Role of standards organizations in the USA and their impact on knowledge representation in the global context

Joseph Busch
Agenda

- U.S. government standards policy
- Types of U.S. standards organizations
- Standards activity drivers
- Standards and knowledge graphs
United States intellectual property and privacy rights

- The Constitution and its Amendments
  - “… promote the progress of science and useful arts, by securing, for limited times, to authors and inventors, the exclusive right to their respective writings and discoveries.” – Article 1, Section 8
  - The right of privacy against unreasonable searches and seizures by the government – Fourth Amendment
  - The right against self-incrimination, which justifies the protection of private information – Fifth Amendment

Constitution of the United States. NARA. Public domain.
Consensus standards policies

- U.S. government policies related to
  - Use and development of voluntary consensus standards.
  - Standards conformity assessment.

---

CIRCULAR NO. A-119 Revised
February 10, 1998

MEMORANDUM FOR HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES

SUBJECT: Federal Participation in the Development and Use of Voluntary Consensus Standards and in Conformity Assessment Activities

Revised OMB Circular A-119 establishes policies on Federal use and development of voluntary consensus standards and on conformity assessment activities. Pub. L. 104-113, the “National Technology Transfer and Advancement Act of 1995,” codified existing policies in A-119, established reporting requirements, and authorized the National Institute of Standards and Technology to coordinate conformity assessment activities of the agencies. OMB is issuing this revision of the Circular in order to make the terminology of the Circular consistent with the National Technology Transfer and Advancement Act of 1995, to issue guidance to the agencies on making their reports to OMB, to direct the Secretary of Commerce to issue policy guidance for conformity assessment, and to make changes for clarity.

National Technology Transfer and Advancement Act, OMB Circular A-119
Cultural heritage

- Officially there is no U.S. culture department

**US Department of Interior**
- Manages largest museum system
- 73 million+ objects
- 86,000 linear feet of archives

**Library of Congress**
- Not officially the U.S. national library but often assumes this role
- Primary mission is to research inquiries for members of Congress
Agenda

- U.S. government standards policy
- Types of U.S. standards organizations
- Standards activity drivers
- Standards and knowledge graphs
## Types of data and information standards organizations in the United States

<table>
<thead>
<tr>
<th>Type of Organization</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Government agencies</td>
<td>Library of Congress, NIST</td>
</tr>
<tr>
<td>Membership non-profits</td>
<td>Dublin Core, NISO</td>
</tr>
<tr>
<td>Private non-profits</td>
<td>Research Institute</td>
</tr>
<tr>
<td>Industry trade associations</td>
<td>SEMI, SIIA</td>
</tr>
</tbody>
</table>
LoC has been a global leader in bibliographic description, controlled vocabularies, and classification – all essential components of reference data.

- Resource description formats
- Digital library standards
- Information resource retrieval protocols
- Information resource retrieval standards
- Controlled vocabularies
Accurate analysis methods, to calibrate measurement systems

NIST TREC has been instrumental in facilitating development of automated categorization of text and image content.

For scientific disciplines including:
- Atomic and molecular physics
- Chemical and crystal structures
- Fluids
- Material properties
- Biotechnology
- Optical character recognition
- …
Text REtrieval Conference (TREC) https://trec.nist.gov/

- Ongoing series of workshops focusing on different information retrieval (IR) research areas.
- Cosponsored by NIST and Intelligence Advanced Research Projects Activity (IARPA)
- Data sets include
  - Legal
  - Medical
  - Chemical
  - Genomics
  - ...
Dublin Core

- Dublin Core is a widely used schema for describing web content, using RDF vocabularies, packaged in application profiles.
- Originated at an OCLC workshop, and hosted by OCLC until 2009.

Dublin Core has become the de facto standard for digital content description in the public web and on private commercial clouds.

