

# Federal Data Strategies Webinar

FDLP Academy: Engage and Empower through Education

GPO's Federal Depository Library Program

February 7, 2019

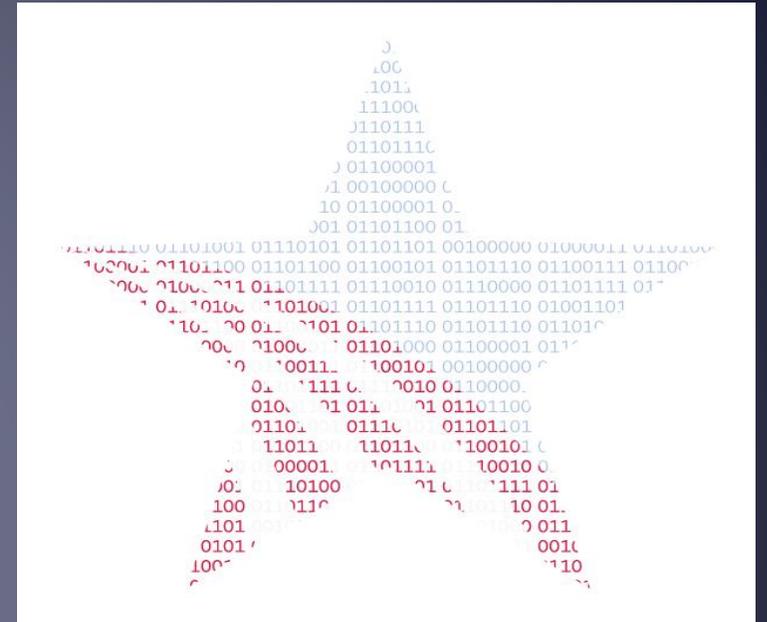
Creator and Presenter: Jennifer C. Boettcher, Georgetown University

Recording will be available on the FDLP website



# What will be covered?

- › Who I am and what I do
- › Data in context
- › Who Creates Data?
- › What is Public Domain and Open Data?
- › Who Creates Federal Data Policy?
- › Current: DATA Act - DAIMS, LEI
- › Changing: OPEN Government Data Act
- › Hopeful changes: Funding, GREAT Act, role for FDLP Librarians
- › Where to find the data



<https://www.mcc.gov/initiatives/initiative/open>

# Jennifer C. Boettcher

Georgetown University 1997-present



M.B.A., Georgetown University, Washington, D.C., 2005

M.L.S., State University of New York, Albany, N.Y., 1992

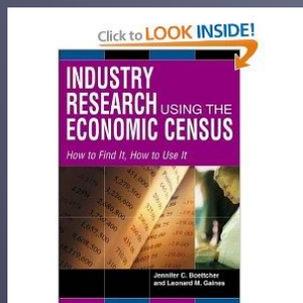
Founder of Business Information Finders (BIF) and Capital Area Business Academic Librarians (CABAL) in DC

2013 Emerald Research Grant: Zombie List (reanimated business sources)

2010 Gale Cengage Learning Award for Excellence in Business Librarianship

## Selected Research

- Jennifer C. Boettcher and Leonard M. Gains. Industry Research Using the Economic Census. Greenwood Press: Phoenix, AZ. 2004
- Jennifer C. Boettcher. “Understanding the Census of Government: 2nd Largest Industry in the U.S.” FDLP Academy, January 2017.
- Jennifer C. Boettcher and K. Matthew Dames. “Government Data as Intellectual Property: Is Public Domain the Same as Open Access?” Online Searcher
- “*Guardians of the (Financial) Galaxy: How Financial Regulators Generate Free Data, Tools, and Information Along the Way,*” ALA Annual, Washington, DC June 21, 2019



