

# Interoperability Through Vocabulary Registries

---

**Dan Gillman**

*US Bureau of Labor Statistics*

*Wednesday, March 28, 2012*



# Outline

---

- Open Government Vocabularies WG
- Vocabularies
- Catalogs and Registries
- Federated Registries
- OGV Registries
- Interoperability

# 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
- Work ended Feb 2012

# OGV WG

- Deliverables (to DAS)
  - ▶ Overview
  - ▶ Registry/Catalog metadata
    - OGV – Registration Model
  - ▶ Registry/Catalog operations
    - OGV – Registration Procedure
  - ▶ Vocabulary model
    - OGV – Content Model
- To be published on [vocab.data.gov](http://vocab.data.gov)

# Vocabularies

---

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

# Vocabularies

---

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

# Catalogs and Registries

---

## ■ Catalog

### ▶ Library Card Catalog

- Basic descriptions
- Promotes discovery
- Allows some comparison

### ▶ Electronic Catalog

- Same idea
- MARC records
- Dublin Core

# Catalogs and Registries

---

- Registry
  - ▶ Includes
    - Metadata
    - Services
    - Procedure
- Metadata
  - ▶ Supports discovery
  - ▶ Some comparability



# Catalogs and Registries

---

- Services (computer oriented)
  - ▶ CRUD operations
  - ▶ Other functions
- Procedures (people oriented)
  - ▶ Management
  - ▶ Provenance
  - ▶ Quality
  - ▶ Harmonization
  - ▶ Standardization

# Federated Registries

---

- What is it?
  - ▶ Multiple instances needed
  - ▶ Single registry not possible
  - ▶ Often each instance has same design
  - ▶ Instances have same purpose
  - ▶ Global search
  - ▶ Local control
  - ▶ Common interface

# Federated Registries

---

## ■ Example

- ▶ US federal statistical agencies
  - Lack of central control
  - Need local instances; local control
  - Global search to support finding data
  - Purpose – same across agencies
    - Manage code sets, classifications

# OGV Registry

- OGV – Registration Model
  - ▶ Vocabulary metadata
    - General cataloging description
    - Promotes discovery
    - Loosely follows Dublin Core
  - ▶ Management metadata
    - Roles
      - Registrar
      - Submitter
      - Steward

# OGV Registry

- ▶ Management metadata – Statuses (6)
  - Administration
    - ❖ Publicly available?
  - Identification
    - ❖ URI's assigned?
  - Metadata
    - ❖ All metadata provided?
  - Quality – content
    - ❖ Content metadata meet quality guidelines?
  - Quality – vocabulary
    - ❖ Vocabulary metadata meet quality guidelines?

# OGV Registry

- ▶ Management metadata– Statuses (cont'd)
  - Registration
    - ❖ Semantic comparisons made?
- OGV – Registration Procedure
  - ▶ Roles defined
  - ▶ Statuses defined
  - ▶ Steps for
    - Each role
    - Each status level

# OGV Registry

- Result - Vocabularies
  - ▶ Described
    - Container
    - Contents
  - ▶ Managed
    - Existence
    - Quality
    - Harmonization
- Quick step to standardization

# Interoperability

---

- Interoperability described
  - ▶ Ability to use a resource without access to the source
- Identification status =>
  - ▶ URI's exist for
    - Vocabulary as a whole
    - Each concept and term / code within



# Interoperability

- Example
  - ▶ Metadata for variables, captured by
    - Value
    - Reference (use URI's)
- URI's guarantee same meaning
- Shared meaning => Semantic interoperability
- Shared codes => Data interoperability

# Conclusion

---

- Registries
  - ▶ Catalogs with services and procedures defined
- Federated registries
  - ▶ Support groups of agencies
  - ▶ No central control
  - ▶ Global access
- Support interoperability

# Contact Information

---

**Daniel Gillman**

Information Scientist

Office of Survey Methods Research

[www.bls.gov/ore](http://www.bls.gov/ore)

+1-202-691-7523

Gillman.Daniel@bls.gov