Taxonomy / Lexicon Project at the US Bureau of Labor Statistics

Dan Gillman
Leader, Taxonomy/Lexicon Team
US Bureau of Labor Statistics
IASSIST 2014
Toronto, CA
Outline

- Goals
- Background
- Work plan
- Future work
- Related Work
Goals

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

- Design System for BLS Terminology
  - Other applications
    - Improve web site design
  - Other possible applications
    - Classification management
    - Harmonize / Standardize terms
Background

- Many dissemination tools
  - Specific to
    - Particular series
    - Office or survey
  - Dependent on
    - Web site design
    - User computer configuration
- Limitations
  - No combining / harmonizing data
Background

- Document tagging
  - Improve searches
  - Document / data consistency

- Current search engine
  - Results inconsistent
  - Results incomplete
Background

- Previous work
  - Pilot thesaurus (1990s)
  - Data Query Team
    - Series / Measures descriptions

- BLS team
  - Formed summer 2013
  - Representation from all offices
Work Plan

Phase 1

- Completed January 2014
- Identify, from LabStat
  - Measures
  - Characteristics
    - Industry, Occupation, Geography, ...
  - Statistics
- Organize
  - Spreadsheets
  - Access DB – pilot thesaurus
Work Plan

- Phase 1
  - Identify plain English
    - Used by public
    - Represents their understanding
      - Our data
      - Their question
  - Interview
    - All regional offices
    - All program offices / surveys
Work Plan

- Phase 1
  - Produce / deliver report
Work Plan

Phase 2

- Begun April 2014
- Build 2 or 3 level hierarchy
  - Organize measures & characteristics
  - Identify commonalities
  - Assign terms to bins
  - Link to individual
    - Measures
    - Characteristics
    - Statistics
Work Plan

Phase 2

- Build plain English mapping
  - Identify commonly used words
  - Determine BLS meaning
    - Measures
    - Characteristics
    - Statistics
  - Assign tendency score
    - Strong
    - Weak
Work Plan

Phase 2

- Integrate plain English
  - Link to hierarchy terms
- Produce spreadsheets
  - Terms / Words per level
  - Mappings between levels
Future Work

- Classifications
  - Detailed links
    - Census Industry & NAICS
    - Census Occupation & SOC
    - MSA / CSA / other Geography
    - Other classifications (men versus male)

- Universes and Variables
  - Collect terms
    - Establishment – the same?
Future Work

- Cognitive tests
  - Hierarchies
    - Measure versus characteristics
  - Effective?
    - Card sorting
    - Focus groups
    - Other techniques

- Improve relationships

- Build database
Related Work

- Harmonize / Standardize terminology?

- Benefits
  - Common understanding
  - Data harmonization
  - Data consistency

- Possible? – New team
  - Determine feasibility, costs, time
  - Possible design problems
  - Series breaks
Questions
Contact Information

Dan Gillman
information scientist
US Bureau of Labor Statistics
Office of Survey Methods Research
www.bls.gov/osmr
202-691-7523
TaxonomyLexiconTeam@bls.gov