# 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

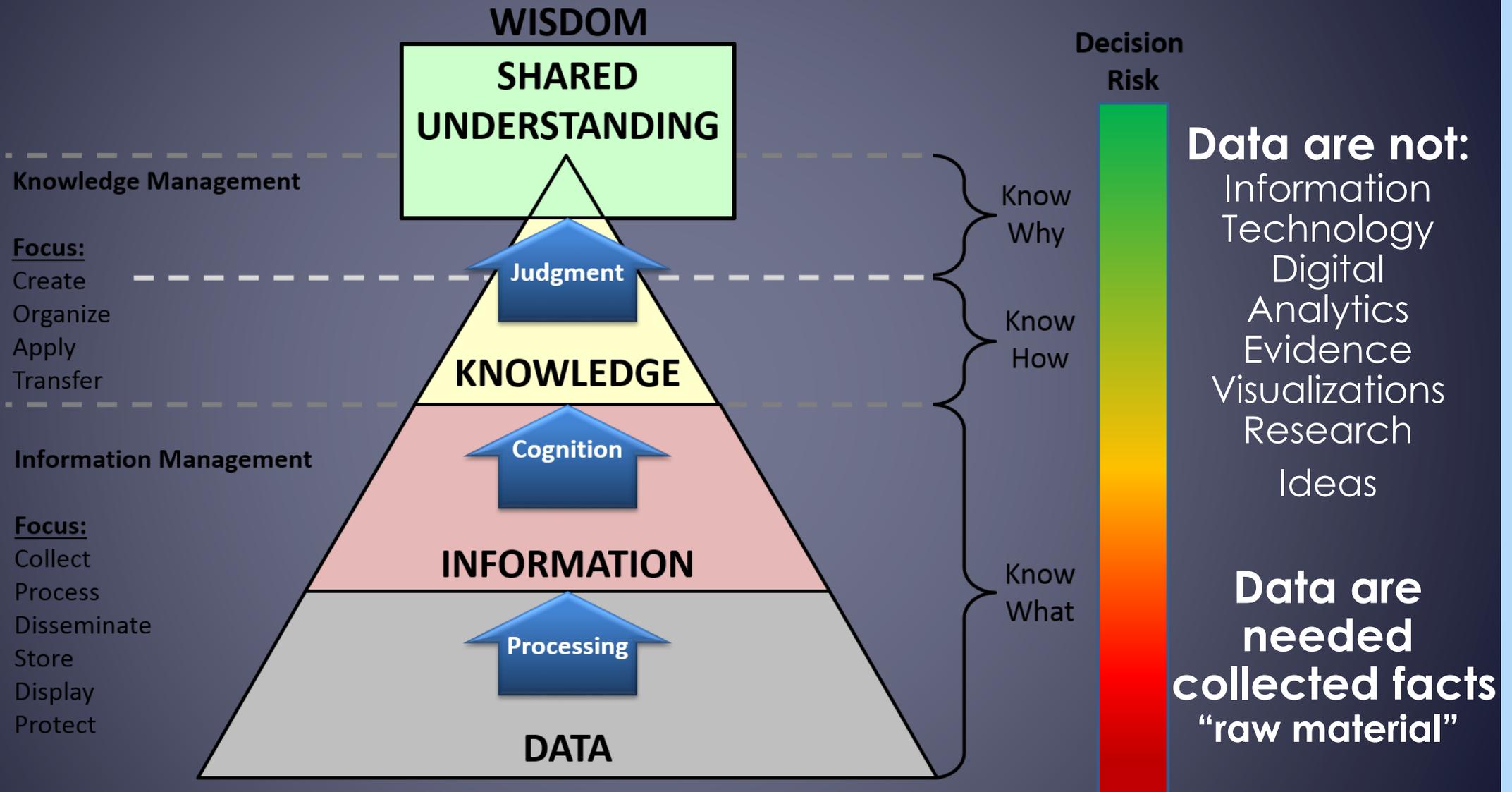
- Have your context or expertise
- Publish the primary source
- Do data entry
- Do statistical analysis
- Make decisions on the data
- Have legal expertise

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

# Should Federal data be open?

- › Free
  - › Transparent
  - › Accountable
  - › Accessible to citizens
  - › Engages all citizens
  - › Machine readable: visualizations, dashboards, and analytical tools
  - › Human readable: simple look ups
  - › When in doubt openness prevails
- › **Why isn't it already?**
    - Classified
    - Personal Information (PII)
    - Not sent to GPO or data.gov
    - Not widely distributed/web
    - Stuck in closed system
    - Perceived lack of supposed interest
    - Not kept
    - **Lack of funding**

# Knowledge Management Cognitive Pyramid



Adaptations of DIKW pyramid by US Army Knowledge Managers, from [https://en.wikipedia.org/wiki/DIKW\\_pyramid](https://en.wikipedia.org/wiki/DIKW_pyramid)

# Works Produced for the U.S. Government: People in the Lifecycle of Data are Data Scientists

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

**Researchers** design methods or experiments to collect the data and create the data and codebooks.

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

**Information Technologists** link graphics, statistical downloads, and application programming interfaces (APIs) to the researcher's raw data.

