

Standards for Taxonomies and Linked Data

+1-505-998-0800

www.accessinn.com

www.dataharmony.com

mhlava@accessinn.com

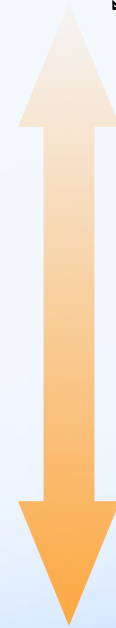
Marjorie M K Hlava, President
Access Innovations, Inc.
Albuquerque, NM



A Taxonomy is a Knowledge Organization System (KOS)

- Uncontrolled list
- Name authority file
- Synonym set/ring
- Controlled vocabulary
- *Taxonomy*
- *Thesaurus*
- Ontology
- Semantic network

Not complex



Highly complex

Structure of Knowledge Organization Systems

Lists Synonyms Taxonomy Thesaurus Ontology Linked data

INCREASING COMPLEXITY and CONTROL



Ambiguity

Ambiguity
Synonym

Ambiguity
Hierarchy

Ambiguity
Synonyms
Hierarchy
Relationships

Specifies
Directions in
Relationships

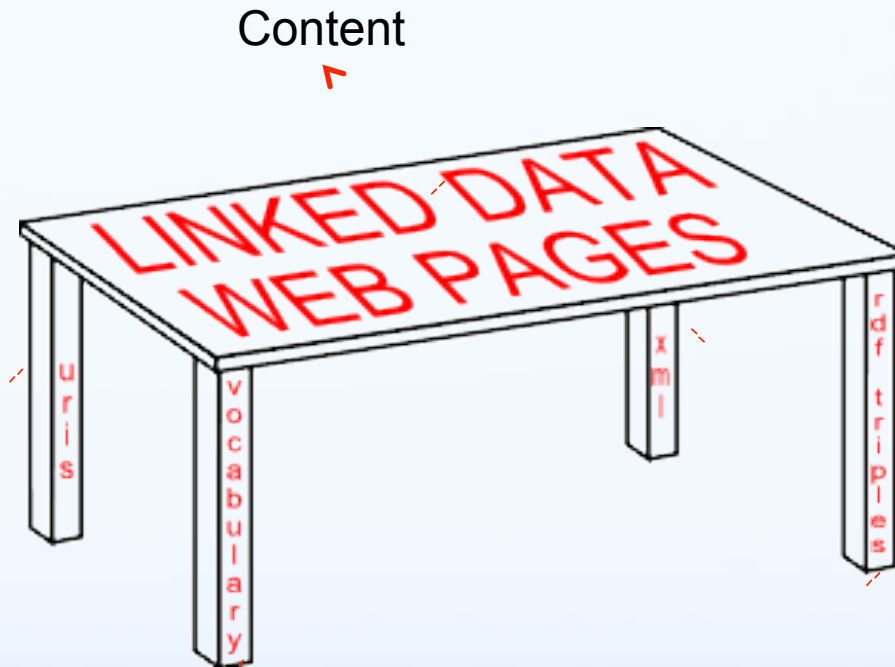
Embedded
Info
Objects

Standards to Enrich and Link Data

- Vocabulary
- XML's
- Export
- Web Standards
- Nice to have
 - Metadata
 - Other terminology
- Search

Like the legs on a table

Persistent links
To external sources



Structured text

Defined relationships
Between objects

Taxonomy, Thesaurus,
Or Ontology

English Language Taxonomy Standards

- ISO 25964 Thesauri and interoperability with other vocabularies
 - Part 1 (2011): Thesauri for information retrieval
 - Part 2 (2013): Interoperability with other vocabularies
 - International Standards Organization
- ANSI / NISO Z39.19 (2010)
 - American national Standards Institute
 - National Information Standards Organization
- BS 8723-4:2007,
 - *Structured vocabularies for information retrieval. Guide. Interoperability between vocabularies.*
 - British Standards Institute

Formats

- Metadata formats
- Data formats
- Export formats
- File Exchange formats
- Coding formats

XML Standards – Data Formats

- XML
- RDF
- JATS
- Other ML's
 - Mark Up languages
 - Math ML
 - Chem ML
 - Etc.

Export formats

- MARC XML
 - Machine Automated Record Cataloging
 - Extensible Mark up Language
- OWL
 - Web Ontology Language
 - DL, Lite, Full
- SKOS
 - Simplified Knowledge Organization System
- CSV
 - Comma Separated Values – like Excel
- JATS
 - Journal Article Tagging System
 - NLM DTD

OWL and SKOS

- OWL 2
 - *Web Ontology Language: Document Overview*. W3C Recommendation, 27 October 2009.
<http://www.w3.org/TR/owl2-overview/>
- SKOS
 - *Simple Knowledge Organization System Reference*. W3C Recommendation, 18 August 2009.
<http://www.w3.org/TR/skos-reference>
- SKOS-XL
 - *Simple Knowledge Organization System eXtension for Labels (SKOS-XL) Namespace Document - HTML Variant*. W3C Recommendation Edition, 18 August 2009.
<http://www.w3.org/TR/skos-reference/skos-xl.html>

