Practicing Data Science in the Government

Ron Prevost
Senior Statistician for Data Integration
Research and Methodology Directorate
U.S. Census Bureau

Roundtable on Data Science Post-Secondary Education: Meeting #3
May 1, 2017 from 9:00am-5:00pm EST
Pew Research Center
1615 L St. NW, Suite 800
Washington, DC

Outline

- General Statistical Agency Challenges
- Transforming Business Processes
- Constraints to Statistical Organization Transformation
- Addressing Knowledge Gaps

Disclaimer: The opinions expressed in this talk are my own.

General Statistical Agency Challenges

- Federal data statistical system data collection and integration challenges
  - Survey response rates dropping & costs rising
  - Information requests complexity growing for more timely, relevant, and local/sector information
  - New/competing products from non-statistical sources
  - Changing conditions requiring new sources and methods
  - Availability and use of alternative (non-survey) data sources to potentially combine with survey data to produce multi-sourced, quality official statistics and products
  - Advent of responsive design techniques, “Big Data” sources, administrative records, and increasingly accurate models

- Survey methodology still necessary, but by itself, is not sufficient to address the challenges
  - Need for additional data collection methods and agile interdisciplinary teams, including mix of survey stats, methodologists, data scientists, operational researchers, math stats, and economists
  - Demands new computing techniques require improved approaches to respondent/provider confidentiality protection

General Statistical Agency Challenges

- Produce information that
  - Is unbiased “gold standard” information
  - Meets the rigors of statistical accuracy
  - Is delivered quickly & cost-effectively
  - Can be used to determine causality
  - Is created by understandable methods
  - Is repeatable & transparent in process and source
  - Protects the confidentiality of its sources

Transforming Business Processes

- Greater focus on model-based estimation and data source acquisition and integration
- Develop models from legacy collection operations

  surveys and administrative records (past\current)
- Use probability samples to validate and correct micro-records, aggregations, and models
- Requires questionnaire revision to collect new information that can be used to measure bias and control/improve models
- Develop a continuous feedback loop for iterative improvement.

Transforming Business Processes

Federal Statistical Data Collection Modernization Drivers

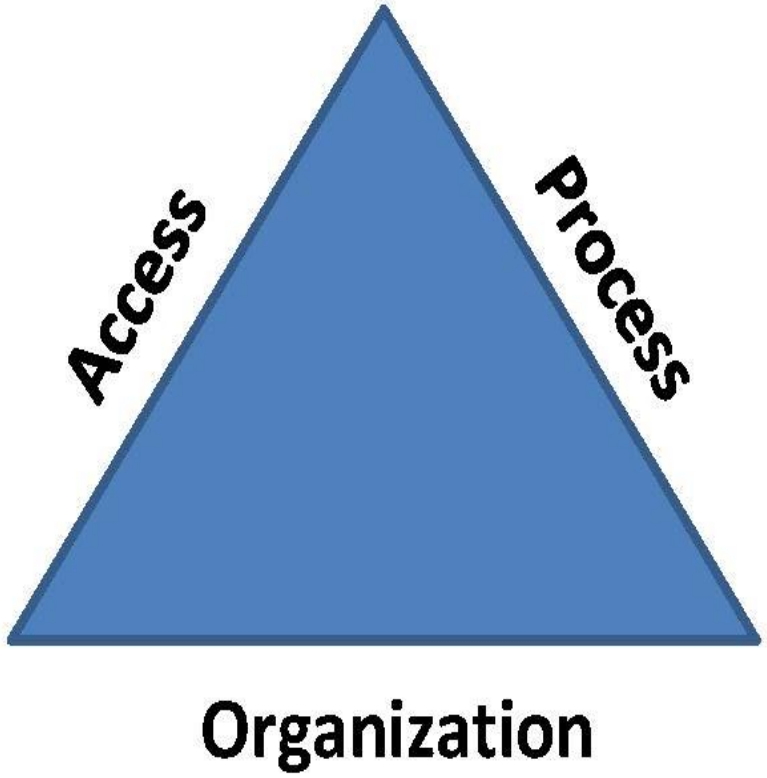
- Scalable methodology to align mix of collection methods to data measurement objectives
- Sufficient knowledge about changing work to determine future staff occupations, competencies, roles, and counts to determine professional development strategies and program investment levels
- Professional Development Program that supports both current and future development through a needs and role-based approach
- Integration Plan that ensures coherence among methods, workforce, and professional development