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

The federal agency is the **Publisher**, putting the synthesized resources on its website for all: primarily for **decision makers** but also for **citizens**, to read.

# 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

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

**Algorithm:** set of rules to be followed in calculations or other problem-solving operations by a computer

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

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

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

**Distributed Ledger Technology (DLT):** **Blockchain** is a type of DLT

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

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

# Copyright provides the owner the exclusive right to:

- › **Reproduce** the work in copies
- › **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
- › Prepare **derivative works** based upon the work
- › **Display** the work
- › Distribute “**collective works**” or compilations
- › **Authorize others** to exercise these exclusive rights, subject to certain statutory limitations

# Copyright and Numeric Data

**Facts are not copyrighted** (In no case does copyright protection for an original work of authorship extend to any idea, procedure, process, system, method of operation, concept, principle, or discovery, regardless of the form in which it is described, explained, illustrated, or embodied in such work.) [17 USC 102b](#)

**In US, collections of facts or data that fail to meet the minimum threshold of creativity also are ineligible for copyright protection,** even if assembling such a collection takes significant time, effort, or resources: “sweat of the brow.”

**Creative expression of data in compilation is protected** (Feist 1991)

Jennifer C. Boettcher and K. Matthew Dames. “Government Data as Intellectual Property: Is Public Domain the Same as Open Access?” [Online Searcher](#), vol. 42, no. 4, (July/August 2018): 42-48.

# Public Domain: No Copyright Restrictions

**Public Domain:** when a creative work is **not protected by intellectual property laws** such as copyright, trademark, or patent laws. The public owns these works, not an individual author or artist. 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](#)

**Includes US Federal Government-produced or funded data**

REMEMBER: Public domain data must be attributed.

# What makes it Open Data?\*

- › **Availability and Access:** the data must be available as a whole and at no more than a reasonable reproduction cost, preferably by downloading over the internet.
- › **Re-use and Redistribution:** the data must be provided under terms that permit re-use and redistribution including intermixing with other datasets.
- › **Universal Participation:** everyone must be able to use, re-use and redistribute - there should be no discrimination against fields of endeavour or against persons or groups.

– \* from Open Data Handbook, <http://opendatahandbook.org>

# What is Creative Commons?

“**Creative Commons (CC)** is an American non-profit organization devoted to expanding the range of creative works available for others to build upon legally and to share. The organization has released several copyright-licenses known as Creative Commons licenses free of charge to the public. These licenses allow creators to communicate which rights they reserve, and which rights they waive for the benefit of recipients or other creators.” Wikipedia, 11/5/18



[creativecommons.org/share-your-work/public-domain](https://creativecommons.org/share-your-work/public-domain)

# Why Open Data Exists is about access without restrictions, like payments

- › Funded research created for a specific purpose
  - US national and some state data
  - Some [Other Countries](#)
  - Non-Government Organizations (NGO)
  - Grants (mainly scientific, e.g. PubMed Central)
  - Publisher-required (mainly scientific, e.g. Science)
- › By-product of research used in decision making
- › Open Access is not intellectual property law. It's a license agreement from the copyright owner and a set of principles: [CC0](#)

# Why Open Access to Government Data?

**Internal Management:** Fraud, waste, and abuse detection through more examination and direct feedback enhance productivity.

**Transparency:** Increases access, decreases FOIA requests, and decreases redundancy.

**Automation:** Saves time for mission, operations, curation, collaboration, and collection, reduce compliance cost, and lets people play with the data.

