Insuring compatibility and crosswalks

Marjorie M. K. Hlava

Access Innovations / Data Harmony

mhlava@accessinn.com

The Problem – KEEPING UP

- Many players we know and don't know
- Between controlled vocabulary standards
  - ISO 2788 and 5964,
  - BSI 8723
- Groups developing guidelines and standards
  - W3C with SKOS and OWL
  - Governments world wide developing and mandating taxonomies
- Communities
  - increase reuse
  - mapping interoperability between controlled vocabularies.

Traditional Standards

- ISO
    - TC 46
        - SC 9
    - ANSI
        - NISO
            - Z39.19
    - BSI
        - BS 8723
- W3C
    - OWL
    - SKOS
- US Government
    - Office of Management and Budget
- European Union

Thesaurus related

- NISO Z39.19 2006 www.niso.org
- BSI (BS 8723) the next revised ISO
- ISO 2788 - Monolingual (1986)
- ISO 5964 - Multilingual (1985) www.iso.ch/iso/en/ISOOnline.frontpage
- ISO 5127, *Information and documentation* — *Vocabulary*
- OWL from W3C
- SKOS the W3C thesaurus standard

DataHarmony

Access Innovations.

ISO TC 46/SC 9

- Information and Documentation - Identification and Description
- **TC 46** is ISO's Technical Committee (TC) for information and documentation standards.
- **SC 9** is the TC 46 Subcommittee (SC) that develops and maintains ISO standards on the identification and description of information resources.

Thesaurus and Indexing Standards – ISO

- **ISO 2788:1986**

  Documentation - Guidelines for the establishment and development of monolingual thesauri

- **ISO 5964:1985**

  Documentation - Guidelines for the establishment and development of multilingual thesauri

- **ISO 5963:1985**

  Documentation - Methods for examining documents, determining their subjects, and selecting indexing terms

- **ISO 999:1996**

  Information and documentation - Guidelines for the content, organization and presentation of indexes

Thesaurus and Indexing Standards – ANSI/NISO

- **ANSI/NISO Z39.19 - 2003** Guidelines for the Construction, Format, and Management of Monolingual Thesauri

- **NISO Z39.19-200x** Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies

- **NISO TR02-1997** Guidelines for Indexes and Related Information Retrieval Devices
  by James D. Anderson

ISO TC 37

Scope of ISO TC 37:
Standardization of principles, methods and applications relating to terminology and other language resources.

- TC 37/SC 1 - Principles and methods
- TC 37/SC 2 - Terminography and lexicography
- TC 37/SC 3 - Computer applications for terminology
- TC 37/SC 4 - Language resource management

Other ISO standards: Concept-oriented terminology

ISO 704:2000 Terminology work - Principles and methods

ISO 860:1996 Terminology work - Harmonization of concepts and terms

ISO 1087-1:2000 Terminology work - Vocabulary - Part 1: Theory and application

ISO 1087-2:2000 Terminology work - Vocabulary - Part 2: Computer applications

ISO 10241:1992 Preparation and layout of international terminology standards

ISO - Data Categories

- **ISO 12200:1999** Computer applications in terminology - Machine-readable terminology interchange format (MARTIF) - Negotiated interchange
- **ISO 12616:2002** Translation-oriented terminography
- **ISO/TR 12618:1994** Computer aids in terminology - Creation and use of terminological databases and text corpora
- **ISO 12620:1999** Computer applications in terminology - Data categories used to create glossaries
- Others in the work codes

W3C

- OWL – Web Ontology Language
- RDF – Resource Description Format
- Topic Maps
- SKOS - Simple Knowledge Organization Systems

- Which community to serve?
- Build on the current standard
- Might make this link next

Other Relevant ISO & W3C Standards

- Markup Languages
- Metadata Resources
- Character Coding
- Access Protocols and Interoperability
- Content Creation, Manipulation, and Maintenance
- Authoring Standards
- Text and Content Markup
- Translation Standards
- Terminology and Lexicography Standards
- ISO TC 37 Standards
- Terminology Interchange Standards
- Controlled Language Standards
- Taxonomy and Ontology Standards
- Corpus Management Standards
- Locale-Related Standards

For translation, terminology and applied linguists go to: http://appling.kent.edu/ResourcePages/LTStandards/Chart/standards.chart.htm#Ontology

Other links

- http://esw.w3.org/topic/SkosDev/ThesaurusLinks/XmlFormats
- MARC-21 XMLSchema.
- Zthes Z39.50 profile for thesaurus navigation (2001).
- TML thesaurus markup language (1999).
- ADL Thesaurus Protocol XML formats (2002).
- MeSH XML format (2001).
- GEMET XML format (2003).
- APAIS XML thesaurus format, an extension of Zthes (2000).
- Open University thesaurus schemas (2002).
- Soergel XML thesaurus specification (2001).

Other things to watch

- ☐ Other W3C and ISO areas
- ☐ Support groups
  - ■ Blogs
  - ■ Communities of Practice
- ☐ SIMILE
- ☐ Web 2.0 activities
- ☐ WSDL – Web Services Digital Library

DataHarmony

Access Innovations.

The Solutions

- ☐ Registry? NKOS KOS of KOS
- ☐ SKOS participants
- ☐ KOS typology - Tudhope
- ☐ Tesauro.com – Spanish - Salama
- ☐ Kent.edu site – Marcia Zeng
- ☐ Taxonomy Warehouse – Factiva - Clarke
- ☐ UMLS - Unified Medical Language System

More Solutions

Simile

- Semantic Interoperability of Metadata and Information in unLike Environments (Open Source

- UK HILT - Nicholson

DataHarmony

Access Innovations.

Good starts

- Link to each other
- Include
  - Thesauri
  - Taxonomies
  - Semantic webs
  - Classification systems
  - Subject headings
  - SKOS
  - OWL and Ontologies
  - Other KOS

What about?

- ☐ Authority Files
- ☐ Other pick lists
- ☐ Roget's and other synonym rings
- ☐ Dictionaries
- ☐ Gazetteers
- ☐ Glossaries
- ☐ Etc.

Thesauri need a place to call home

- Open contribution
- Thesaurus metadata contributions
- Comments on the contributions
- Examples of implementation
- A clearing house to keep track of
    - all the initiatives and
    - suggested standards,
    - a means to allow input from and to those initiatives, and
    - publishing of best practices or lessons learned from implementations
    - perhaps a WikiKOS

Discussion??

Thank you for your attention!

Marjorie M. K. Hlava

Access Innovations / Data Harmony

mhlava@accessinn.com

Z39.19 - What's new?

The old standard

- Coverage
  - documents
- Types of vocabularies
  - Thesauri
- Single BT
- Post-coordinated
- Printed formats
- Monolingual vocabularies

The revised standard

- Coverage
  - Content objects
- Types of vocabularies
  - lists, synonym rings, taxonomy
- Pre-coordinated
- Web format
- Multilingual vocabularies (general)
- Poly hierarchical
- Interoperability
- Facet analysis

DataHarmony

Access Innovations.