

Standards Inventory Resource: Standards Hub Proposal

**Baris Suzek
Brian Davis**

February 15, 2012

Outline

- **Background**
 - What is my use case?
 - What is the motivation?
- **Initial Survey**
 - Community questions around standards
 - Current resources
- **Objective**
- **Implementation Ideas**
 - Content and Organization
 - Resources
- **Conclusion and Challenges**
- **Discussion**

Background

- **John Speakman:** What would be the role of workspaces to help community?
- **My use case:**
 - Support cohort or research data discovery in multi-site consortium
 - Semantic normalization of lab names needed
 - What is the target standard terminology? LOINC
 - What are the tools/procedures for transforming “legacy values” to LOINC?
 - What are the procedures/mechanisms tested/used?
- **Available resources:**
 - People, Google/Bing, LOINC pages
 - caBIG community built in years

Motivation

- **Many types of standards**
 - Vocabulary, Interface, Format, Messaging, Documentation ... etc.
- **Standards as practical /adapted vs. academic exercises**
 - Limited adaption by geographical, institutional context
 - Never adapted or evolved before adaption
- **Standard adapter and/or governance bodies often face the questions:**
 - What?
 - Are there standards (or even candidates) that can or should be used?
 - Is there a gap to fill?
 - Why?
 - What is the value proposition for adopting a standard? Does it have community buy-in?
 - Are there rules/policies that I need to watch for (e.g. meaningful use)
 - How?
 - What is the path (tools/process) to adopting this standard? e.g. are there transformations for the terminologies of interest?
 - What are the paths tested/well-defined?

Initial Survey - Community questions around standards

- “A person is looking for a distinct canonical representation of demographics (race/religion/country of birth). What would be alternative options for representing this information? What are available mappings? ”
- “Are there any ontologies that are suitable for annotation of transcriptomics data from a Nuclear Receptor Signaling Project?”
- “There are so many standards, I don’t know where to start”
- “ I am using an in-house, legacy system based on homegrown vocabulary for Lab data that I call “Lab test name”, “Lab value”, “Blood count”, “Oxygen content”, etc. How do I transition from the legacy system to LOINC?”
- “How can we create standards profiles to use in compliance? How can we guide people in using standards without knowing what is available?”
- “I have a specification for a software application from Stanford. I want to compare it to NCI CBIIT software specifications for LexEVS to see that they’ll be interoperable for some of the data they are going to pass to NCI CBIIT. What standards should I be looking at to make sure they are compliant?”
- “What are the options for me to represent Protocol Lifecycle objects: as an ontological representation or a UML representation or some other representation? What do you have that I can leverage and how?”
- “Where can I submit my commentaries or see others’ about a standard?”

Initial Survey – Current Resources

- **Google**
- **SMEs**
- **NLM**
- **NCBO/Bioportal**
- **Standard organizations**
- **Blogs/literature**
- **Similar projects**

Initial Survey – Similar Projects

Cancer InfoMatrix



The Cancer InfoMatrix (CIM) provides information about work being conducted to support good data management in cancer research through the development of [data standards](#). Biomedical domains are arranged across the horizontal axis of the CIM while on the vertical axis are broad areas of data management that enable data sharing, access and exchange. Clicking on a cell within the CIM will take you to a list describing each initiative in that area as well as other relevant information.

The cells in the Cancer InfoMatrix are colour coded using a "traffic-light-battery system" to indicate the degree of progress in each area:

You can make suggestions for additions to the CancerInfo Matrix by [contacting us](#).

