Why Reorganize US Official Statistics?

Erica L. Groshen
Visiting Senior Scholar, Cornell—ILR
Research Fellow, Upjohn Institute for Employment Research
Former Commissioner of Labor Statistics

July 10, 2019
Statistical agency priorities

• Provide data as a public good
  – **Accurate, Objective, Relevant, Timely & Accessible**
  – Help nation’s policymakers, businesses, and families make good decisions

• Maintain trust of respondents and users
  – Transparency and confidentiality

• Get the most from each US data dollar
Agenda

• Context
• Rationale
• Caveats and support
Current context

• Advancing data analytics
  – Burgeoning big data (transactional, administrative)
  – Cheaper computational power
  – New software (AI, linking, modelling...)

• Within federal stats system
  – Falling household response rates
  – Declining funding
  – Barriers to data-sharing
  – Great opportunities to improve data
    • Production use of non-survey data
    • Validation
    • Modeling and granularity
Reorganization proposal

• Move BLS to Commerce
  – Sit alongside Census and BEA
  – Step toward separate agency (ala UK?)

• Advanced in President’s Management Agenda
  – Proposed often: since 1970s or before, including previous administration

• Supported (conditionally) by
  – NABE
  – AEA
  – Friends of BLS
Reason to reorganize: modernization

- Share data across agencies
  - New data products
  - New efficiencies
- Coordinate federal data assets
  - Reduce public burden
  - Easier access to data
- Leverage buying power and capacity
- Increase staff capacity for data analysis
Barriers to government data-sharing

• Legal restrictions
  – Congressional restrictions on access to IRS data
  – State ownership and laws (UI wage records, NDNH)
• MOU process between departments
• Risk aversion + lack of clarity (re-interpretation needed)
  – Informed consent (e.g., web-scraping)
  – Responsibility for protecting confidentiality
Key data-sharing opportunity

• Synchronize BLS and Census business registers
  – Maintain separate registers now
  – Requires BLS access to IRS info in Census business register

• After CIPSEA (2002), 17 years without passing narrow synchronization permission law
  – Congress resists broader (especially Dept. of Labor) access to IRS data
Benefits to business register synchronization

• Reduce industry discrepancies between BLS and Census data
  – Improve BEA and private sector products that use both
• Improve quality of registers and dependent programs
  – Dynamics (BED, LEHD)
  – Establishment surveys from BLS and Census
• New products through linking
  – Entrepreneurship
  – Future of work
  – Job impacts of trade
  – Regional detail
• Enhance efficiency
Operational opportunities for sharing

• Computer storage and computational capacity
• Privacy protection
• Computation and analysis
  – Human capital
  – Software
• Cross-validation
• Linking files
  – Especially for modeling
Conditions/caveats/risks

Legislation carries risks along with opportunities...so, must ensure:

- **Independence** enhanced
- **Missions** of BLS, Census, and BEA preserved
- **Funding** adequate to improve data
- **Service** to data users maintained or improved
Bottom line: Reorg offers major opportunity

• Barriers to stats agency modernization can be reduced by reorganization
  – Data-sharing
  – Operational efficiencies
• Caveats important
How public data users can support statistical modernization and good reorganization

• Speak up for official statistics
  – Informally and formally
  – Encourage participation in surveys
  – Join Friends of BLS [www.friendsofbls.org](http://www.friendsofbls.org)
    • Linkedin: Friends of the Bureau of Labor Statistics
    • Twitter: [@Friends_of_BLS](http://twitter.com/Friends_of_BLS)

• Monitor developments

• Communicate
  – Needs, suggestions, observations
  – Compare/validate findings
Erica L. Groshen

erica.groshen@gmail.edu
Big Data challenges

• Quality
  – Coverage
  – Selection bias
  – Curation
  – History

• Access
  – Costs—in advance and ongoing (purchase and oversight)
  – Legal or administrative restrictions
  – Continuity risk

• Respondent protection
  – Informed consent
  – Privacy
  – Disclosure control

• Operations
  – Computer storage and computational capacity
  – Computation and analysis (human capital and software)
  – Linking files across sources (often critical for modeling)

• Comparability as you merge different strategies and sources
  – Time
  – Sectors

• Mission compatibility
  – Data manipulation
  – Front running
  – Curation quality
  – Continuity
  – Pricing
3. Respondent protection

• Informed consent
  – Well defined in government surveys
  – Poorly defined in purchased micro-data and web-scraped data

• Privacy
  – Responsibility to indirect respondents
  – Who bears risks?

• Disclosure control
  – If DP, who manages privacy budget?
  – Less researcher access to microdata
    • Reduced transparency and improvements
    • Less total value
Challenges for BLS

• Measuring a dynamic economy
  – Continuity vs. change, speed vs. quality
• U.S. federal budget
  – Declining funding and certainty
• Maintaining participation in surveys
• Confidentiality and trust concerns
  – Squeezing public use files
  – Legislative restrictions on sharing
  – Public attacks on integrity
• Attracting young people to public service
Government statistics: Headlined, not sidelined

• With private sector innovation, the need for statistical agencies is greater!
  – As experts, organizers, and facilitators
  – As the “gold standard”
  – For transparency, weights and continuity

• So is the need for conversations--like this--between us
How BLS already uses Big Data

• Quarterly Census of Employment and Wages
  – Sample frame
  – Firm dynamics
  – Match profit/nonprofit status

• Consumer Price Index
  – Web-scrape characteristics for hedonics
  – Transaction data

• Producer Price Index
  – Bid/ask prices for securities indexes
  – DOT baggage fees

• Survey of Occupational Injuries and Illnesses
  – Auto-code text for illnesses and injuries
6. Mission compatibility

- Curation and continuity
  - Who responsible
  - Standards and incentives
  - Methodology, software upgrades
  - Institutional memory
  - Provider training and commitment

- Pricing
  - Altruism
  - Monopoly power
  - Market fluctuations

- Insider misbehavior
  - Data manipulation
  - Front-running
  - Providers become government agents?
Big Data opportunities for stat agencies

• Reduce respondent burden
• Improve statistics
  – Detail
  – Coverage
  – Timeliness
  – Precision
  – Accuracy
• Raise efficiency and resilience
2. Access

• Costs
  – Evaluation
  – Design
  – Operational
    • Purchase
    • IT hardware and software
    • Oversight, relationship management
    • Quality control and validation

• Legal or administrative restrictions

• Continuity risk
  – Will provider continue to supply data or give adequate notice
  – Back-up source required?
Barriers to government data-sharing

• Sharing restrictions
  – Legal (require legislation to change)
    • E.g., Congressional restrictions on access to IRS data
    • E.g., state ownership and laws (UI wage records, NDNH)
  – Risk aversion + lack of clarity (may just require re-interpretation)
    • Informed consent (e.g., web-scraping)
    • Responsibility for protecting confidentiality

• Resources
  – To curate, document & store data; program code; train staff, etc.
    • e.g., UI wage vs. employer records; UI initial claims

• Priority of operational mission
  – Existence may be hidden
  – Immediate demands impede timeliness, continuity, quality, enhancements, documentation needed for statistical uses
  – Imperfections may be embarrassing