants, institutional training awards, etc.).  Each of these includes not only an award mechanism (e.g., R01, R43, and T32), but also administrative procedures, policies, eligibility requirements, review criteria, etc.  Together these form frameworks that vary for each type of research or training activity.  By working within their respective frameworks, many different types of research and training activities can be supported in routine ways. Such frameworks are now lacking to support community-based standards efforts; BD2K will establish such frameworks[2].

---

[1] Where "community-based" refers to any of the full panoply of stakeholders, such as domain experts, technical experts, librarians and information scientists, vendors, publishers, funders, etc.  And, where "standards efforts" may include any of many activities associated with data-related standards, such as: identifying, describing, mapping, monitoring, or updating existing standards, modifying and repurposing standards, developing new standards, disseminating standards, etc.

[2] A framework of policies, governance, administrative procedures, funding mechanisms, etc. will be developed by NIH to support community-based standards efforts.  It is paramount to make clear that this component of BD2K would not support comprehensive NIH data standards efforts; rather, it would focus on particularly opportune community-based standards efforts related to research data that would make an important difference across broad aspects of NIH-supported research.  Nor would this component assume responsibility for any particular standards effort for the long term; rather, it would provide time-limited assistance, mostly in convening, organizing and logistics, to facilitate a community of practice address standards-related issues.  These frameworks would provide catalytic support.

**Community Standards:  Towards Truly Reproducible Science**
The expectations of funding organizations, those who benefit from biomedical research, and society in general are changing.  There is increased interest in making the results of research more widely accessible and in increasing the return on research investment. For data, these interests align when data that are accessible are also usable by employing data-related standards.  This promotes interoperability and adds significant value to their use.  As investigators add this value to their data, their attention is turned to finding standards that will best serve their scientific needs.

With widespread use of community-accepted standards, not only will it be possible to share conceptual products of research, such as interpretations of results in the context of a particular theoretical perspective, but also to share the data in a meaningful way.  With the use of metadata to put any given data set in its proper context, and with the use of other data-related standards, reanalysis, aggregation with other data sets, and reproducibility of interpretations, or of whole experiments, becomes possible, and the return on investment is multiplied.  Without the use of data-related standards, none of this is possible.

**Mapping the Landscape of Community Standards**
As the promise of data sharing and interoperability become more compelling, investigators are increasingly interested in identifying data-related standards that will give them a competitive advantage by working with key data sets, tools or repositories.  Since so many standards relevant to biomedical research data begin as "boutique" efforts that later gain a foothold across a community.   Because standards are developed and can be described from so many different starting and vantage points, and because there are no conventions for describing standards, it is often difficult for investigators to identify the standard that would work best for them.  Toward this end, BioSharing.org has assembled information about standards, data sharing-related policies and data resources that aim to help investigators identify useful standards.

**Formulating, Conducting and Maintaining Community-Based Standards Efforts**
Workshop participants represented more than two dozen community-based standards efforts relating to a wide swath of biomedical data, including, but not limited to:  omics, structural biology, anatomical, clinical, behavioral, imaging, and a range of ontologies.  They represented large, longstanding standards organizations as well as smaller, more loosely organized groups. Some of these efforts are highly formal, while others are informal.  A number of questions were posed to the participants regarding the formulation, conduct and maintenance of community-based standards efforts; below is a distillation of themes that emerged as common answers across this very diverse collection of community-based endeavors.

• *How is the need for a particular standards effort identified?* Needs for data-related standards efforts are initially identified by a community of practice through a number of different ways, including encountering barriers to progress, active deliberation across the community, proactive engagement of and surveying the community of practice regarding unmet needs, and standards groups receiving suggestions from the community about unmet needs.  At the outset of such efforts a crucial question to consider is whether an existing standards solution can address the need (as is, or in slightly modified form) to avoid unnecessary duplication of effort.

• *What is the process used to assess and prioritize selected activities?*  A major driver of prioritization is the impact on enabling research that the effort will have on its own or in concert with related activities (i.e., leveraging other efforts).  Another determinant of priorities is how the activity matches the mission or goals of the organization.  Well described use cases were noted as being helpful in prioritization.

• *How do participants contribute to the standards effort?*  Typically, standards efforts are a team activity, or perhaps more correctly, a multidisciplinary team activity, with each team comprising a range of expertise and perspectives.  The core team ideally includes members who collectively have domain, technical and standards expertise.  While the manner in which the core team is chosen varies, the characteristics of its members include not only a range of expertise, but also a devotion to the goal.  This team engages with the community of practice (i.e., those who will use the standard in their research, tools or repository) through formal or informal means,  as well as with stakeholders who represent interests beyond those active in hands-on research (such as software developers, patient advocacy organizations, funders, publishers, etc.).  The interactions among these different groups are typically iterative as the effort moves forward, to allow feedback

to affect its course.  Feedback from those outside the core group is universally encouraged, though the manner in which feedback is received varies.  Engagement of the community of practice and other stakeholders often occurs in association with scientific or professional meetings, but should be ongoing.

• *What are the characteristics of the ongoing discussions/meetings?*  The frequency of meetings varies greatly depending upon the activity being undertaken.  Many efforts are conducted with meetings both virtual (conference call, video/voice over internet, wiki, email, etc.) and face-to-face; having some physical meetings is paramount.  Face-to-face meetings have a particularly large impact as the first meeting and/or as a meeting near the mid-way point of the particular effort.

• *Are milestones or similar indicators of progress used, and if so, how?* Larger and more formal efforts tended to use milestones more often and rigorously than smaller, less formal efforts.  These milestones were internal, such as comparison of progress with a timetable created at or near the outset, as well as external, such as comparison of deliverables with the dates of deliverables described in a contract supporting the effort.  When measures of progress included an indication of how successful adoption or use was, the lag between such adoption and use and public report (e.g., noting the use of a standard in a methods section of a journal article) was identified as a challenge.

• *How is the product of the standards effort updated and assessed?*  Many of the standards efforts used explicitly identified groups or committees to conduct the hands-on maintenance of the standard.  Most of the follow up and assessment of impact was informal and based on how widely the standard is used.

Regardless of the domain or data type, standards are important to allow data (and its contextual information), tools and data resources to work with each other.  When data-related standards come from the community of practice, broadly defined and actively engaged, their adoption and use is increased.  While details vary, many successful community-based standards efforts:  are driven by well-formulated needs of the community, include a core group of implementers with diverse expertise, engage a wide array of stakeholders from the outset and in an ongoing way, iteratively evolve based on broad input, and are recognized as long term and dynamic.

**Conclusion**
BD2K can help transform the biomedical research enterprise from one that is concept-centric to one that is more data-centric, and reap the value produced by that transformation.  This change would be hastened through the wide use of data-related standards.

To help investigators choose standards that would best serve their science, key information about available data-related standards is needed.  Such information would be most useful if it were easy to find, systematically organized and presented in a consistent way in a curated, trusted and publicly available resource.  For communities that have identified standards efforts with broad impact, having frameworks of policies, administrative procedures, award mechanisms, etc. would allow NIH to routinely catalyze such efforts and facilitate the broad use of data in that community.