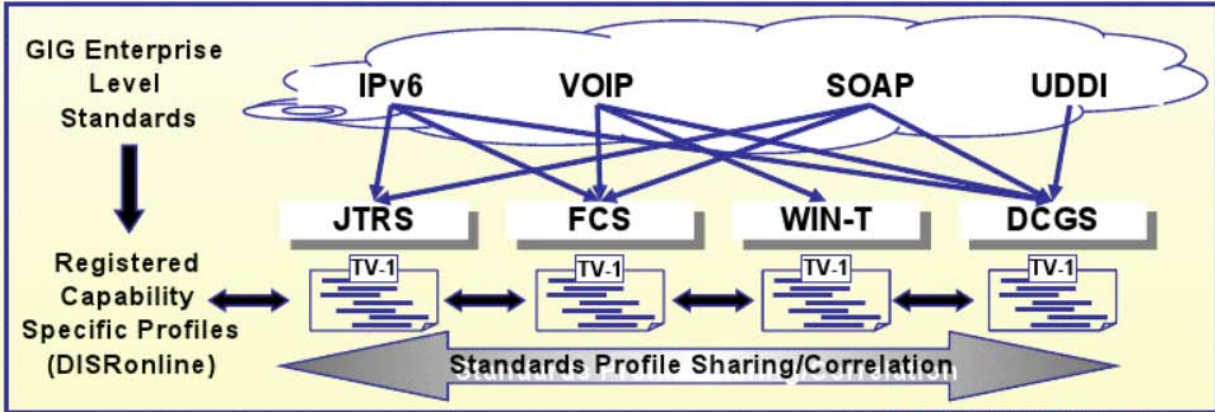
	DNA	RNA	Proteins	Metabolites	Tissue	 Models	 Systems	Images	Clinical	Population
Data Reporting Guidelines										
Ontologies & Controlled Vocabularies										
Data Exchange Formats										
Standards Tools										

Initial Survey – Similar Projects

 **UNCLASSIFIED** 

DISRonline Capability

- DISRonline is one DoD's **only electronic WWW database of IT standards** that provides the minimal set of information technology standards adopted by Services, Agencies, and Combatant Commands consensus to enable Joint/Combined interoperability
- DISRonline database structure ensures that PMs use a **structured process to develop their system's IT Standards Profile (TV-1)** before JCIDS documents are submitted to the Gatekeeper IAW CJCSM 3170.01



The diagram illustrates the structure of DISRonline. At the top, a cloud contains four enterprise-level standards: IPv6, VOIP, SOAP, and UDDI. Below this, four project-specific standards are listed: JTRS, FCS, WIN-T, and DCGS. Each project standard is linked to a 'Registered Capability Specific Profile (DISRonline)' represented by a document icon with 'TV-1' on it. A large double-headed arrow at the bottom indicates 'Standards Profile Sharing/Correlation' between these profiles. A vertical arrow on the left points from 'GIG Enterprise Level Standards' down to the 'Registered Capability Specific Profiles (DISRonline)'.

UNCLASSIFIED

Initial Survey – Similar Projects



[POLICIES](#) |
 [STANDARDS](#) |
 [BLOG](#) |
 [CONTACT](#) |
 [COMMUNITIES](#) |
 [ABOUT](#) |
 [LOGIN OR REGISTER](#)

Home

A CATALOGUE OF STANDARDS

You can **sort** columns and **browse** the reporting guidelines content, or you can view [all the standards](#), or [reporting guidelines](#), or [terminological artifacts](#) or [exchange formats](#) only; or go back to the [catalogue main page](#).

BioSharing ID	STANDARD	FULL NAME ▲	TYPE	DOMAIN(S) COVERED	MAIN PUBLICATION	ORG
bsg-000297	.ACE format	.ACE format	exchange format	sequence assembly		
bsg-000208	Abstract Syntax Notation One	Abstract Syntax Notation One	exchange format	syntax		ISC
bsg-000197	ABCD	Access to Biological Collection Data	exchange format	biological diversity		TD
bsg-000204	ABCDDNA	Access to Biological Collection Data DNA extension	exchange format	biological diversity; DNA collection		TD
bsg-000219	Affymetrix Raw Intensity Format	Affymetrix Raw Intensity Format	exchange format	affymetrix microarray raw data		Aff
bsg-000019	AAO	Amphibian gross Anatomy Ontology	terminology artifact	anatomy (amphibian)		AA
bsg-000005	ATO	Amphibian Taxonomy Ontology	terminology artifact	anatomy		AT
bsg-000001	ADaM	Analysis Data Model	reporting guideline	analysis datasets; clinical; non clinical (regulatory submission)		CD
bsg-000004	AEO	Anatomical Entity Ontology	terminology artifact	anatomy		
bsg-000035	ARRIVE	Animals in Research: Reporting In Vivo Experiments	reporting guideline	animal research	Kilkenny et al; PLoS Biol; 2010	NC
bsg-000207	ARLEQUIN Project Format	ARLEQUIN Project Format	exchange format	population genetics information		
bsg-000109	AMIS	Article Minimum Information Standard	reporting guideline	literature standard	Tassy et al; Genome Res; 2010	AN
bsg-000128	APO	Ascomycete Phenotype Ontology	terminology artifact	phenotype		AP
	Axt Alignment		exchange			

