Business Sources from the Federal Government:
Exciting and New

Federal Depository Library Program: FDLP Academy
March 26, 2020: Webinar

Presenter: Jennifer C. Boettcher, Georgetown University
Jennifer C. Boettcher


M.B.A., Georgetown University, Washington, D.C., 2005
M.L.S., State University of New York, Albany, N.Y., 1992
B.A., University of New Hampshire, Durham, N.H., 1987

ALA RUSA BRASS Member since 1991
ALA GODORT member since 1990
Georgetown University 1997-present
Founder of Business Information Finders (BIF) and Capital Area Business Academic Librarians (CABAL) in DC
2013 Emerald Research Grant: Zombie List (reanimated business sources)
Seeking contributors: [https://boettcher.georgetown.domains/HisBusColl](https://boettcher.georgetown.domains/HisBusColl)
2010 Gale Cengage Learning Award for Excellence in Business Librarianship
Librarian & Information Scientist

• As a Librarian, I
  – Understand the source
  – Know how to find the source
  – Know the related subjects
  – Know how it’s connected to other sources
  – Know how to read it
  – Make connections between publisher and researcher

• As a Librarian, I don’t
  – Publish the primary source
  – Have your context or expertise
  – Do statistical analysis
  – Interpret the data
  – Do data entry
  – Have legal expertise

These are my views and do not reflect those of Georgetown or GPO.
Outline

• Public Domain
• Open Government
• Questions
• Government and Business
• Economic Census
  – General
  – Geography
  – NAICS/NAPCS
  – 2017 Developments
• Questions

• Starting Points
• Major Sources
  – Social Science
  – Natural Science
• Non Federal Sources
  – Sub-national
  – International/NGO
• Fun
• Next Steps
• Questions
Copyright provides the owner of copyright with the exclusive right to

- **Reproduce** the work in copies
- **Prepare** derivative works based upon the work
- **Distribute** copies of the work to the public by sale or other transfer of ownership or by rental, lease, or lending
- **Perform** the work publicly live or by means of a digital transmission
- **Display** the work
- Distribution of **Public Domain collections** called “collective works” or compilations
- Copyright also provides the owner of copyright the right to **authorize others** to exercise these exclusive rights, subject to certain statutory limitations.
Public Domain= No Copyright Restrictions

Public Domain is not protected by intellectual property laws, like copyright. Anyone can use a public domain work without obtaining permission, but no one can ever own it.

Example: no longer protected due to age of creative work.

Works produced for the U.S. Government by its officers and employees should not be subject to copyright. The provision applies the principle equally to unpublished and published works. 17 USC 105

REMEMBER: Public domain data must be attributed.
It should be Public Domain, but I can’t get it

• Privacy Concerns
  – Personally Identifiable Information (PII) and personnel
  – Health Insurance Portability and Accountability Act (HIPAA)
  – Family Educational Rights and Privacy Act (FERPA)

• Security Concerns
  – Generally military and intelligence related - classified
  – Controlled Unclassified Information (CUI)

• Financial Concerns
  – Contains propriety data or trade secrets
  – Requires cost recovery
  – Contracts with creator don’t allow it, licensing issues
  – Civil litigation or law enforcement
  – Unpublished information and data concerning oil wells

Why FOIA Requests?
• Know what you are looking for!
• What to ask for
  – Anything unpublished by US government
  – Controlled Unclassified Information (CUI),
    • replaces “For Official Use Only” and other categories
• Who can ask,
  – OPEN Government Act of 2007, citizen journalist

Other potential access problems
• Not widely distributed/web
• Contracted or grant work
• Lack of supposed interest, DoJ
• Lack of funding, Census
• Not kept, Record Retention
Open Government

• Free
• Transparent
• Accountable
• Accessible to citizens
• Engage all citizens
• Data is Machine Readable,
  – API v Human
• When in doubt openness prevails
Open policy in the Federal Government

• Federal government policy
  – Passed by Congress
  – Implemented by Executive Branch
  – Refined by Courts