Provides Decision Makers and Citizens with FACTS

# 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
- 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

# Administrative Data and the Freedom of Information Act (FOIA) 5 U.S.C. § 552, 1966

- › What to ask for
  - Anything unpublished by US government
  - Controlled Unclassified Information (CUI)

[Read this from National Archives and Records Administration](#)

File here [FOIAonline](#)

Help with [FOIA.gov](#)

Oversight: [NARA's Office of Government Information Services](#)

OPEN Government Act of 2007

Citizen Journalist

20 day response

## Watchdogs

[Project On Government Oversight \(POGO\)](#)

[Reporters Committee Freedom of the Press](#)

[FOIAProject](#)

[PublicCitizen](#)

[MuckRock](#)

# States and Public Domain

Some States have data in public domain: California, Indiana\*, Louisiana\*, Florida, North Carolina\*, New Jersey\*, New York\*, Massachusetts, Ohio\*, and Virginia

\*Check with issuing agency

[copyright.lib.harvard.edu/states](http://copyright.lib.harvard.edu/states)

› Works of the governments of the District of Columbia, the Commonwealth of Puerto Rico, and the rest of the organized territories (Guam, Northern Mariana Islands, and the U.S. Virgin Islands) are considered U.S. government works and therefore have no copyright protection.

› U.S. Copyright Office, Compendium of U.S. Copyright Office Practices, Third Edition (2014), Section 313.6.

Questions?



CC0, <https://pixabay.com/en/hedgehog-child-young-hedgehog-1759027>

# Public Domain vs. Open Access

## › Public Domain

- Copyright Law
- No restrictions
- More powerful than Open Access
- Government products
- Data at any stage can be retrieved by FOIA
- Not for some sub-nationals

## › Open Access

- Because of ownership of copyright
- Principles and license
- No restrictions on who can access and use data released
- Allows derivative works as open only
- Electronically transferable
- **Machine-readable**

“..not only strengthens our democracy and promotes efficiency and effectiveness in government, but also has the potential to create economic opportunity and improve citizens’ quality of life.” [Project Open Data](#)

# Data Policy in the Federal Government

- › Federal government policy
  - Passed by Congress
  - Implemented by Executive Branch/Whitehouse
  - Refined by Courts
- › Policy on data collection (top to bottom hierarchy)
  - Mandated by law (in CFR) H.R.4174 - Foundations for Evidence-Based Policymaking Act of 2017, Public Law: 115-435
  - Implemented by regulations (Federal Register) 82 FR 52213
  - Directed by memorandum (Presidential) M-13-13
  - Standard of practice: Data plans (Agencies)

# Data Policy from the OMB

- › Office of Management and Budget
  - 1<sup>st</sup> responsibility is to create the Presidential Budget
  - OMB evaluates the effectiveness of agency programs, policies, and procedures, assesses competing funding demands among agencies, and sets funding priorities.
- › Oversight of paperwork and statistical gathering (1980)
  - Chief Statistician of the United States: Nancy Potok
  - Interagency Council on Statistical Policy
  - Advisory Committee on Data for Evidence Building
  - Each agency will have a Chief Data, Chief Information, Chief Privacy, Chief Performance, and Evaluation Officers.

# OMB Statistical Policy Directive No. 1

79 FR 71609

- Produce and disseminate relevant and timely information
- Conduct credible and accurate statistical activities
- Conduct objective statistical activities
- Protect the trust of information providers
- Govern and manage data as a strategic asset
- Protect and secure data
- Promote efficient use of data assets
- Build a culture that values data as an asset
- Honor stakeholder input and leverage partners

# Other Presidential Management Agenda (PMA) Cross Agency Priority (CAP) @ Performance.Gov

- › Modernize IT to Increase Productivity and Security
- › Improving Customer Experience with Federal Services
- › Sharing Quality Services
- › Shifting From Low-Value to High-Value Work
- › Leveraging Common Contracts and Best Practices
- › Oriented Accountability for Grants
- › Getting Payments Right
- › Federal IT Spending Transparency
- › Improve Management of Major Acquisitions
- › Modernize Infrastructure Permitting
- › Improve Technology Transfer

# Leveraging Data as a Strategic Asset

The Federal Data Strategy will define Principles, Practices, and a Year 1 Action Plan to deliver a more consistent approach to federal data stewardship, use, and access government wide.