- Internet Engineering Task Force standard RFC 5791 (2010) Encoding Dublin Core Metadata in HTML
- ANSI/NISO Z39.85-2012 The Dublin Core Metadata Element Set
- ISO 15836-1:2017 Core Elements
- ISO 15836-2:2019 Properties and Classes
NISO

- Develops, maintains and publishes technical standards related to publishing, bibliographic and library applications.

NISO has been instrumental in promoting foundational library standards to enable information description and protocols for sharing bibliographic information systems and services.

- Z39.5: Abbreviation of Titles in Periodicals
- Z39.7-2013: Information Services and Use
- Z39.29-2005 (R2010): Bibliographic References
- Z39.50-2003 (S2014): Information Retrieval
- Z39.84-2000: Digital object identifier (DOI)
- Z39.85-2012: Dublin Core
- Z39.87-2006 (R2017): Digital Still Images
- Z39.88-2004 (R2010): OpenURL
- Z39.96-2015 JATS: Journal Article Tag Suite
- Z39.102 -2017: Standards Tag
The Getty Research Institute

- Getty Vocabularies are available as Linked Open Data, XML, Relational Tables, and through APIs.
- Components for community based terminology management
  - Editorial guidelines
  - Training materials
  - Authorized contributors
  - International Terminology Working Group

Getty Vocabularies are widely used as the basis for semantic interoperability in the culture sector.
Industry association representing electronics manufacturing and design supply chain.

Location: Milpitas, CA, USA (Silicon Valley)

More than 1,000 standards and guidelines related to all aspects of automated fabs including:
- Documentation & training
- Terminology
- XML

The ability to communicate documentation so that equipment can be used and maintained is critical for the economic viability of manufacturing and other sectors.
Software & Information Industry Association

- Trade association for the entertainment, consumer and business software industries.

- Public policy
  - Intellectual property protection
  - Privacy and data security

- Advocacy
  - Data driven innovation (business intelligence, data analytics, big data)
  - AI and automation (future of work, ethics, fairness)

- Best practice recommendations
  - Derived data products
  - Usage of data
  - Service levels
  - Billing and invoicing
  - Contracts
  - Vendor scorecards

Effective intellectual property policies are critical to economic viability of creating information products and services.
Agenda

- U.S. government standards policy
- Types of U.S. standards organizations
- Standards activity drivers
- Standards and knowledge graphs
### Standardization in relation to information management, publishing, reference data, and semantic interoperability

<table>
<thead>
<tr>
<th>Organization</th>
<th>Standards Types</th>
<th>Info Mgmt</th>
<th>Publish</th>
<th>Ref Data</th>
<th>Sem Interop</th>
</tr>
</thead>
<tbody>
<tr>
<td>LoC</td>
<td>Vocabularies, schema, protocols</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>NIST TREC</td>
<td>Data sets</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DCMI</td>
<td>Vocabulary, schema, encoding</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>NISO</td>
<td>Schema, vocabularies, protocols</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Getty</td>
<td>Vocabularies, schema, tools</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>SEMI</td>
<td>Documentation, terminology, schema</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>SIIA</td>
<td>Best practices, briefs, whitepapers</td>
<td>✓</td>
<td>✓</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Glossary
- **Vocabulary.** A way to organize knowledge for retrieval including value sets and schema labels.
- **Schema.** Description of how to model real world entities as objects in a database.
- **Protocol.** A system of rules that allow two or more entities to effectively communicate.
- **Encoding.** The process of converting data from one form to another.

- **Information Management.** The process of collecting, managing, preserving, storing and delivering information.
- **Publishing.** The activity of making information and other content publicly available.
- **Reference Data Management.** The process of managing classifications and hierarchies across systems.
- **Semantic Interoperability.** The ability of computer systems to exchange data with unambiguous, shared meaning.
Standards activity drivers

- Standards have an economic impact on operations.
  - GDPR – Most large US companies with EU operations comply or aim to comply with the General Data Protection Regulation (GDPR).
  - ISO/IEC 2700 family of standards to keep information assets secure.