• Policy on information collection (priorities)
  – Mandated by law (in CFR) Public Law No: 115-435 (signed 1/14/19)
  – Implemented by regulations (Federal Register) 82 FR 52213
  – Directed by memorandum (Presidential) M-13-13
  – Standard of practice: Data plans (Agencies)
Public Domain Vs. Open Access

• Public Domain
  – US Law
  – Federal Government products
  – can be retrieved by FOIA
  – Some sub-nationals
  – Some countries

• Open Access
  – Decision of publisher/owner
  – Because of ownership of copyright
  – License (CC, GNU, etc)
  – Some NGOs
  – Principles
    • Reuse and redistribution of the data
    • Allows derivative works as Open only
    • No restrictions on who can access and use
    • Electronically transferable
    • Machine-readable
Questions?

Use Cases and User Stories

- Entrepreneurs opening their first business
- Existing business owners looking to expand their markets or better understand them
- Grant writers needing data for their proposal
- Academics doing research for their class
- Regional planners researching business gaps
- Emergency managers and staff doing preparedness, response, and recovery work

Slide provided by Census, 2019
Doing business **WITH** the U.S. Federal Government

### Procurement/Contracts
- **Authoritative site for Assistance Listings, Wage Determinations, and Contract Opportunities** *(beta.sam.gov)*
  - Old: Federal Business Opportunities *(FedBizOpps.gov)* to Replace the *Commerce Business Daily* *(CBD)*
  - Old: Catalog of Federal Domestic Assistance *(CFDA)*

### Disasters Assistance/Grants
- **Disaster Assistance**
- **Federal and State Benefits**
- **Search CFDA**
Federal Government Information for Business

**Executive Branch**
- Regulations
- Data, API
- Data, Statistics or Indicators
- Press releases or Trade press or Bulletins
- Handbooks or Fact Books
- Departments’ Reports
- Industry Reports or Surveys or Reviews
- State of the Industry and Outlooks
- Geographic specific
- New Technology and direction

**Congressional Branch**
- Forthcoming laws
- Committee reports/prints
- Congressional Research Section
- Library of Congress, Reading Rooms
- Government Publishing Office

**Judicial Branch**
- Interpreting laws
- Courts, transcribes
- Courts, defendants

Knowledge Management Cognitive Pyramid

Knowledge Management

Focus:
Create
Organize
Apply
Transfer

Information Management

Focus:
Collect
Process
Disseminate
Store
Display
Protect

DATA

Processing

INFORMATION

Cognition

KNOWLEDGE

Judgment

SHARED UNDERSTANDING

WISDOM

Decision
Risk

Know
Why

Know
How

Know
What

Data are not:
- Information
- Technology
- Digital Analytics
- Evidence
- Research
- Visualizations
- Ideas

Data are collected facts “raw material”
The Economic Census (EC)

- NAICS: Nearly every 2- thru 6-digit code covered by the Census Bureau
  - Excludes Agriculture and other types of business (see [https://www.census.gov/programs-surveys/economic-census/guidance/understanding-naics.html#par_textimage](https://www.census.gov/programs-surveys/economic-census/guidance/understanding-naics.html#par_textimage))
- Geography: National, state, metro, county, place, and ZIP Code levels
- Other characteristics, including business size and Franchise status
- Over 200 data variables shown (common and sector-specific ones)
- Product Lines data (revenue by products sold or manufactured or services provided)

See [https://www.census.gov/programs-surveys/economic-census.html](https://www.census.gov/programs-surveys/economic-census.html) and [https://www.census.gov/data/academy/webinars/2019/2017-econ-census.html](https://www.census.gov/data/academy/webinars/2019/2017-econ-census.html) for more information, including release schedule for 2017 and what’s new.