Programmatic/Mission data  
(material)

Support data (operational)

Statistical data (curated)

- › **Enterprise Data Governance**
- › **Access, Use, and Augmentation**
- › **Decision Making & Accountability**
- › **Commercialization, Innovation, and Public Use**

[strategy.data.gov](https://strategy.data.gov)

# Digital Accountability and Transparency Act (DATA) Act of 2014: Federal Spending only

1. Treasury & OMB create and maintain standard data elements & formats.
    - › [Bureau of Fiscal Services](#)
    - › [USASpending.gov](#)
  2. Create a schema to use data elements.
    - › [DATA Act Information Model Schema \(DAIMS\)](#)
  3. Tie performance reporting to grant spending.
    - › [National Information Exchange Model \(NIEM\)](#)
  4. OMB oversees a fully automated pilot program (Section 5) for grants & contracts to checkbook level replacing FPDS and FAADS and CFFR
    - › [Common Data Element Reporting \(CDER\) Library](#)
    - › [Legal Entity Identifier \(LEI\)](#)
    - › [eXtensible Business Reporting Language \(XBRL\)](#)
    - › [datalab.usaspending.gov](#)
- [Problems are happening](#)

# Open, Public, Electronic, and Necessary (OPEN) Government Data Act [H.R.4174, title II](#)

- › Make Data Open by Default
- › Data Inventory and Federal Data Catalogue
  - Data.gov
- › Machine readable mandate
- › Chief Data Officer
- › Chief Data Officer Council
- › Title 1 Federal Evidence Building Activities
  - Power to OMB
- › Title III Confidential Information Protection and Statistical Efficiency
  - Reauthorizing Confidential Information Protection and Statistical Efficiency Act (CIPSEA)
  - Creates a common portal for restricted data

# Why Can't I find it? It's not what I want.

## PROBLEMS THAT MIGHT COME WITH GOVERNMENT DATA

- › Beggars can't be choosers
  - Too old
  - Not to the geographic level needed
  - Too detailed
  - Have to file a [FOIA](#) request
  - Machine readable only
- › Compatibility
  - Standardization
  - Combining two datasets even from same source **might not be possible**
  - Combining two different sources must **look at methodology**

## WHY ISN'T DATA AVAILABLE?

- Classified
- Personal Information (PII)
- Not sent to GPO or data.gov
- Not widely distributed/web
- Stuck in closed system
- Perceived lack of supposed interest [70% fewer crime statistics Tables](#)
- Not kept
- **Lack of funding**

# Funding for Federal Data Collection

	FY15	FY16	FY17	FY18		FY19	
				Final	Change from FY17	Request	Change from FY18
<b>Research Agency (amounts in millions of dollars)</b>							
NIH*	30311	32311	34229	37084	8.3%	34767	-6.2%
NSF	7344	7463	7472	7767	4.0%	7472	-3.8%
AHRQ	364	334	324	334	3.1%	256	-23.4%
FDA	2597	2730	2771	2812	1.5%	3524	25.3%
<b>Statistical Agency (amounts in millions of dollars)</b>							
BEA	96.3	105.1	103.8	99.0	-4.9%	98.0	-1.0%
BJS	41.0	41.0	45.5	48.0	5.5%	41.0	-14.6%
BLS	592.2	609.0	609.0	612.0	0.5%	609.0	-0.5%
BTS	26.0	26.0	26.0	26.0	0.0%	26.0	0.0%
Census	1088.0	1370.0	1457.0	2814.0	93.1%	3797.0	34.9%
EIA	117.0	122.0	122.0	125.0	2.5%	115.0	-8.0%
ERS	85.4	85.4	86.8	86.8	0.0%	45.0	-48.1%
NASS	172.4	168.4	171.2	191.7	12.0%	165.0	-13.9%
NCES	232.1	261.0	258.5	258.5	0.0%	261.5	1.2%
NCHS	155.4	160.4	160.4	160.4	0.0%	155.0	-3.4%
NCSES	58.3	58.3	59.7			59.8	
ORES	29.0	25.9	24.0	31.0	29.2%	31.0	0.0%
SOI	36.8	37.8	34.3	33.6	-2.0%	35.2	4.8%

Image from AmStat (with permission)

NIH- National Institutes of Health (HHS)

NSF- National Science Foundation

AHRQ- Agency for Healthcare Research & Quality (HHS)

FDA- Food & Drug Administration (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)

