Taxonomy ShareSpace

Marjorie M.K. Hlava
President, Access Innovations, Inc.
Albuquerque, New Mexico
USA

NKOS/CENDI Joint Workshop
October 22, 2009

DataHarmony

Access Innovations.

Mission

- "Taxonomy ShareSpace -- access, deposit, share, save, and discuss taxonomy resources"
- There is no theoretical limit to the scope of subject matter coverage in ShareSpace.

DataHarmony

Access Innovations.

Additional purposes

- To provide visitors with informative examples of taxonomy structure
- To show the variety and scope of taxonomies (and other terminologies)
- To reveal the possibilities that can be achieved with a taxonomy
- To implement and take advantage of JISC recommendations

DataHarmony Access Innovations.

JISC recommendations

- Published in Terminology Registry Scoping Study (TRSS): Final Report
  - By Koraljka Golub (UKOLN, University of Bath) and Douglas Tudhope (University of Glamorgan)
- JISC = the Joint Information Systems Committee of the Higher and Further Education Funding Councils (UK)

DataHarmony Access Innovations.

Outline of the talk

- JISC terminology recommendations (as implemented in Taxonomy ShareSpace)
- Tour of Taxonomy ShareSpace website
- Search and discovery of the metadata
- NICEM thesaurus for navigation and search
- Contributions of taxonomy metadata

DataHarmony Access Innovations.

What a terminology registry is

- TRSS: “A terminology registry (TR) lists, describes, identifies and points to sets of vocabularies available for use in information systems and services.”
- “The registry allows discovery of suitable schemes for information or, potentially, use, by exposing rich metadata about them for navigation and retrieval.”

DataHarmony Access Innovations.

Implemented recommendations

- Which JISC/TRSS recommendations are followed in Taxonomy ShareSpace
- How the metadata types are implemented on the website
  - Viewer interface
  - Contributor template

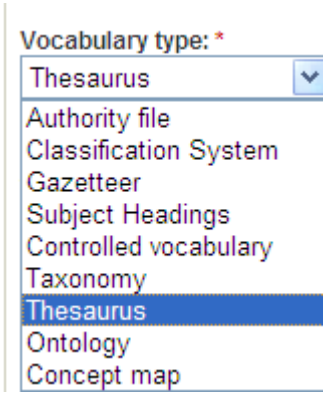
DataHarmony Access Innovations.

Group 1 – General information

- Vocabulary name
  - No-brainer – essential metadata
  - Also serves as a heading
- Vocabulary alternative name or acronym
  - Originally omitted; acronym included w name
  - Decided to include field to improve search
- Vocabulary type
  - Pick list in entry template for contributors

DataHarmony Access Innovations.

Group 1 (continued)



DataHarmony

Access Innovations.

Group 1 (continued)

- Author or editor
  - Credits individuals
  - Includes institutional authors
- Current version/edition
- Date of current version
- Update frequency

DataHarmony

Access Innovations.

Group 1 (concluded)

- Available format(s)
- Omitted – Available terminology services
- Vocabulary identifier (we used "Vocabulary URL"
- Vocabulary sample URL
- Vocabulary description (essential)

DataHarmony Access Innovations.

Group 2 – Scope and Usage

- Language(s)
  - We sought out multiple language vocabularies to enrich the variety of the registry.
  - Pick list or free text? Problems w. short list
  - ISO language codes list - ISO 639-2 or -3
  - Numerous languages, some quite obscure (example: OmegaWiki)

DataHarmony Access Innovations.

Group 2 (continued)

- Major subjects covered
  - In cases for which major subjects were not included in the descriptive material on the terminology website, we used the top terms.
- Minor subjects covered
  - Included. However, the major subjects are often fully indicative of the subject matter scope.

DataHarmony Access Innovations.

Group 2 (continued)

- Purpose as given by author/publisher
  - We used the label “Purpose” for succinctness.
  - The purpose is sometimes discernible, but no explicitly stated by the author/publisher.
- Used by
  - Sometimes clearly stated in a separate paragraph on the terminology website
  - Often covered in the description or purpose

DataHarmony Access Innovations.

Group 2 (continued)

- Related vocabularies
- Overlap with related vocabularies
- Mappings to other vocabularies: which vocabularies, intellectual or automated

DataHarmony Access Innovations.