Note: Census also have programs related to the EC, including the Economic Census of Island Areas, the Survey of Business Owners (and the Annual Survey of Entrepreneurs / Annual Business Survey) and the Commodity Flow Survey

Slide provided by Census, 2019
Series and Reports

Geographic Area Series
  All Geographic data
Puerto Rico & Island Area
Industry Series (US only)
  Product Lines NAPCS X NAICS
  Establishment and Firm Size
  Concentration Ratios
  Legal Form of Organization
Subject Series (ST)
  Miscellaneous Subjects
  Business Expenses
First Look
Franchise Statistics Report
Comparative Statistics Report between 12 & 17
Nonemployer Statistics
Annual Business Survey (ABS)
  Survey of Business Owners (SBO)
  Annual Survey of Entrepreneurs (ASE),
  Business Research and Development and Innovation Survey for Microbusinesses (BRDI-M)
Commodity Flow Survey
Materials Consumed
  Mining and Manufacturing
Nonemployer Statistics

Includes the Gig workers

Businesses w/o paid employees account for
75% of all businesses
3.5% of all sales

Excluded from other census reports

Updated annually

Gig Economy
Geographic/TIGERweb
What is new with 2017 Geography?

https://www.census.gov/programs-surveys/economic-census/geographies/change-notes.html

- Metro Areas – the Core Based Statistical Areas in 15 states had some type of change
- Counties – Alaska, South Dakota, and Virginia had changes to 1 or more counties
- Economic Places – every state had places with area gain, loss, and/or code or name changes. There are 442 new Economic Places being recognized and 171 places that are being dropped for 2017, primarily due to population decline
- Offshore for drilling and mining along the coasts. <3 miles with state data
- For example: North Carolina has 474 Total Places

<table>
<thead>
<tr>
<th>Gains</th>
<th>Losses</th>
<th>Name Change</th>
</tr>
</thead>
<tbody>
<tr>
<td>245 area size</td>
<td>109 area size</td>
<td>3 name changes</td>
</tr>
<tr>
<td>22 places</td>
<td>6 places</td>
<td></td>
</tr>
</tbody>
</table>
Which Industry?
February 26, 2020 OMB asked for permission to kill the Standard Industrial Classification (SIC). Long live the North American Industry Classification System (NAICS) 85 FR 11120, parts V and VI

V. OMB Statistical Policy Directive No. 8, Standard Industrial Classification of Establishments The ECPC is soliciting public comments at the request of OMB on the advisability of formally updating Statistical Policy Directive No. 8, Standard Industrial Classification of Establishments, and seeking comments on the proposed text of the update.

NAICS Sectors

11 Agriculture, Forestry, Fishing and Hunting
21 Mining
22 Utilities
23 Construction
31-33 Manufacturing
42 Wholesale Trade
44-45 Retail Trade
48-49 Transportation and Warehousing
51 Information
52 Finance and Insurance
53 Real Estate and Rental and Leasing
54 Professional, Scientific, and Technical Services
55 Management of Companies and Enterprises
56 Administrative and Support and Waste Management and Remediation Services
61 Education Services
62 Health Care and Social Assistance
71 Arts, Entertainment, and Recreation
72 Accommodation and Food Services
81 Other Services (except Public Administration)
92 Public Administration
# North American Industry Classification System

## Example

<table>
<thead>
<tr>
<th><strong>Level</strong></th>
<th><strong>Code</strong></th>
<th><strong>Description</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Sector</td>
<td>52</td>
<td>Finance and Insurance</td>
</tr>
<tr>
<td>Subsector</td>
<td>524</td>
<td>Insurance Carriers &amp; related activities</td>
</tr>
<tr>
<td>Industry Group</td>
<td>5241*</td>
<td>Insurance Carriers</td>
</tr>
<tr>
<td>Industry</td>
<td>52411</td>
<td>Direct Life Health &amp; Med Ins Carriers</td>
</tr>
<tr>
<td>U.S. Industry</td>
<td>524113</td>
<td>Direct Life Insurance Carriers</td>
</tr>
</tbody>
</table>

* Mexico & Canada Compatible
What and why NAPCS?

