Perspective from the US Bureau of Labor Statistics

Dan Gillman

Data Management & Curation: Lessons from Government, Academia, and Research

3 June 2014
US Federal Statistical System

- Not like any other country
- 14 major US federal statistical agencies
  - BLS is one
- One subject matter
  - BLS – Labor economics
  - Exception – Census Bureau
    - Censuses
    - Federal household data collection
BLS Mission

- The Bureau of Labor Statistics of the U.S. Department of Labor is the principal Federal agency responsible for measuring labor market activity, working conditions, and price changes in the economy. Its mission is to collect, analyze, and disseminate essential economic information to support public and private decision-making. As an independent statistical agency, BLS serves its diverse user communities by providing products and services that are objective, timely, accurate, and relevant.
BLS Subject Areas

- Employment / Unemployment
- Prices
  - Consumer, Producer, Import / Export
- Wages / Benefits
- Workplace health
- Productivity
- Time use
- International comparisons
BLS Data

- Most data –
  - Establishment surveys + Indexes
  - 2 Censuses: QCEW, CFOI
  - Other state programs: LAUS, MLS, SOII
  - Demographic surveys: CE, CPS, ATUS, NLS

- Publishing microdata
  - CE, NLS only
  - Census Bureau publishes CPS
BLS Data

- Most data
  - Highly aggregated measures
  - Over 50 measure families
  - Thousands of series
  - Managed in LabStat database

- Other data
  - Published separately
  - Migrating to LabStat
Data Availability

Data

- Tools on-line
  - Current – tools per subject area
    - One-screen (needs Java)
    - Multi-screen
  - Data Finder – one-stop shop - prototype

- Text files
  - Formerly thru FTP, now HTTP
    - http://download.bls.gov/pub/time.series/
    - FTP not supported by all Federal agencies
    - Session limits
Data Availability

- **Common API**
  - Access to LabStat database
  - JSON element set

- **Data tables**
  - HTML, PDF
    - All tables
  - Text, Excel, SAS, Stata
    - In addition
Data Availability

- Metadata
  - Mapping files ~ Classifications
  - Definitions in BLS Taxonomy
  - SDMX
    - Not in wide use
    - BLS uses own data transfer to LabStat
    - IMF
      - National indicators
      - Metadata – SDDS
      - US project
Data Availability

- Metadata
  - DDI
    - CE PUMS metadata
    - Plans / Needs Assessment
Management

- Memory – cheaper
- Storage needs – going up
  - Data steadily increasing
  - Need for copies / Older snapshots

- Backups
  - Regular database backups
  - New pay structure
    - By the byte!
Management

- Data files
  - Constantly growing
  - About 40M rows per year
    - Duplicate copies
      • Inside
      • Outside
    - 6GB new data each year
    - Does not include
      • Reports
      • Articles
      • Documentation
Archiving

- By program office
  - Data to NARA (national archive)
  - General – records management

- Data libraries
  - No general policy
  - Public data
    - On web
Contact Information

Dan Gillman
Information Scientist
Office of Survey Methods Research
www.bls.gov/osmr
202-691-7523
Gillman.Daniel@BLS.Gov