Terminology and Bottom-Up Standardization at the US Bureau of Labor Statistics

Daniel Gillman

Information Scientist

US Bureau of labor Statistics

Modern Stats World Conference

Friday 13 April
Overview

- Terminology projects at BLS
- Standardization
  - Top-Down and Middle-Out
- Terminology standards
  - Bottom-Up
Terminology Projects at BLS

- Glossary
  - Complete set of technical Definitions
  - Synonyms identified
    - through links
  - Homographs identified
    - through repeated term entries
    - With different definitions
Terminology Projects at BLS

- A-Z List
  - List of topics
  - Current web site guide is too program specific
  - Contains some terms in glossary and taxonomy
  - Higher level than taxonomy
  - Organized sometimes similar to taxonomy
Terminology Projects at BLS

Taxonomy

- Multi-hierarchy
  - High level
    - Measures – quantitative variables/estimates
    - Characteristics – classifications used as dimensions
  - Each characteristic is its own hierarchy

- Every series or time-oriented data set
  - Has at least 2 paths to it
    - One from Measures
    - At least one from characteristics
Terminology Projects at BLS

- **Taxonomy**
  - **Planned uses**
    - Interface for new DataFinder dissemination tool
      - Select as many series as needed
      - Independent of subject matter
    - Guide to reorganize BLS web site
      - Currently, can’t answer questions similar to
        - “Give me all the data you have about Boston”
        - No way to know if the user has ALL such data
Terminology Projects at BLS

- **Taxonomy**
  - **Planned uses**
    - Classification for documents, reports, and articles
      - Ensure classification is consistent with data
    - Classification of records for official archiving
      - Same problem
    - Resource for disambiguating similar terms
      - Homographs
      - Near synonyms
        - Wages, earnings, pay, compensation, income
Standardization

- Standards produced by process
  - Open
    - Any stakeholder may join the effort
  - Balanced
    - Stakeholders fairly represent all constituencies
  - Fair
    - No stakeholder has an advantage over others
  - Transparent
    - Development process is open to any observer
Standardization

- Typical scenario -
  - Some business practice needs standardizing
    - Choice and meaning of colors in traffic lights
    - Turning direction of threads in a light bulb socket
    - Statistical metadata information models
  - Problem identified
  - Relevant SDO is petitioned to begin process
  - Enough stakeholders identified to support work
Standardization

- Technical Work
  - General problem presented
  - If solutions are in existence
    - Ideal solution determined
    - Best practices identified
    - Common elements fit to ideal and use cases
    - Example – GSBPM
      - Every NSO and survey has its own practice
      - But, each step maps to the theory of statistical surveys
    - Middle-Out Approach
Standardization

- Technical Work
  - General problem presented
  - If solutions are not already in use
    - Problem solved from first principles
    - Development occurs from general to specific
    - Example – GSIM
      - Again, theory of statistical surveys known
      - Good, complete information models don’t exist
    - Top-Down approach
Terminology Standards

- For any statistical survey
  - Many terms used
  - Some come from other standards
  - Others invented or borrowed for specialized use
    - i.e., within domain of each survey
  - For example, in US Current Population Survey
    - Labor Force
    - Unemployed
Terminology Standards

What happens when the NSO

- Needs a comprehensive glossary
- Wants to promote data
  - Interoperability –
    * shared understanding
  - Harmonization –
    * ability to combine data from multiple sources
- across the agency?
Terminology Standards

- Building terminology systems
  - Glossary, Taxonomy

- Collect existing terms
  - No notion of ‘best practices’

- Definitions
  - Create, if don’t have
  - Make sense, if confusing
  - Follow rules
Terminology Standards

Order terms

- Alphabitically
- Also, link to appropriate surveys
- Link similar terms
  - Wages
  - Earnings
- Remove unnecessary terms
  - Some terms mean the same as word in dictionary
  - Example: illness, injury
Terminology Standards

- Disambiguate homographs
  - Homograph – words/terms that are
    - spelled the same
    - different in meaning
    - Example: employment
    - Current Population Survey
      • People holding jobs
    - Current Employment Statistics survey
      • Employers with filled positions
Terminology Standards

That is Bottom-Up

Questions
Contact Information

Daniel Gillman
Office of Survey Methods Research
US Bureau of Labor Statistics
www.bls.gov/OSMR
+1.202.691.7523
Gillman.Daniel@BLS.Gov