What is the purpose of NAPCS?
Comprehensive demand/market-oriented classification framework for services and goods.
NAPCS was developed as a reference classification to facilitate the analysis of economic data based on markets rather than production processes and allows comparisons of products among the three North American countries.

What is the relationship between NAPCS and NAICS?
NAPCS, a product classification system, and NAICS, an industry classification system, are independent but complementary.
A product produced/sold/distributed by multiple industries carries the same title, definition, and code in NAPCS, regardless of its industries of origin.
Products can be linked to the industries that produce/sell/distribute them, and industries can be linked to the products.

What else can NAPCS be used for?
BEA’s National Accounts (NIIP), Harmonized Commodity Description and Coding System (HS) 2012 and Extended Balance of Payments Services (EBOPS) 2010.
North American Product Classification System 2017 (NAPCS 2017) version Beta 1.0

Hierarchical Structure

<table>
<thead>
<tr>
<th>Level</th>
<th>Digits</th>
<th>Aggregation level</th>
<th>Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>Section</td>
<td>24</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>Subsection</td>
<td>61</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>Division</td>
<td>172</td>
</tr>
<tr>
<td>4</td>
<td>7</td>
<td>Group</td>
<td>276</td>
</tr>
<tr>
<td>5</td>
<td>9</td>
<td>Subgroup</td>
<td>497</td>
</tr>
<tr>
<td>6</td>
<td>11</td>
<td>Trilateral product</td>
<td>1167</td>
</tr>
</tbody>
</table>

Example (library services)

<table>
<thead>
<tr>
<th>Code</th>
<th>Product title</th>
</tr>
</thead>
<tbody>
<tr>
<td>24</td>
<td>Home entertainment, recreation and culture products, including household pets and related products</td>
</tr>
<tr>
<td>241</td>
<td>Home entertainment, recreation and culture products</td>
</tr>
<tr>
<td>24110</td>
<td>Other home entertainment, recreation and culture products</td>
</tr>
<tr>
<td>2411001</td>
<td>Other home entertainment, recreation and culture products</td>
</tr>
<tr>
<td>241100101</td>
<td>Gambling products</td>
</tr>
<tr>
<td>241100102</td>
<td>Leisure, recreational, and athletic instructional programs</td>
</tr>
<tr>
<td>241100103</td>
<td>Library and archive services</td>
</tr>
<tr>
<td>24110010301</td>
<td>Library and archive services</td>
</tr>
</tbody>
</table>
Subject Series NAPCS x NAICS

• Used to be
  – Merchandise Line
  – Product Line
  – Revenue Line
  – Service Line

Other Changes with the 2017 Economic Census

• New disclosure rules -- new privacy rules will result in the establishment count being suppressed when fewer than 3 establishments in that industry in that geography.
  – You will be seeing a lot more of D Withheld to avoid disclosing data for individual companies; data are included in higher level totals

• Local areas published by sector – Place-level data will no longer be available for the Manufacturing sector and the NAICS and geographic levels published for other sectors may be adjusted based on data quality and privacy issues.
  – Fewer geographic details

• Introduction of the Annual Business Survey (ABS) -- The Survey replaces three existing surveys: the five-year Survey of Business Owners (SBO) for employer businesses, the Annual Survey of Entrepreneurs (ASE), and the Business Research and Development and Innovation Survey for Microbusinesses (BRDI-M).
Ways to Access Data

- **QuickFacts**
- **My Congressional District**
- **Global Market Finder**
- **Census Business Builder**
- **OnTheMap** and **OTM for Emergency Management**
- **American FactFinder** [going away]
- **USA Trade Online**
- **data.census.gov**
- **DataFerrett** [gone away]
- Application Programming Interface (API)

https://www.census.gov/programs-surveys/economic-census/guidance/finding-data.html

