

Initial Gap Analysis: Standards Hub and Biosharing Catalogue

Background

From www.biosharing.org, the Biosharing Catalogue is a resource for standards aiming to:

“ 1. centralize community-developed bioscience standards, linking to policies, other portals, open access resources and lists of tools and databases implementing the standards;

2. develop and maintain a set of criteria for assessing the usability and popularity of the standards, also the interoperability and relations among them;

3. foster interoperability, addressing overlaps and duplication of efforts that hamper their wider uptake and interfere with the creation of standards-compliant systems.”

The Standards Hub was proposed to develop a web based community resource:

- to centralize valuable and practical information around standards to facilitate adoption and prioritization
- to identify gaps and/or overlaps in standards
- that will evolve content and organization-wise by input from community

The hub was planned to be more than a dry list of standard bodies or standards and handle/advertise following content in addition to typical identifying information (e.g. name, description, organization):

- Adoption experience (contributed by “community” or curated)
 - Level of difficulty in adoption/costs associated with adoption
 - Prerequisites for adoption
 - Level/extent of adoption or maturity
 - Description of tools/processes associated with Standard (including ones that facilitate adoption)
 - References to peer-reviewed journals
 - Links to other resources Blogs, Abstracts, Forums
- Accessibility/availability/licenses associated with adoption
- Relationship among standards (leading creation of standard profiles)
- Related regulations and mandates (e.g., Meaningful use in US).

Introduction

This document is prepared as a follow-up to the meeting held on Jun,6, 2012 to discuss the overlap between Biosharing Catalogue and the proposed Standards Hub project. The meeting attendees were:

- Juli Klemm, NCI-CBIIT

- Sherri de Coronado, NCI-CBIIT
- Gilberto Fragoso, NCI-CBIIT
- Susanna-Assunta Sansone, University of Oxford e-Research Centre
- Philippe Roca-Serra, University of Oxford e-Research Centre
- Pascale Gaudet, Swiss Institute of Bioinformatics
- Baris Suzek, Georgetown University

The document provides the function and content gap between the Biosharing Catalogue and proposed Standards Hub project. This initial gap analysis can be used to identify potential next steps in how best leverage, repurpose or expand on the existing Biosharing Catalogue. Gaps potentially are the future developments planned for Biosharing Catalogue.

The next section provides potential gaps organized into two categories; information coverage , and organization/access mechanism.

Potential Gaps

Information Coverage

- Coverage of clinical, imaging and biomedical informatics standards. For example HL7 standards such as Clinical Document Architecture
- Coverage of national (specific to country) standards, such as US. For example OMB Standards for Data on Race and Ethnicity
- Relationships between standards.
 - Some exchange formats naturally use some terminologies. For example Generic Feature Format Version 3 use GO.
 - Some can be transformed to other with/without information loss. For example mzXML to mzML
 - Some predates or expected to be replaced by the other. For example mzData+mzXML=mzML
 - Some relationships could be a gap to start with which can also shown. For example UniProt TissList and SNOMED-CT
- A distinction between standard and used terminologies/formats. For example Uniprot Tisslist is not a standard.
- Level of maturity or adoption/use (e.g. are there nationwide projects using this standard?). BioDBCore-based information is not visible at standard summary.
- Tools or processes to facilitate adoption. For example RELMA for legacy Lab Term to LOINC mapping or tools for picking ontologies (http://www.bioontology.org/technology?magic_tabs_callback_tab=1)
- List of tools/resources to help transformation between formats/terminologies. For example Mass Spectrometry specific converters (<http://www.ms-utils.org/wiki/pmwiki.php/Main/SoftwareList>) or standard terminology mapping (SNOMED-CT to

ICD-9 by NIH/NLM for reimbursement purposes

(http://www.nlm.nih.gov/research/umls/mapping_projects/snomedct_to_icd9cm_reimburse.html)

- List of policies mandating use of standard. For example Meaningful Use rules around standards (http://healthit.hhs.gov/portal/server.pt?open=512&objID=1195&parentname=CommunityPage&parentid=97&mode=2&in_hi_userid=11673&cached=true)
- List of projects (not necessarily database) using standard. BioDBCore is a great start for databases but there could be tools/software or large projects (such as ENCODE, 1000genomes etc) as well. For example 1000genomes used formats (<http://www.1000genomes.org/data>)
- Past experience, best practices, lessons learned, critique around adoption, use, interactions with governing bodies, cost etc.. For example literature like “What is mzXML good for?” (<http://www.ncbi.nlm.nih.gov/pubmed/16307524>) or tool FAQ pages like <http://loinc.org/faq/mapping-local-codes-to-LOINC> or literature around terminology mapping <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1839416/>.

Organization/Access Mechanisms

- Simple text search and/or advanced text search (providing limits on fields such as type, description)
- Crosslinks between standards based on relationships
- Interactive maps to navigate standards
 - This was proposed as an alternative mechanism but the details were not well thought off. The idea was to make categorization easily visible/navigable. Possibly another ontology to categorized/classify and provide different entry points.