



IP and Government Data: Understanding Open Access vs. Public Domain

BEYOND THE NUMBERS: The Economic Data Ecosystem

Federal Reserve Bank of St. Louis

November 9, 2018

Presenter: Jennifer C. Boettcher, Georgetown University

Slides available on the BTN website

Jennifer C. Boettcher

Jennifer C. Boettcher and Leonard M. Gains. Industry Research Using the Economic Census. Greenwood Press: Phoenix, AZ. 2004



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

Georgetown Univ 1997-present

Catholic Univ of America, Adjunct Faculty 03-07

Texas A&M Univ 94-97

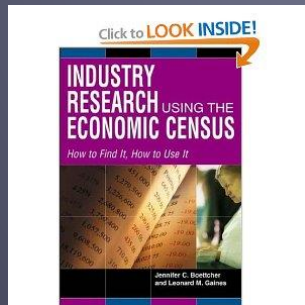
ALA RUSA BRASS Member since 1991

SLA Member since 1992

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

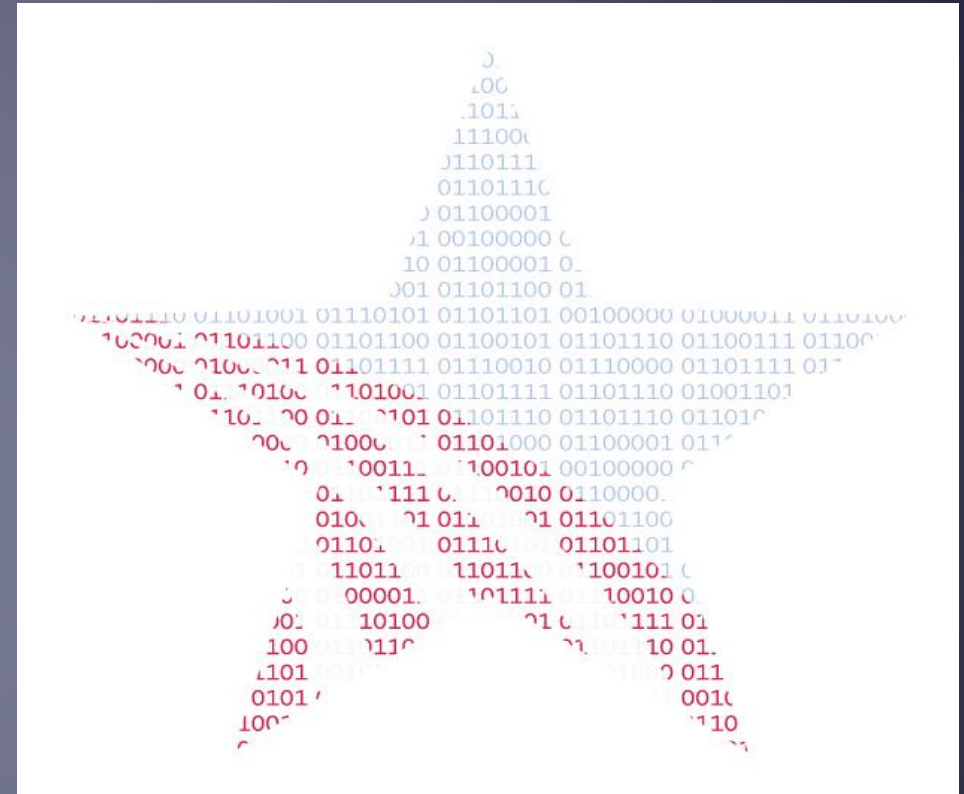
2013 Emerald Research Grant: Zombie List (reanimated business sources)

2010 Gale Cengage Learning Award for Excellence in Business Librarianship



What will be covered?

- › Who I am and what I do
- › Data in context
- › What is Open Data?
- › What Is IP?
 - › Data as IP
- › What is Public Domain?
- › Who Creates Data?
- › Who Creates Federal Data Policy?
- › Where to get involved and learn more?



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

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

Boettcher, J. C., & Dames, K. M. (2018). [Government data as intellectual property: Is public domain the same as open access?](#) *Online Searcher*, 42(4), 42-48.