Group 2 (concluded)

- Optional Group 2 suggestions not used:
  - Description of collections where used
  - Usage case study
  - Use in application profiles
  - Rating
  - URL to vocabulary users' discussion board
  - Change notification details
  - URL to tutorial for applying vocabulary

DataHarmony Access Innovations.

3 – Vocabulary characteristics

- Type of display
  - Check list
  - Hierarchical; Alphabetical; etc.
- Description of overall structure
- Type of terms
  - TRSS: “e.g. concept terms, geographic names

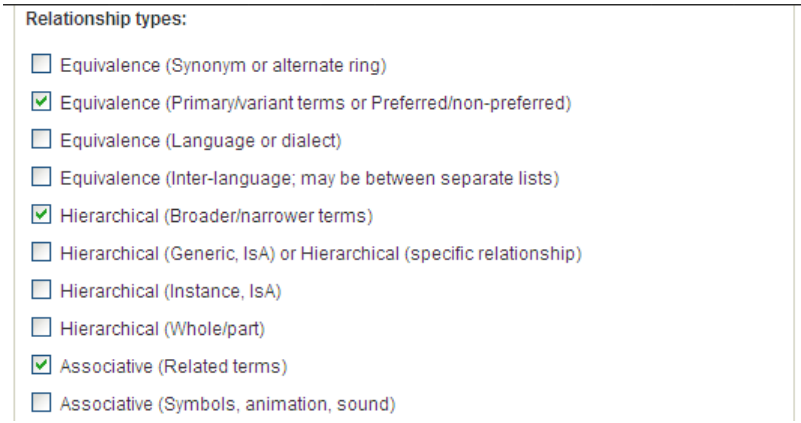
DataHarmony Access Innovations.

Group 3 (continued)

- Types of relationships
  - Check list (see next slide)
  - Hierarchical; associative; equivalence (etc.?)
  - Two major types of equivalence found
    - Preferred/Non-preferred
    - Equal synonyms (in synonym rings, termbanks)
- Total number of terms
- Total number of classes

DataHarmony Access Innovations.

Group 3 (continued)



DataHarmony

Access Innovations.

Group 3 (concluded)

- Number of preferred terms
- Number of non-preferred terms
- Depth of hierarchy
  - On rare occasions, is stated on website
  - We sample branches (“at least n levels”)
- Not used – Notes field (Optional)
- Not used Information given (Optional)

DataHarmony

Access Innovations.

Group 4 – Term and conditions

- Availability: free for all, free for regsitered users, costs
  - Initial focus was on terminology websites freely available for viewing
- Import/download instructions
- Purchase/subscription price
- Licensing options

DataHarmony

Access Innovations.

Group 5 - Provider

- Vocabulary provider name
  - We generally listed the organization hosting the vocabulary website
- Vocabulary provider URL
- Vocabulary provider contact details

DataHarmony

Access Innovations.

Current architecture

- Mainly an enriched directory
- Names and links
- Metadata
- Seed/starter entries to attract users and encourage outside contributions
- Is set up (in Drupal) for entries by contributors

DataHarmony

Access Innovations.

Collaboration opportunities

- Users can create the following content:
  - Blog entries
  - Event announcements
  - Stories/articles
  - Terminology metadata
  - Eventually, downloaded terminologies
    - Preserve for endangered thesauri

DataHarmony

Access Innovations.

