Standards to promote data interchange in the life sciences

George M. Garrity
Michigan State University
NamesforLife, LLC

Charles T. Parker
NamesforLife, LLC
Culture collections and BRCs

• Repositories
  • biological materials
  • data/information of materials in holding
• Potential and unrealized value
  • high potential
  • largely unrealized
• Barriers
  • expectations/needs of the market
  • cost of delivery
  • implementation details
• Not unique to culture collections or BRCs
Standards

• What are they?
  • definition
  • benefits
    • business
    • government
    • society
Standards

• How are created, implemented and maintained?
  • Response to market need
  • Expert global opinion
  • Mutual agreement among stakeholders
  • Community consensus
Setting standards (the ISO model)

• Proposed stage
• Preparatory stage
• Committee stage
• Enquiry stage
• Approval stage
• Publication
Standards setting bodies

• International/National
  • ISO/NISO/ANSI/DIN
    • ISO 9000, DOI

• Industry specific
  • W3C/IEEE/OMG

• Specialized
  • Codes of nomenclature
  • MIGS/MIMS/MIxS
A consortium of small collections

• Problems
  • bioresources hard to find unless...
  • large collections are the “go-to” suppliers
    • marketing
    • size
    • publication hegemony

• Goals
  • improve visibility of smaller collections
    • increase business
    • increase citation frequency
  • facilitate cooperation between collections
A consortium of small collections (cont.)

• Proposed solutions
  • create standard to reduce the cost of staying on top of web and search
• An Open Collection standard
  • URL conformance
  • collection description standard
  • metadata standard for individual resources
URL conformance

• Collection accessions should map consistently into URLs
  
  http(s)://www.my-agency.org/{COLLECTION}/{ACCESSION}

  where (COLLECTION) is a string representing individual collection with an institution.

  where (ACCESSION) is URI-mapped accession to that collection

• Rationale - provides support of multiple collections within a single organization, maps directly to Search Engine Optimization, and to REST
Examples from actual collections

http://www.atcc.org/Products/All/{ACCESSION_SUFFIX}.aspx

ATCC accessions are “ATCC 3944”, URL leaves off collection ID.

https://www.dsmz.de/catalogues/details/culture/{ACCESSION_PREFIX}-{ACCESSION_SUFFIX}.html

DSMZ accessions are “DSM 7”, URL form is “DSM-7”.

Many smaller collections have accessions, but offer no way to link directly to a resource:

```plaintext
```

Above example uses Google Search to find resources on their web site.
• Many smaller collections have accessions, but offer no way to link directly to a resource:

   http://fat.org.br/catalogo-de-culturasy?termo=&letra=c

This only links via the first letter of the “name” associated with the strain. This is only one view of a collection, and it ignores the fact that names change. This will leave many resources “stranded” in the collection, not findable.
Collection Description Standard

• Document appropriate contacts at the organization level
  • provenance
  • country info, links to appropriate MTA forms
• Document individual collections at an organization
  • accession->URI mapping.
  • availability.
  • updated contact information.
  • short description of collection.
• If merged include mappings between accessions
• Create a schema.org description
Metadata standard for resources

• Provide standard method of documenting source material
  • employ other source collection URI mapping schemes were possible
  • employ other standards (ORCID, Research-ID, N4L)
• Document available phenotypic data consistently.
  • employ other standards (SKOS, ontologies)
This work was funded through the Small Business technology transfer program of the United States Department of Energy under grants number DE-FG02-07ER86321 and DE-SC0006191. Funding for business development was provided through grants and loans from the Michigan Economic Development Corporation and the Michigan Universities Commercialization Initiative. NamesforLife semantic resolution technology is covered under US Patent 7,925,444 B2. The SOSCC systems and methods is covered under US Patent 8,036,997. Semantic markup technology and semiotic fingerprinting are subject of pending US and EPO patent applications.