- Standards or best practices have emerged at the right time and become widely adopted
  - Dublin Core is the de facto standard for web content description.

- Standards that facilitate data interchange
  - FpML (Financial products Markup Language) by the International Swaps and Derivatives Association (ISDA) enables business-to-business over-the-counter financial derivative transactions online by following W3C standards.
  - SKOS (Simple Knowledge Organization System) and OWL (Web Ontology Language) for concepts and relationships are enabling standards for semantic interoperability.
Agenda

- U.S. government standards policy
- Types of U.S. standards organizations
- Standards activity drivers
- Standards and knowledge graphs
What is a knowledge graph?

- Representation of an organization’s knowledge assets, content, and data—people, places, documents, multimedia, data, etc.
What is a Knowledge Graph?

- Representation of an organization’s knowledge assets, content, and data—people, places, documents, multimedia, data, etc.—and how these things are related to each other.
What is a knowledge graph?

- Framework that
  - Defines the **things** – people, documents, data, multimedia, etc.
    - Classes of things
    - Subclasses of things
  - Defines **properties** to describe a class of things or a subclass of things, ex: Name
  - Defines **relationships** between the things or between the classes and subclasses

- Typically, this is an **ONTOLOGY** that defines classes for the things, properties for the things, and relationships between the things.

- **Knowledge Graph = an ontology + instances**
Knowledge Graph = Ontology + Instances

THINGS with RELATIONSHIPS!

- Nancy, France
- Metz, France
- Patrice
- Nov 2020 sales Report
- Enterprise Cloud Storage
- 2020 Sales Reports
- Executive Presentation
- Lives in
- Has work location
- Created
- Stored
- Part of
- Talked About
Same concept, two knowledge graphs – Fluid Dynamics

**PhySH - An ontology for the physics domain**

**Google knowledge graph for the same concept**
### Types of knowledge graphs

<table>
<thead>
<tr>
<th>Type</th>
<th>Content</th>
<th>Goal</th>
</tr>
</thead>
<tbody>
<tr>
<td>Knowledge base</td>
<td>Repository of structured and unstructured information.</td>
<td>Discover and manage resources so they can be used to reason about and draw conclusions about the world.</td>
</tr>
<tr>
<td>Social network</td>
<td>Social structure determined by the interactions between individuals, groups, and organizations.</td>
<td>Identify patterns, locate influential entities, and examine network dynamics.</td>
</tr>
<tr>
<td>Data catalog</td>
<td>An organized inventory of the data assets in an organization.</td>
<td>Discover, use, interpret, and govern data entities.</td>
</tr>
<tr>
<td>Combinations</td>
<td>Combinations of content.</td>
<td>Combinations of goals.</td>
</tr>
</tbody>
</table>
Common uses for knowledge graphs

<table>
<thead>
<tr>
<th>Uses</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enhanced search results</td>
<td>Google</td>
</tr>
<tr>
<td>Personalized search results</td>
<td>Facebook, Google</td>
</tr>
<tr>
<td>Entity linking</td>
<td>Amazon</td>
</tr>
<tr>
<td>Recommender systems</td>
<td>Amazon, Netflix, Spotify</td>
</tr>
<tr>
<td>Ad targeting</td>
<td>Google, Facebook</td>
</tr>
<tr>
<td>Enhanced data analytics</td>
<td>Any business</td>
</tr>
</tbody>
</table>
There are not a lot of knowledge graphs available off the shelf … but there are many ontologies available (think framework)

<table>
<thead>
<tr>
<th>DDC</th>
<th>EuroVoc</th>
<th>Language</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

704 records found.

