

Taxonomy / Lexicon Project at BLS

Dan Gillman

Leader, Taxonomy/Lexicon Team
US Bureau of Labor Statistics
FedCASIC 2014



Outline

- Goals
- Background
- Work plan
- Progress
- Issues
- Applications

Goals

- Design System for BLS Terminology
 - ▶ Technical BLS terms
 - ▶ Plain English words
 - ▶ Supports series dissemination
 - taxonomy
 - ▶ Supports document tagging
 - lexicon
 - ▶ Other possible applications

Goals

- Taxonomy
 - ▶ Terms
 - ▶ Concepts
 - Manage synonyms, homonyms
 - ▶ Relationships
 - Broader/Narrower
 - Related to
- Lexicon
 - ▶ Alphabetical listing of terms

Background

- Many dissemination tools
 - ▶ Specific to
 - Particular series
 - Office or survey
- Document tagging
 - ▶ Improve searches
 - ▶ Document / data consistency

Background

- Previous work
 - ▶ Pilot thesaurus (1990s)
 - ▶ Data Query Team
 - Series descriptions
- BLS team
 - ▶ Formed summer 2013
 - ▶ Representation from all offices

Taxonomy / Lexicon Project at BLS

Work Plan

- Produce pilot Taxonomy and Lexicon
 - ▶ Gather terms
 - Time series measures
 - Time series dimensions (classifications)
 - Highest level concept only
 - Grouping concepts (e.g., productivity)
 - ▶ Place terms and definitions in database
 - ▶ Find plain English equivalents
 - ▶ Relate English to BLS terms

Work Plan

- Produce pilot Taxonomy and Lexicon
 - ▶ Add some dimensional details
 - ▶ Establish relationships
 - ▶ Gather and add additional terms
 - E.g., establishment, household
 - Missing measures
 - ▶ Validate plain English
 - ▶ Repeat as needed

Progress

- Measures
- Classifications – highest (root) level
 - ▶ Some detail obtained
- Grouping terms for measures
- Two main facets
 - ▶ Measure
 - ▶ Characteristic
 - Classifications
 - Statistics

Progress

- High level groupings per facet
- Plain English

Issues

- Classification/Categorization variants
 - ▶ E.g., geography
- Definitional variants
 - ▶ E.g., establishment
- Harmonization? Standardization?

Applications

- Original
 - ▶ Dissemination
 - ▶ Document tagging
 - ▶ Coherent search
- Question development
 - ▶ Wording depends on concept
 - ▶ Answer choices
- Interviewer instructions

Applications

- Sampling
 - ▶ Strata selection
- Data dictionary construction
- Database design
 - ▶ Entities / attributes
- Table construction
 - ▶ Tabs – labels

Applications

- Web site design
 - ▶ Organization
 - Across pages
 - Within each page
 - ▶ Labels
- Help desk
 - ▶ Find information faster

?

Contact Information

Dan Gillman

information scientist

Office of Survey Methods Research

www.bls.gov/osmr

202-691-7523

Gillman.Daniel@BLS.Gov