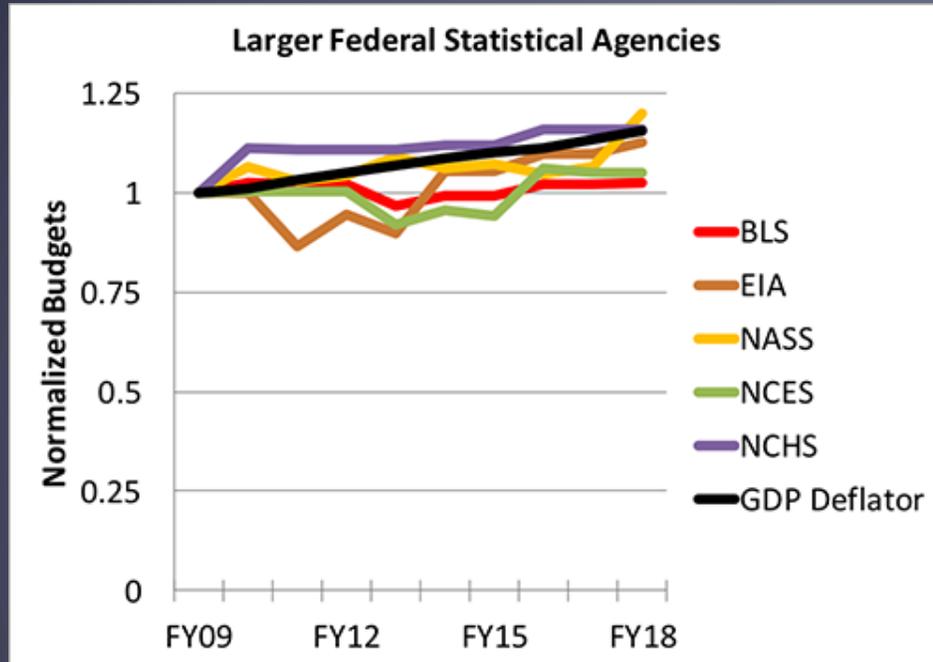
NCHS- Nat. Center for Health Stat. (HHS)

NCSES- Nat. Center for Science and Engineering Sat. (NSF)

ORES- Off. of Research, Evaluation, and Statistics (SSA)

SOI- Statistics of Income (IRS)

# Future of the Bureau of Labor Statistics



**In danger:** Nat. Longitudinal Sur., JOLTS, Am. Time Use Sur., Employee Benefits Sur., Cen. of Fatal Occupational Injuries, [Evaluation \\$27M > \\$2M](#)

## 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](#)

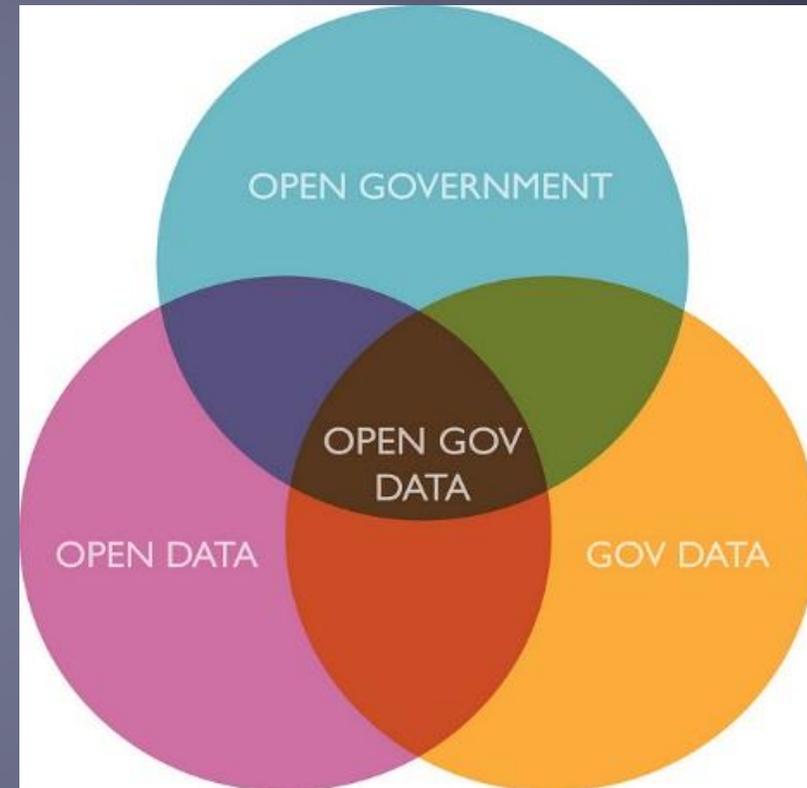
[Center for Data Innovation](#)