Web Standards - 1

- HTTP /2.0
 - RFC 7540 - May 2015
- URI - Uniform resource identifier (URI)
RFC 3986
 - Uniform resource locator (URL) web address
 - Uniform resource name (URN) ID resources, i.e. namespaces
 - RDDL (Resource Directory Description Language)
 - Information about a particular namespace

Web Standards - 2

- PURL -Persistent uniform resource locator
 - DOI – Digital Object Identifier
 - CrossRef – publisher deposits
 - DataCite – research sets
 - mEDRA - Multilingual
 - Airti Inc .Traditional Chinese
 - Wanfang Data - Chinese
 - LSIDs Life Sciences Identifiers
 - Info URI
- RDF – Resource Description Framework

Web Standards – RDF - 1

- Classical conceptual modeling approaches
 - entity relationship or class diagrams
- Statements about resources = Triples
 - Subject – object – predicate expressions
 - Subject denotes the resource
 - Predicate gives traits or aspects of that resource links the two

Web Standards – RDF - 2

- Object oriented systems (Java)
 - Entity attribute value model
 - Triples
- Relational database SQL
 - Entity attribute value model
 - Triple stores
- Several serialization formats (file formats)
 - Triple or resource coding varies with format.

RDF File Formats

- **The way to share data between applications**
- **Turtle**,^[12] a compact, human-friendly format
- **N-Triples**,^[13] a very simple, easy-to-parse, line-based format that is not as compact as Turtle
- **N-Quads**,^{[14][15]} a superset of N-Triples, for serializing multiple RDF graphs
- **JSON-LD**,^[16] a JSON-based serialization
- **N3** or Notation3, a non-standard serialization that is very similar to Turtle, but has some additional features, such as the ability to define inference rules
- **RDF/XML**,^[17] an XML-based syntax that was the first standard format for serializing RDF

Coding Formats

- Relational Database
 - Tables
 - SQL
 - SPARQL
 - Triple-stores
- Object oriented programs
 - Each resource as an object with links
 - Triples

Metadata Standards

- MADS
 - US Library of Congress. *MADS (Metadata Authority Description Schema)*. Washington, DC: Library of Congress. Available at <http://www.loc.gov/standards/mads/>
- Dublin Core
 - IETF [RFC 5013](#)
 - ISO Standard 15836-2009
 - NISO Standard Z39.85

Other Terminology Standards

- ISO 704:2009
 - *Terminology work — Principles and methods*
- ISO 1087-1:2000
 - *Terminology work — Vocabulary — Part 1: Theory and application*
- ISO 15489-1
 - *Information and documentation — Records management — Part 1: General*
- ISO/TR 15489-2
 - *Information and documentation — Records management — Part 2: Guidelines*
- ISO 30042:2008
 - *Systems to manage terminology, knowledge and content — TermBase eXchange (TBX)*

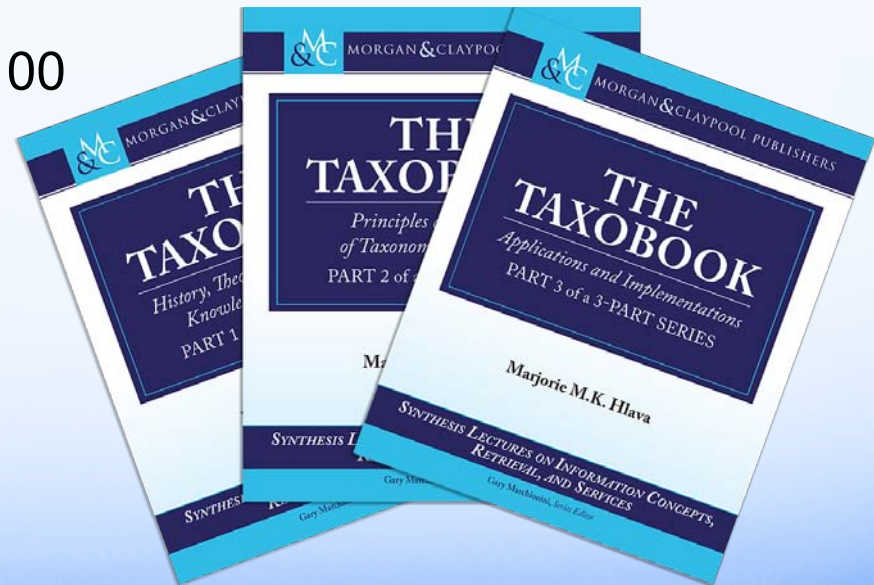
Many options

- Still evolving
- What do I use today?
 - NISO Z39.19 taxonomy
 - Java – object oriented system
 - Triples
 - Json
 - Contributed Metadata set
 - Dublin Core for DOI

Questions?

Thanks!

Marjorie M K Hlava, President
Access Innovations, Inc.
4725 Indian School Rd, Ste. 100
Albuquerque, NM 87110
+1-505-998-0800
www.accessinn.com
www.dataharmony.com
mhlava@accessinn.com



Slides will be posted to my website and the conference site.
Or give me a business card to receive via email



DataHarmony
from Access Innovations, Inc.