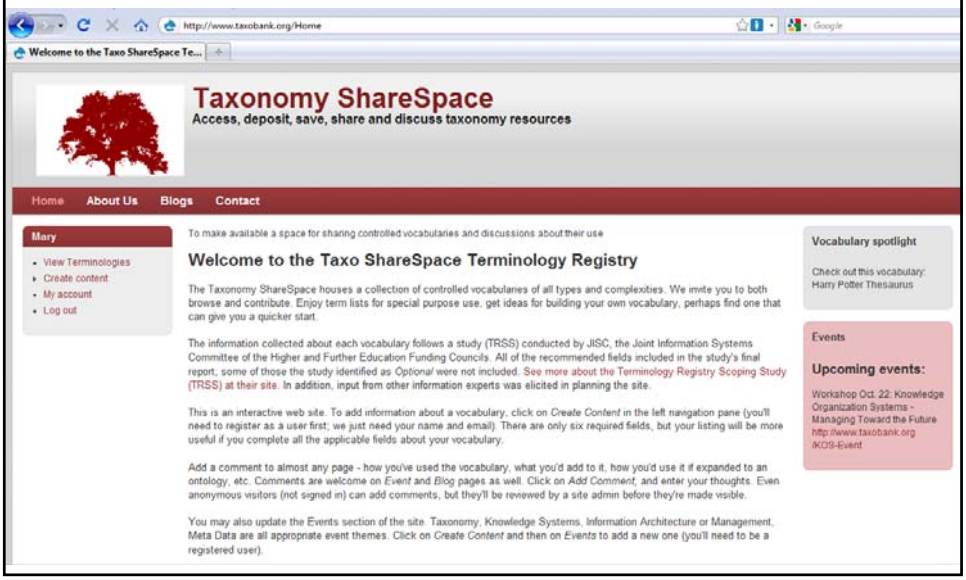
Special case – student thesauri

- Thesauri usually on university sites
- Therefore in danger of extinction
- We offer to preserve thesauri/web pages
- Interesting subjects
  - Harry Potter thesaurus (by Dr. Stasa Milojevic)
  - Sailspotting (Sails of tall ships)
  - Thesaurus of Bellydancing (ThoBe) by the League of Bellydancing Librarians, Indexers, and Archivists (LoBeLIA)

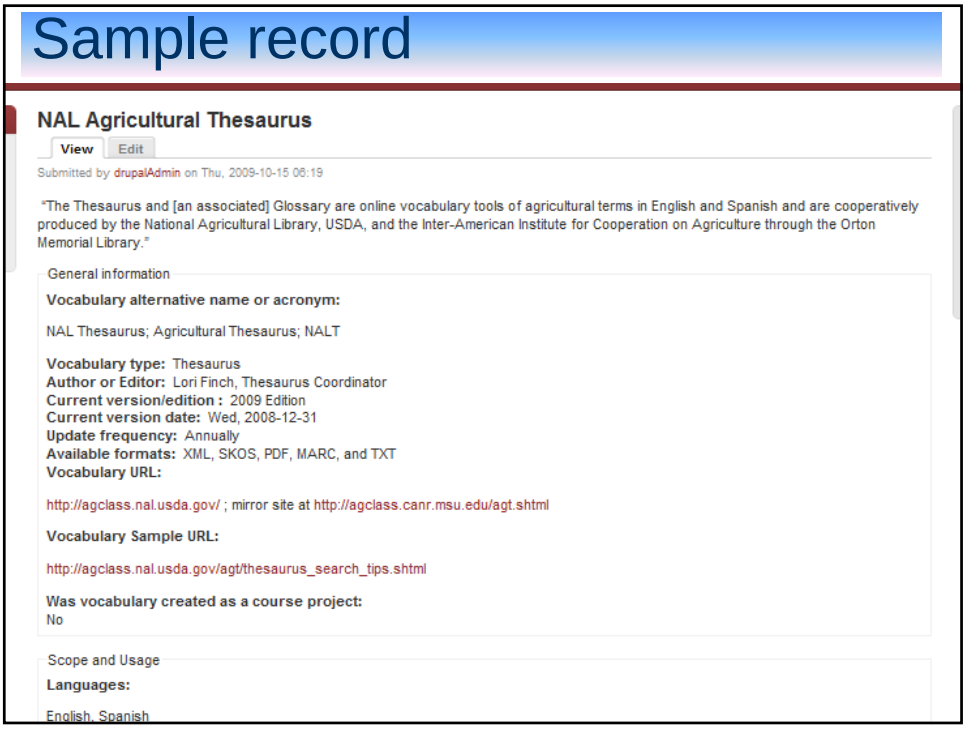
Possible eventual architecture

- Storage of full contributed terminologies
- Use of SKOS wherever practicable
- Search capability across terminologies
- Automated indexing of descriptions
- Navigation tree based on NICEM thesaurus as “spine” for collection

Home page



Sample record



Sample record (continued)

Major subjects covered:

Agricultural subjects, biological sciences

Minor subjects covered:

Rural and agricultural sociology; physical sciences

Purpose:

"The NAL Agricultural Thesaurus (NALT) was originally prepared by staff of the National Agricultural Library to meet the needs of the United States Department of Agriculture (USDA), Agricultural Research Service (ARS). ... The thesaurus is primarily used for indexing and for improving retrieval of agricultural information."