Objective

- **An organized hub of standards for clinical, biomedical informatics and bioinformatics**
 - To facilitate standard adoption by developers
 - To assist prioritization of "high-recommended" standards for compliance at organizations (creation of standard profiles)
- **Addressing issues around**
 - Multiple places to find standards. Particularly challenging in the life sciences space as there are fewer of the large organizations like HL7
 - Competing/complementary standards
 - Difficulty accessing "maturity", "benefits", "implementation costs" and knowing the "tooling/best practice/process availability"
- **An organization to collect valuable community input**
 - SMEs, Researchers, IT specialists, Standards bodiesetc.

Implementation Ideas - Content

- **Content**
 - Standard Name
 - Standard Description
 - Authoritative resource for this standard (eg, URL).
 - **Adoption Experience (some contributed by “community”)**
 - Level of difficulty in adoption/costs associated with adoption
 - Prerequisites for adoption
 - Level/extent of Adoption
 - Description of tools/processes associated with Standard
 - Accessibility/availability/Licenses associated with adoption
 - Relationship among standards
 - Regulations (eg, Meaningful use).
 - References to peer-reviewed journals
 - Links to other resources Blogs, Abstracts, Forums

Implementation Ideas - Organization

- **Templates**
- **Multiple perspectives**
 - Domain of interest
 - Type of standard
 - .. others based on use cases and identified standards
- **Multiple entry points**
 - Tables (e.g. Infomatrix)
 - Text search
 - Mind/concept/interaction maps to navigate space

Implementation Ideas - Resources

- **Use cases from stakeholders:**
 - ICR Workspace
 - Imaging Workspace
 - Tissue Banks and Pathology Tools Workspace
 - CTMS Workspace, including the Population Sciences SIG
 - NCI CBIIT Architecture Review Board
 - Other community members
- **Initial resources to mine:**
 - Some existing inventory of Standards and Standard Bodies in caBIG (see <https://wiki.nci.nih.gov/x/Lh4hAQ>)
 - Existing Standards outlined by ARB (see https://docs.google.com/spreadsheets/ccc?key=0A0oguv4DzhdmdEd1aDcwQTE5TEVLM1RiLVZtQ1JYVIE&hl=en_US#gid=0)
 - caDSR metadata standards (see <https://cabig.nci.nih.gov/concepts/caDSR/>)
 - Vocabulary Standards (e.g., see NCBO BioPortal: <http://bioportal.bioontology.org/> see NCI Thesaurus: <http://ncit60.nci.nih.gov/>)
 - caGrid Standards (e.g., see <http://cagrid.org/display/projects/Home>)
 - Federal Registry
 - DoD and NCRI Resources
 - Biosharing Catalogue

Conclusion

- **A proposal for a web based community resource**
 - to centralize valuable and practical information around standards to facilitate adoption and prioritization (Shorter term)
 - also to identify gaps and/or overlaps in standards (Longer term)
- **A proposal for a community resource that will evolve**
 - Content, organization, governance... etc.
- **It should be more than a dry list of standard bodies or standards**

Possible Challenges

- **Scale and scope**
- **Expectation management and prioritization**
- **Governance**
- **Noise**
- **Resources availability**
- **Time**
- **More...**

Discussion

Backup