Slide provided by Census, 2019
What is not covered in the Economic Census

<table>
<thead>
<tr>
<th>NAICS</th>
<th>Industry categories excluded from the economic census</th>
<th>Coverage in CBP</th>
</tr>
</thead>
<tbody>
<tr>
<td>11</td>
<td>Agriculture, Forestry, Fishing and Hunting</td>
<td>CBP (except 111,112)</td>
</tr>
<tr>
<td>482</td>
<td>Rail Transportation</td>
<td></td>
</tr>
<tr>
<td>491</td>
<td>Postal Service</td>
<td></td>
</tr>
<tr>
<td>525</td>
<td>Funds, Trusts, and Other Financial Vehicles</td>
<td>CBP (except 5251)</td>
</tr>
<tr>
<td>6111</td>
<td>Elementary and Secondary Schools</td>
<td>CBP (private schools only)</td>
</tr>
<tr>
<td>6112</td>
<td>Junior Colleges</td>
<td>CBP (private schools only)</td>
</tr>
<tr>
<td>6113</td>
<td>Colleges, Universities, and Professional Schools</td>
<td>CBP (private schools only)</td>
</tr>
<tr>
<td>8131</td>
<td>Religious Organizations</td>
<td>CBP</td>
</tr>
<tr>
<td>81393</td>
<td>Labor Unions and Similar Labor Organizations</td>
<td>CBP</td>
</tr>
<tr>
<td>81394</td>
<td>Political Organizations</td>
<td>CBP</td>
</tr>
<tr>
<td>814</td>
<td>Private Households</td>
<td></td>
</tr>
<tr>
<td>92</td>
<td>Public Administration</td>
<td></td>
</tr>
</tbody>
</table>
Economic Indicators from Census

County Business Patterns
Annual Survey of Manufacturers
Advance Monthly Retail Sales
Monthly Wholesale Trade
Manufacturers’ Shipments, Inventories and Orders
Quarterly Services Survey
Value of New Construction Put in Place
Housing Completions
Housing Vacancies and Homeownership
US International Trade in Goods and Services
Quarterly Financial Report
E-Commerce Monthly
Census of Governments – Public Sector
Public Education Finances
Where to start for Open Data

- United States [https://www.data.gov](https://www.data.gov)
  - Advanced
- States [http://copyright.lib.harvard.edu/states](http://copyright.lib.harvard.edu/states)
- Inter-university Consortium for Political and Social Research repository [https://www.icpsr.umich.edu/icpsrweb/ICPSR/](https://www.icpsr.umich.edu/icpsrweb/ICPSR/)
  - Themes
Major Sources of Social Science Information from the US Government

- **Census** (Demographics, Public Finance, Manufacturing, Retail, Wholesale, Construction, all other industries)
- **Labor**
- **Companies**
- **Economic**
- **Federal Budget Projections**
- **Taxes**
- **Spending**
- **Social Security**
- **Justice**
- **Banking**
- **Mortgage**
- **Trade**
- **Transportation**
- **Telecommunications**
- **Education**
- **Libraries and Museums**
- **Economic Analysis**
Major Sources of Natural Science Information from the US Government

- Agriculture and Fisheries
- Defense
- Health
- Energy
- Environment
- Intellectual Property
- Land Management and Minerals
- Pharmaceuticals
- Science & Engineering
- Weather

https://www.flickr.com/photos/notbrucelee/6897137283/in/photostream
Problems that come with government Info

- Beggars can’t be choosers
  - Too old
  - Not to the geographic level needed, going to get worst
  - Too detailed
  - Have to file a FOIA request
States and Cities

• Public Domain in California and New York City
• https://www.google.com/search?q=site%3A.gov+"open+data"+[YOUR STATE HERE]
• Remember to search using site:.gov
• There are a lot of consultants making money on cool visualizations using open data.
## Major International Data Sources