Used By:

Currently, the thesaurus is the indexing vocabulary for the National Agricultural Library's bibliographic database of citations to agricultural resources, AGRICOLA. The Food Safety Research Information Office (FSRIO) and Agricultural Network Information Center (AgNIC) also use the NALT as the indexing vocabulary for their information systems. In addition, the NALT is used as an aid for locating information at the Agricultural Research Service, Economic Research Service and AgNIC web sites and databases.

Related vocabularies:

The NAL Glossary provides definitions for 2,574 of the thesaurus terms in English and 2,574 terms in Spanish.

Overlap with related vocabularies:

It can be assumed that NALT has considerable overlap with AGROVOC, another agriculture related thesaurus.

Mappinigs to other vocabularies:

NALT was aligned with AGROVOC by participants in the OAEI 2007 food task. See http://www.few.vu.nl/~wrvhage/oaei2007/food.html

Vocabulary characteristics

Type of display :
Hierarchy
Description of overall structure:

See http://agclass.nal.usda.gov/structure.shtml#equiv

Type of terms:

Agricultural and scientific concepts

Sample record (continued)

Relationship types:
Equivalence (Primary/variant terms or Preferred/non-preferred)
Hierarchical (Broader/narrower terms)
Associative (Related terms)
Number of classes:

17 subject hierarchies

Number of terms:
English: 73,194. Spanish: 69,118
Number of preferred terms:
English: 44,857; Spanish: 44,857
Number of Non-Preferred terms:
English: 28,337; Spanish: 24,261
Characteristics Comments:

Terms and Conditions

Availability:
Free to use
Comments, Availability:

Downloadable from the website in XML, SKOS, PDF, MARC and
TXT formats. Other formats may be obtained by contacting the NAL Thesaurus
Staff by email, mail, phone, or fax. Browseable online.

Sample record (concluded)

Licensing Options:

No license is required to obtain the data contained in the NAL Agricultural Thesaurus and Glossary. If the use is not personal, (1) the U.S. National Agricultural Library must be identified as the creator, maintainer, and provider of the data; (2) the version of the data must be clearly stated by the year, e.g. 2004 NAL Thesaurus; and (3) if any modification is made in the content of the file, this must be stated, along with a description of the modifications.

Purchase/Subscription price :
$0.00
Import/download instructions :

http://agclass.nal.usda.gov/download.shtml Downloads involve the acceptance of the NAL's Usage Agreement

Provider

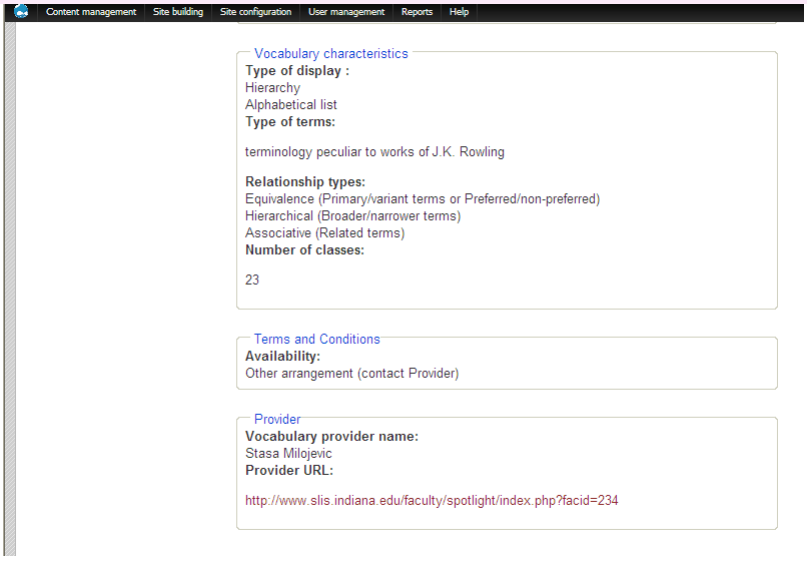
Vocabulary provider name:
National Agricultural Library, U.S. Department of Agriculture
Provider URL:

http://www.nal.usda.gov/

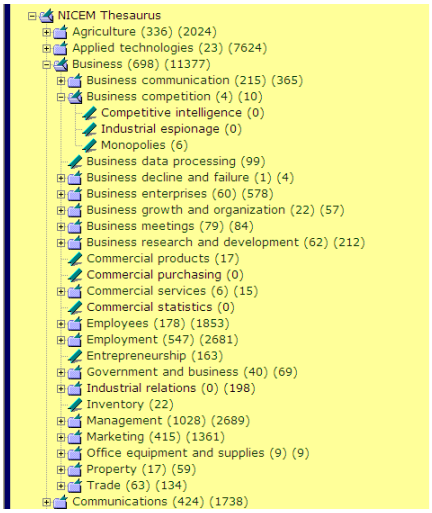
Provider contact details:

Vocabulary provider contact details: E-mail: thes@nal.usda.gov. Mailing address:
National Agricultural Library, c/o Lori Finch, Thesaurus Coordinator, 10301 Baltimore Ave., Beltsville, MD 20705. Phone number: (301) 504-6853. Fax number: (301) 504-5213

Results from data base show all fields



Navigation by Taxonomy



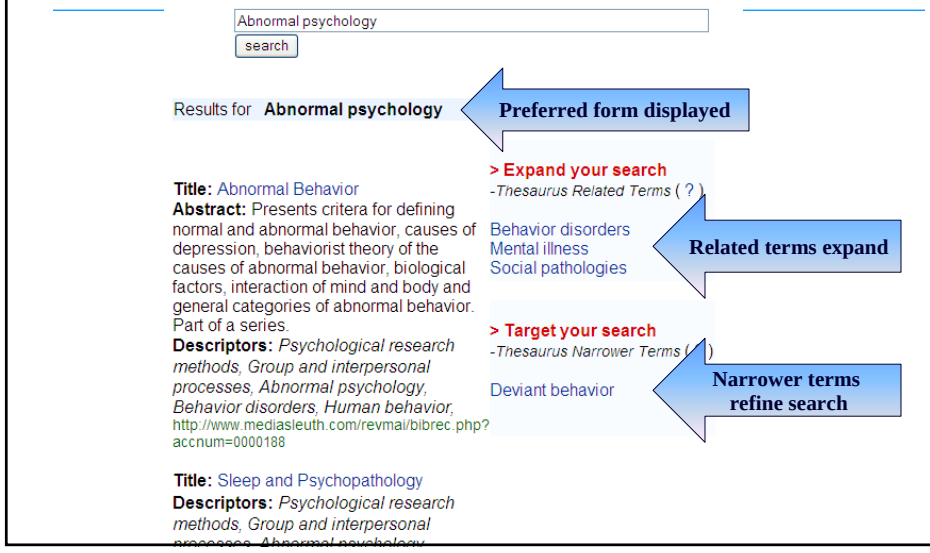
Automatic Completion of Synonyms and Preferred Terms

Search by Taxonomy

psy

Psychiatric hospitals *(nonpreferred)*
Psychic phenomena *(nonpreferred)*
Psychological dependence
Psychological disorders
Psychopathology *(nonpreferred)*
Psychotropic drugs *(nonpreferred)*
Abnormal psychology
Adolescent developmental psychology
Adult developmental psychology
Animal behavioral psychology
Animal psychology
Child developmental psychology
Clinical psychology
Cognitive psychology
Community psychology
Clinical psychopharmacology

Results by Taxonomy



Search By Facet

Next steps

- Add navigation taxonomy and search
- Decide how to accommodate uploads
- Populate as KOS become known
- Allow others to add their KOS
- Spread the word on the registry
- Encourage others to join and participate
- Build the community

DataHarmony

Access Innovations.

Thank you!

Marjorie M.K. Hlava
President
Access Innovations, Inc.
mhlava@accessinn.com
505-998-0800

DataHarmony

Access Innovations.