Open Government Vocabularies and Metadata

Dan Gillman

US Bureau of Labor Statistics

Wednesday, March 28, 2012

BLS

BUREAU OF LABOR STATISTICS
U.S. DEPARTMENT OF LABOR

www.bls.gov

Outline

- Vocabularies
- Open Government Vocabularies WG
- Basics
- Connections to metadata
- Benefits

BLS

Vocabularies

- Linked Open Data term
- Used by Data.Gov
- Means
  - Lexicons / Glossaries
  - Taxonomies / Hierarchies
  - Thesauri
  - Models / Schemas
  - Ontologies

BLS

Vocabularies

- In Statistics, vocabularies are
  - Code sets
  - Classifications and Taxonomies
  - Database models
  - XML schemas
  - Questionnaires (!)
  - Datatypes (!)
    - Base types
    - Arrays, structures, and classes

BLS

OGV WG

- CIO Council
  - Architecture and Infrastructure Committee
    - Data Architecture Sub-committee
      - Open Government Vocabularies Working Group
- OGV
  - Government + Contractors
  - 30 members
  - 6 regular, 4 sometimes, rest lurk

BLS

OGV WG

- Work ended Feb 2012
- Deliverables (to DAS)
  - Overview
  - Registry/Catalog metadata
    - OGV – Registration Model
  - Registry/Catalog operations
    - OGV – Registration Procedure
  - Vocabulary model
    - OGV – Content Model

BLS

Basic Assumptions

- Underlying theory
  - ISO 704 – *Principles of terminology*
  - ISO 1087-1 – *Terminology* – Part 1: *General vocabulary*
  - ISO/IEC 11179-6 – *Metadata registries* – Part 6: *Registration*
- Build abstract model
- Map to specific models
  - Many possibilities

BLS

Basic Assumptions

- Simple Knowledge Organization System
  - SKOS
  - W3C recommendation
  - Actually, SKOS + Extensions
    - DDI / Semantics Workshop
    - Schloß Dagstuhl, Germany, Sept 2011
- Vocabulary = Concept System
    - As defined in SKOS and ISOs

BLS

Connections to Metadata

- Uses of Codes sets and Classifications
  - Questions in questionnaires
    - Response choices
  - Variables
    - Allowed values
  - Tables
    - Row headings
  - Time series
    - Dimensions

BLS

Connections to Metadata

- Uses of
  - Lexicons
  - Glossaries
  - Thesauri
  - Taxonomies
- Variables
  - Characteristic
  - Population / Universe

BLS

Connections to Metadata

- Questions
  - Definitions of special terms
- Tables
  - Units of analysis
- Time series
  - Measure
- Support for searching and discovery
  - Including search engines

BLS

Connections to Metadata

- Uses of
  - Models
  - Schemas
- Questionnaires
  - Skip pattern -> relationships
- Databases
- XML based applications
  - XML-Schema

BLS

Connections to Metadata

- Uses of ontologies
  - Ontology defined as
    - Concept system + computational model
- Datatypes
  - Combine code sets and glossaries
    - Value space -> Code set
    - Computational model
      - Axioms -> Glossary
      - Characterizing operations -> Glossary

BLS

Connections to Metadata

- Other uses of ontologies
  - Formalized statistical systems
    - Fully automated
    - Metadata driven processing
    - Inference capabilities
  - Harmonization systems
    - Complete metadata repository
    - Including definitions of all terms

BLS

Benefits

- Metadata collection
  - Reference
  - Not value
- Semantic interoperability
  - Automatically know meaning is the same
  - Data sets, automatically
    - Compare
    - Combine
    - Harmonize

BLS

Benefits

- Variables
  - Universe
  - Characteristic
  - Allowed values
- Use vocabularies
  - Just relationships
  - LOD ready

BLS

Benefits

- Questionnaires
  - Terms in questions defined
  - Response choices managed
- Database documentation
- Table generation
- Time series indexing and search
- Others

BLS

Conclusion

- Vocabularies
  - Fundamental to metadata management
  - First step
  - Subsequent steps use vocabularies
- Vocabulary management
  - Registry / Catalog for Vocabularies
  - See talk on Registries this afternoon

BLS

Contact Information

Daniel Gillman
Information Scientist
Office of Survey Methods Research
www.bls.gov/ore
+1-202-691-7523
Gillman.Daniel@bls.gov

BLS
BUREAU OF LABOR STATISTICS
U.S. DEPARTMENT OF LABOR