### By topic

<table>
<thead>
<tr>
<th>Topic</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social &amp; Economic</td>
<td><a href="http://data.un.org">http://data.un.org</a></td>
</tr>
<tr>
<td>Human body</td>
<td><a href="http://data.un.org">http://data.un.org</a></td>
</tr>
<tr>
<td>Labor</td>
<td><a href="http://data.un.org">http://data.un.org</a></td>
</tr>
<tr>
<td>Telecommunications</td>
<td><a href="http://data.un.org">http://data.un.org</a></td>
</tr>
<tr>
<td>Governance</td>
<td><a href="http://data.un.org">http://data.un.org</a></td>
</tr>
<tr>
<td>Developed Countries</td>
<td><a href="http://data.un.org">http://data.un.org</a></td>
</tr>
</tbody>
</table>

### By Country

<table>
<thead>
<tr>
<th>Source</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>National Statistical Offices</td>
<td>More data available in national language</td>
</tr>
<tr>
<td></td>
<td>Some charge for access</td>
</tr>
<tr>
<td></td>
<td>Citizens of that country might have free access</td>
</tr>
<tr>
<td>National Repositories/Archives</td>
<td>Historical</td>
</tr>
<tr>
<td></td>
<td>Datasets</td>
</tr>
</tbody>
</table>
P.E.S.T. Analysis for Industry*

• Political
  Legislative
    Congress.gov
  Executive
    Regulations.com
  Judicial
    United States Courts

• Economic
  Sector Inflation
    BLS’s Producers Price Index
  Microeconomic trends
    FRB’s FRED
    Census BEA’s Interactive Tables

• Socio-cultural
  Norms & Ratios
    IRS’s Statistics of Income
  Peers and partners
    Economic Census
    Annual Surveys from Census

• Technology
  Patents
    Citation Analysis
  Tech Transfer
    Government funded research to license

* Scanning the Business Environment (Aguilar)
• Create maps of our state with all the geography in it.
  – Make it a puzzle with 3-D printers
  – Bring it to gerrymandering parties
• Invite the Historical Society to track the industries in your communities.
  – Have Wikipedia parties with new data
• Contact the Chamber of Commerce and other clubs
  – give them quiz questions
• Have a release party with your State Data Center
• Dedicate a computer just to entrepreneurship with CBB and relevant links.
• Arm citizen journalists with the Census of Government and School data.

• YOUR idea here
# Open Government

## US Federal
- [DATA.GOV](https://www.data.gov)
- [Performance.gov](https://performance.gov)
- [Center for Effective Government](https://www.centerforeffectivegov.org) (formerly OMB Watch)
- [Government Accountability Project](https://www.gagov.org)
- [Open Government Foundation](https://www.opengovfoundation.org)

## International
- [Center for Data Innovation](https://www.centerfordatainnovation.org)
- [Data Transparency Coalition](https://www.datatransparencycoalition.org)
- [Open Government Partnership](https://www.opengovernmentpartnership.org)
- [Open Knowledge](https://www.okfn.org)
- [Sunlight Labs](https://www.sunlightlabs.com)
Learning more

Government Sources

FDLP Academy
Accidental Government Librarian
DigitalGov from Digital Government Division of GSA
Standards for Born Digital images

Numerical Data

• Inter-university Consortium for Political and Social Research (ICPSR) repository
• API University from OMB
• DataRefuge

Public Knowledge: Access and Benefits (Information Today, 2016)
Where to learn **MORE**

**From Librarians**

- ALA ‘s Government Documents Round Table (GODORT)
- International Association for Social Science Information Services and Technology (IASSIST)
- ACRL/DSS Numeric and Geospatial Data Services Discussion Group
- Scholarly Publishing and Academic Resources Coalition (SPARC)

**From Federal data**

- Government Printing Office (GPO)
- Association of Public Data Users (APDU)
- Council of Professional Associations on Federal Statistics (COPAFS)
- American Statistical Association
- Project Open Data
- Data Coalition
Let’s discuss

boettcher@georgetown.edu

202 687-7495
Twitter: @jennywombat

These slides are licensed under a Creative Commons Attribution-NonCommercial 4.0 International License. © Bill Waterhouse, with permission

All links last checked 3/25/20
Selected Boettcher bibliography

Books

Reviews

Publications


Free Webinars


One Statistical Office in US: Why Not?