# Major Sources of Social Science Data in the US Government

- › Census
- › Labor
- › Economic
- › Federal Budget Projections
- › Taxes
- › Spending
- › Social Security
- › Justice
- › Banking
- › Trade
- › Transportation
- › Telecommunications
- › Education
- › Libraries and Museums
- › Economic Analysis

# Major Sources of Natural Science Data from the US Government

- › Agriculture
- › Health
- › Energy
- › Environment
- › Land Management
- › Science & Engineering
- › Weather
- › Defense



<https://www.flickr.com/photos/notbrucelee/6897137283/in/photostream>

- Visible, accessible, understandable, linking, and trustworthy (VAULT)

# Change is Coming

## **Grant Reporting Efficiency and Agreements Transparency (GREAT ACT) H.R. 150** Strengthens the DATA Act

- (1) establish data standards for information reported by grant recipients,
- (2) issue guidance directing federal agencies to apply those standards,
- (3) require the publication of recipient-reported data collected from all agencies on a single public website.

# Change is Coming

## Roles for FDLP Librarians

Become a Data Scientist

Become a Public Domain expert

Help folks with FOIA

Watch OMB or track an Agency

Comment in Regulations.gov

Educate the public on spending through community groups

# One Statistical Office in US: Why Not?

## No to National Secure Data Service

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 among departments laws and regulations 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.

# Open Government

## US FEDERAL

- › [DATA.GOV](#)
- › [Performance.gov](#)
- › [Center for Effective Government](#)
- › [Government Accountability Project](#)
- › [Bipartisan Policy Center's Evidence-Based Policymaking Initiative](#)
- › [Socrata](#)

## INTERNATIONAL

- › [Center for Data Innovation](#)
- › [Data Coalition](#)
- › [Open Government Partnership](#)
- › [Open Government Foundation](#)
- › [OpenKnowledge](#)
- › [SunlightLabs](#)
- › [Standard Business Reporting \(SBR\)](#)
  - › In Australia and Netherlands

# Associations: Blogs and Conferences

## FOR 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](#)

## FOR FEDERAL DATA POLICY

- › [Association of Public Data Users \(APDU\)](#)
- › [Council of Professional Associations on Federal Statistics \(COPAFS\)](#)
- › [American Statistical Association](#)
- › [Project Open Data](#)

# 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](#)

› [What is an API](#)

› [DataRefuge](#)

[Public Knowledge: Access and Benefits](#) (Information Today, 2016)  
[Innovation in Federal Statistics](#) (National Academics, 2017)

# Let's discuss

[boettcher@georgetown.edu](mailto:boettcher@georgetown.edu)

202 687-7495

@jenny.wombat

If you like this webinar go to the [Center for Data Innovation's discussion about the future of open data in the United States](#)