*EuroVoc is a multilingual, multi-specialised thesaurus covering the activities of the EU, the European Parliament in particular. It contains terms in 25 EU languages: Bulgarian...*
Commonly used public knowledge graph data resources (for instances)

<table>
<thead>
<tr>
<th>Knowledge Graph</th>
<th>Domain</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>DBpedia</td>
<td>Cross-domain</td>
<td>Wikipedia</td>
</tr>
<tr>
<td>Wikidata</td>
<td>Cross-domain</td>
<td>Wikipedia, (Metaweb Freebase)</td>
</tr>
<tr>
<td>Google Knowledge Graph</td>
<td>Cross-domain</td>
<td>Web data</td>
</tr>
<tr>
<td>Facebook Entities Graph</td>
<td>Cross-domain</td>
<td>Wikipedia, Facebook data</td>
</tr>
<tr>
<td>The Linked Open Data Cloud</td>
<td>Cross-domain</td>
<td>Various</td>
</tr>
</tbody>
</table>

Standards are the foundation for knowledge graphs

<table>
<thead>
<tr>
<th>Standard</th>
<th>Type</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>RDF (Resource Description Framework)</td>
<td>Protocol</td>
<td>Standard model for data interchange on the web based on triples and persistent identifiers (URIs)</td>
</tr>
<tr>
<td>RDFS (Resource Description Framework Schema)</td>
<td>Schema</td>
<td>Set of classes with properties for the description of ontologies, e.g., a representation of the Dublin Core metadata terms, or SKOS vocabularies.</td>
</tr>
<tr>
<td>Dublin Core</td>
<td>Vocabulary</td>
<td>Fifteen properties (metadata elements) used to describe resources.</td>
</tr>
<tr>
<td>SKOS (Simple Knowledge Organization System)</td>
<td>Vocabulary</td>
<td>Used to represent, publish and use thesauri, classification schemes, taxonomies, subject-heading systems, or any other type of structured controlled vocabulary as linked open data.</td>
</tr>
<tr>
<td>OWL (Web Ontology Language)</td>
<td>Vocabulary</td>
<td>Knowledge representation language for describing taxonomies and classification networks.</td>
</tr>
<tr>
<td>Linked Open Data (LOD)</td>
<td>Protocol</td>
<td>Use of RDF and URIs to interlink structured data on the web.</td>
</tr>
</tbody>
</table>
Summary

This icon is licensed as Creative Commons CCBY
Questions

❖ Joseph A. Busch, (415) 377-7912, jbusch@taxonomystrategies.com
US Standards Bibliography

- Dublin Core Metadata Initiative. [https://www.dublincore.org/](https://www.dublincore.org/)
- Financial products Markup Language (FpML) [https://www.fpml.org/](https://www.fpml.org/)
- General Data Protection Regulation (GDPR). [https://gdpr-info.eu/](https://gdpr-info.eu/)
- Getty Vocabulary Program. [https://www.getty.edu/research/tools/vocabularies/](https://www.getty.edu/research/tools/vocabularies/)
- Library of Congress. [https://www.loc.gov/](https://www.loc.gov/)
- SEMI. [https://www.semi.org/en](https://www.semi.org/en)
- Simple Knowledge Organization System (SKOS) [https://www.w3.org/2004/02/skos/](https://www.w3.org/2004/02/skos/)
- Software and Information Industry Association (SIIA). “About SIIA.” [https://www.siia.net/About/About-SIIA](https://www.siia.net/About/About-SIIA)
- Web Ontology Language (OWL) [https://www.w3.org/OWL/](https://www.w3.org/OWL/)
Knowledge Graph Bibliography

- OWL 2 Web Ontology Language. [https://www.w3.org/TR/owl2-overview/](https://www.w3.org/TR/owl2-overview/).
- RDF. [https://www.w3.org/RDF/](https://www.w3.org/RDF/).
- RDF Schema 1.1. [https://www.w3.org/TR/rdf-schema/](https://www.w3.org/TR/rdf-schema/).
- SPARQL 1.1 Query Language. [https://www.w3.org/TR/sparql11-query/](https://www.w3.org/TR/sparql11-query/).