Constraints to Statistical Organizational Transformation

Access- Logistics driven delivery system, MOU's/agreements, legal authority, metadata availability

Process – IT Security, appropriate hardware/software, project management, system/process governance

Organization – Staff requirements, roles, and competencies, organizational structure, organizational culture



IT Constraints to Statistical Organizational Transformation

- Managing data assets in central curated repositories
- Managing metadata assets
- Managing appropriate use
- Cultural change - training staff to effectively locate and use these assets

IT Constraints to Statistical Organizational Transformation

- Cloud provisioning computational resources
- Security compliance issues with cloud usage
- Security compliance with installation of new software tools – for both production and training
- Sandbox testing with real data
- Application of these components to transform business processes

Professional Development and Workforce Challenges and Proposed Solution

- Rapidly changing technology and insufficient knowledge about changing work to determine how to staff our competency framework to identify future occupations, roles, or counts to determine professional development strategies and program investment levels

- Program use case exercises could help converge on dominant competencies, roles, and staffing levels to project resource requirements for business transformation activities
  - Transformations vary by type
  - Transformations have common phases

Transforming Business Processes

Consolidation	Supplementing or Supplanting	New Product Development	New Capability Development	Optimizing current business processes
• Combining common business processes and systems and generalized solutions	• Supporting current business processes with new process)	• Creating new products based on statutory or customer needs	• Building new capabilities to support business infrastructure (e.g., Strategic Workforce Planning, Portfolio Investment Management, Activity Based Management)	• Improving current processes to improve quality and timeliness

Addressing Knowledge Gaps

- Identified technical competencies for each of four initial academic program areas (statistics, survey methodology, data analytics/data science, and operations research) and began to link to needed competencies to specific courses

- New Production Models
  - Data science
  - Business and Data Analytics
  - Reproducible science

- Supporting Activities
  - Software design and engineering
  - Data storage and retrieval models and distributed computing environments

- Transitioning from research activities to production operations
  - Operations Research

Addressing Knowledge Gaps

- Examined local and online universities with graduate programs and certificates in statistics, survey methodology, operations research, and data analytics/data science

- Examined high-tech companies with expertise in data analytics/data science and operations research for potential sources of training/short courses

- Developed a comprehensive initial catalog of about 80 academic programs (PhD, MS, Citations, Certificates) and training short course sources

Addressing Knowledge Gaps

Local and On-line Courses of Study Included in Our Research

Program	Survey Methods	Statistics	Operations Research	Data Analytics
Certificate	5	6	4	17
Doctorate	1	6	1	1
Master's	2	15	3	17
Short Courses	1	1	0	2
Grand Total	9	28	8	37

Addressing Knowledge Gaps

- Developing approach for baseline and specialized training through Massive Open Online Courses and potential combination with traditional courses
  - Link training to transformation activities
  - Cost effective, quick, flexible, available

  - Capacity for numerous other training courses
  - Potentially couple with degree courses
  - Constraints if software/data dependent – security and IT capacity concerns

Addressing Knowledge Gaps

Training Policy Issues

- Degree/certificate competitive opportunities
- Duty time versus personal time for the catalog courses (consistent practice)
- Continuing service obligations for time and dollars
- Funding source(s)/availability
- Criteria for approving employee requests/selection for courses/programs
- Maintenance and need for distance learning classrooms?

Thank you.

Ronald.C.Prevost@Census.Gov