© [Bill Waterhouse](#), with permission

AMSTAT images from  
<http://magazine.amstat.org/blog/2018/05/01/fy18fedbudget>

All Slides 

# Sunlight Foundation Guidelines for Open Data Policies

## What Data Should Be Public

Proactively release government information online

Reference and build on existing public accountability and access policies

Build on the values, goals and mission of the community and government

Create a public, comprehensive list of all information holdings

Specify methods of determining the prioritization of data release

Stipulate that provisions apply to contractors or quasi-governmental agencies

Appropriately safeguard sensitive information

## How to Implement Policy

Create or appoint oversight authority

Create guidance or other binding regulations for implementation

Incorporate public perspectives into policy implementation

Set appropriately ambitious timelines for implementation

Create processes to ensure data quality

Ensure sufficient funding for implementation

Create or explore potential partnerships

Mandate future review for potential changes to this policy

## How to Make Data Public

Mandate data formats for maximal technical access

Provide comprehensive and appropriate formats for varied uses

Remove restrictions for accessing information

Mandate data be explicitly license-free

Charge data-creating agencies with recommending an appropriate citation form

Require publishing metadata

Require publishing data creation processes

Mandate the use of unique identifiers

Require code sharing or publishing open source

Require digitization and distribution of archival materials

Create a central location devoted to data publication and policies

Publish bulk data

Create public APIs for accessing information

Optimize methods of data collection

Mandate ongoing data publication and updates

Create permanent, lasting access to data

# States and Cities

- › Public Domain in [California](#) and [New York City](#)
- › [https://www.google.com/search?source=hp&q=site%3A.gov+"open+data"+\[YOUR STATE HERE\]](https://www.google.com/search?source=hp&q=site%3A.gov+%27open+data%27+[YOUR STATE HERE])
- › Remember to [search](#) using site:.gov
- › There are a lot of consultants making money on cool visualizations.



<https://data.sonomacounty.ca.gov/dataset/SoCo-Data-PNG/3m9t-bc35>

# Major International Data Sources

## BY TOPIC

<http://data.un.org>

[Social & Economic- World Bank](#)

[Financial & Economic- International  
Monetary Fund](#)

[Human body- World Health Org](#)

[Labor- International Labour Org](#)

[Telecommunications- International  
Telecommunications Union](#)

[Governance- Transparency  
International](#)

[Developed Countries- Organization  
for Economic Co-operation and  
Development \(OECD\)](#)

## BY COUNTRY

### [National Statistical Offices](#)

More data available in national language

Some charge for access

Citizens of that country might have free access

### [National Repositories/Archives](#)

Historical

Datasets

# Legal issues

## Data and IP

- › <https://www.lib.umn.edu/data-management/copyright>
- › <https://data.research.cornell.edu/content/intellectual-property>
- › [https://en.wikipedia.org/wiki/Copyright\\_status\\_of\\_work\\_by\\_U.S.\\_subnational\\_governments](https://en.wikipedia.org/wiki/Copyright_status_of_work_by_U.S._subnational_governments)

## Licensing Data

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

# Offer API: Freedom from interfaces

- › **Application Programming Interface as opposed to Bulk Data**
  - Read-only machine to machine querying
- › Accessing, Searching, Finding, and Transferring data
- › Steps to happiness
  - 1. Find the webpage where the API lives
  - 2. Read the documentation to make sure it has the data
  - 3. Might have to register to query, to get a key or token
  - 4. “On top of good data, a well-designed data API provides granular access, deep filtering, typed values, normalized tables, RESTful interfaces, multiple output formats, useful validation messages, use-case or intent-oriented URLs, documentation, client libraries, versioning, fast results, high uptime, easy on-boarding, interactive documentation, and a developer community hub.”
    - › <https://opengovdata.io/2014/bulk-data-an-api>
- › [example](#)

# Numeric Data: who is responsible

- › Datum (single of data)

- Researchers

- › Data
- › Creation of numbers

- › Statistics

- Publishers

- › Information
- › Relations among numbers

- › Analytics

- Analysts

- › Knowledge
- › Context of numbers

- › Tacit Knowledge

- Expert

- › Wisdom
- › “I Just know it”

- › Curation

- Repository

- › Preserving access
- › Set up API or Bulk Access

# P.E.S.T. Analysis for Industry

- Political
  - Legislative
    - [Congress.gov](https://www.congress.gov)
  - Executive
    - [NARA's Federal Register](https://www.nara.gov)
    - [Regulations.com](https://www.regulations.gov)
  - Judicial
    - [United States Courts](https://www.uscourts.gov)
- Socio-cultural
  - Norms & Ratios
    - [IRS's Statistics of Income](https://www.irs.gov)
  - Peers and partners
    - [Economic Census](https://www.census.gov)
    - [Annual Surveys from Census](https://www.census.gov)
- Economic
  - Sector Inflation
    - [BLS's Producers Price Index](https://www.bls.gov)
  - Microeconomic trends
    - [FRB's FRED](https://fred.stlouisfed.org)
    - [Census BEA's Interactive Tables](https://www.census.gov)
- Technology
  - Patents
    - [Citation Analysis](https://www.uspto.gov)
  - Tech Transfer
    - [Government funded research to license](https://www.stta.gov)