1. **Privacy:** The Privacy Act of 1974, Confidential Information Protection and Statistical Efficiency Act of 2002 (CIPSEA), and Statistical Policy Directive No. 1 (2014) require agencies to ensure that the collection and maintenance of citizens' data is accurate, confidential, and within legal restrictions. With different offices having access to those records, there would be less possibility of everything being leaked.

2. **Security:** Along the lines of fewer offices having access to data records. The more servers that hold the data, the safer it is. The times when an exchange of information is necessary laws and regulations among departments allow to protect access to data.

3. **Integrity:** The income you report to IRS might be different from what you report to the Census Bureau.

4. **Methodology:** Sometimes data must have a higher number of people questioned so the accuracy will be better; different methods of collection or sampling may be required.

5. **Popularity:** Anything being done by the government has a political dimension, especially funding for employees and for modernizing and updating technology, attractiveness of the research, and repetition of statistical programs by agencies.
Vocabulary: Tools, Process, and Products

**Datasets or compilation**: Raw or statistical numbers, can be flat file such as Comma Separated Variable (CSV) or proprietary like Excel

**Metadata**: Includes field descriptions for the dataset, found in codebooks

**Schema**: How data is organized or structured using standards, like classification

**Application Program Interface (API)**: Read-only machine to machine querying, generally from JSON or XML files

**Big data**: Raw, unstructured data; normally transactional (example: each check out)

**Natural Language Processing (NLP)**: Use for text analysis, not numeric data

**Artificial Intelligence (AI)**: Includes predictive analytics and machine learning

**Reports**: Usually aggregated statistics based on big data (example: how many checkouts)

**Data Visualization**: Using software to visually communicate relationships and context of data

**Open Data**: Freely accessible data, created for a specific purpose; by-product of decision making or research
Works produced for the U.S. Government: Lifecycle of Data

**Policy Makers** who ask the questions about what has to be found or measured

**Researchers** who design methods or experiments to collect the data and where the data and codebooks are created.

**Statisticians** who manipulate datasets using models and algorithms to see trends in longitudinal data and to interpret data at a moment of time in cross-sectional studies.

**Analysts** who see patterns using predictive analytics, seek the emerging relationships between the numbers, transforming data into information by giving it context.

Other **Data Scientists** will link graphics, statistical downloads, and application programming interfaces (APIs) to the researcher's raw data.

**Writers** and **Data Visualization Designers**, who uses their imagination to apply their knowledge to make data understandable in reports, press releases, and other resources.

The federal agency will act as **Publishers**, putting the synthesized resources on its website for all—primarily for **decision makers** but also for **citizens**—to read.
## Funding for Federal Data Collection

<table>
<thead>
<tr>
<th>Research Agency (amounts in millions of dollars)</th>
<th>FY18</th>
<th>FY19</th>
<th>Change from FY17</th>
<th>Request</th>
<th>Change from FY18</th>
</tr>
</thead>
<tbody>
<tr>
<td>NIH</td>
<td>30311</td>
<td>32311</td>
<td>34229</td>
<td>37084</td>
<td>8.3%</td>
</tr>
<tr>
<td>NSF</td>
<td>7344</td>
<td>7463</td>
<td>7472</td>
<td>7767</td>
<td>4.0%</td>
</tr>
<tr>
<td>AHRQ</td>
<td>364</td>
<td>334</td>
<td>324</td>
<td>334</td>
<td>3.1%</td>
</tr>
<tr>
<td>FDA</td>
<td>2597</td>
<td>2730</td>
<td>2771</td>
<td>2812</td>
<td>1.5%</td>
</tr>
<tr>
<td>Statistical Agency (amounts in millions of dollars)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BEA</td>
<td>96.3</td>
<td>105.1</td>
<td>103.8</td>
<td>99.0</td>
<td>-4.9%</td>
</tr>
<tr>
<td>BJS</td>
<td>41.0</td>
<td>41.0</td>
<td>45.5</td>
<td>48.0</td>
<td>5.5%</td>
</tr>
<tr>
<td>BLS</td>
<td>592.2</td>
<td>609.0</td>
<td>609.0</td>
<td>612.0</td>
<td>0.5%</td>
</tr>
<tr>
<td>BTS</td>
<td>26.0</td>
<td>26.0</td>
<td>26.0</td>
<td>26.0</td>
<td>0.0%</td>
</tr>
<tr>
<td>Census</td>
<td>1068.0</td>
<td>1370.0</td>
<td>1457.0</td>
<td>2814.0</td>
<td>93.1%</td>
</tr>
<tr>
<td>EIA</td>
<td>117.0</td>
<td>122.0</td>
<td>122.0</td>
<td>125.0</td>
<td>2.5%</td>
</tr>
<tr>
<td>ERS</td>
<td>85.4</td>
<td>85.4</td>
<td>86.8</td>
<td>86.8</td>
<td>0.0%</td>
</tr>
<tr>
<td>NASS</td>
<td>172.4</td>
<td>168.4</td>
<td>171.2</td>
<td>191.7</td>
<td>12.0%</td>
</tr>
<tr>
<td>NCES</td>
<td>232.1</td>
<td>261.0</td>
<td>258.5</td>
<td>258.5</td>
<td>0.0%</td>
</tr>
<tr>
<td>NCHS</td>
<td>155.4</td>
<td>160.4</td>
<td>160.4</td>
<td>160.4</td>
<td>0.0%</td>
</tr>
<tr>
<td>NCSES</td>
<td>58.3</td>
<td>58.3</td>
<td>59.7</td>
<td>59.8</td>
<td>1.2%</td>
</tr>
<tr>
<td>ORES</td>
<td>29.0</td>
<td>25.9</td>
<td>24.0</td>
<td>31.0</td>
<td>29.2%</td>
</tr>
<tr>
<td>SOI</td>
<td>36.0</td>
<td>37.0</td>
<td>34.3</td>
<td>33.6</td>
<td>-2.0%</td>
</tr>
</tbody>
</table>

NIH - National Institutes of Health (HHS)
NSF - National Science Foundation
AHRQ - Agency for Healthcare Research & Quality (HHS)
FDA - Food & Drug Agency (HHS)
BEA - Bureau of Economic Analysis (DoC)
BJS - Bureau of Justice Statistics (DoJ)
BTS - Bureau of Transportation Stat. (DoT)
Census - DoC
EIA - Energy Information Admin. (DoE)
ERS - Economic Research Service (DoA)
NASS - Nat. Agricultural Stat. Service (DoA)
NCES - Nat. Center of Education Stat. (DoE)
NCHS - Nat. Center for Health Stat. (HHS)
NCSES - Nat. Center for Science and Engineering Sat. (NSF)
ORES - Off. of Research, Evaluation, and Statistcs (SSA)
SOI - Statistics of Income (IRS)

AMSTAT image from http://magazine.amstat.org/blog/2018/05/01/fy18feddbudget
Future of the Bureau of Labor Statistics


Protected
Principal Federal Economic Indicators (PFEI) and programs written into or referenced by law for allocation or other purpose. 85% of budget

Will it move to Commerce?

Whitehouse Plan
Center for Data Innovation

AMSTAT image from http://magazine.amstat.org/blog/2018/05/01/fy18fedbudget
OMB’s Statistical Policy Directive No. 1
Executive agencies must:

1. Produce and disseminate relevant and timely information
2. Conduct credible and accurate statistical activities
3. Conduct objective statistical activities
4. Protect the trust of information providers
Legal issues

Data and IP

- https://www.lib.umn.edu/data-management/copyright
- https://data.research.cornell.edu/content/intellectual-property

Licensing Data

- http://opendefinition.org/guide/data
- http://library.duke.edu/data/guides/data-management/copyright-licensing