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



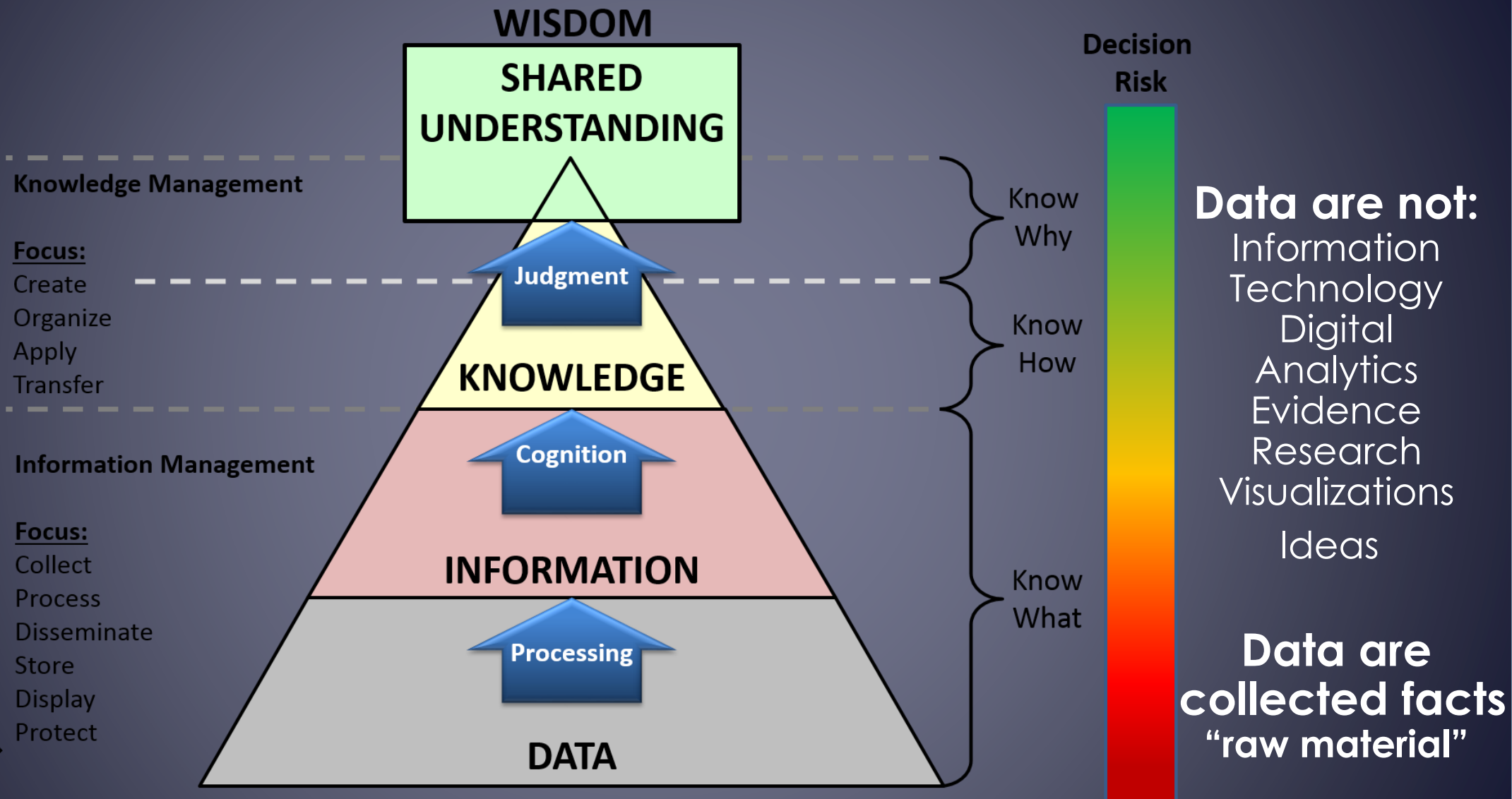
Should data be open?

- › Free
- › Transparent
- › Accountable
 - › Accessible to citizens
 - › Engages all citizens
 - › Machine Readable
 - › When in doubt openness prevails

› **Why isn't it already?**

- › Not sent to Government Publishing Office (GPO)
- › Classified
- › Not widely distributed/web
- › Lack of supposed interest
- › Lack of funding
- › Not kept

Knowledge Management Cognitive Pyramid



Adaptations of DIKW pyramid by US Army Knowledge Managers,
from https://en.wikipedia.org/wiki/DIKW_pyramid

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



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>



Why Open Data Exists

- › 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](#)

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



Intellectual Property (IP):

Legal term for works that can be protected from infringement.
Works can be bought, sold, assigned, or licensed

Patents: Legal protection for a new invention: an application of a new idea, discovery, or concept that is useful.

Trademarks: provide rights to use symbols, particular words, logos, or other markings that indicate the source of a product or service.


Trade Secrets: Rather than disclose the idea, simply keep it secret. Trade secrets have the advantage that they never expire, but special measures are required to ensure continued secrecy, and should one be violated, there is little legal protection for the owner .

Copyright: Legal protection from copying any creative work (e.g., works of art, literature [fiction or nonfiction], music, lyrics, photographs), as well as compilations of information.

No registration is required.

Copyright does not protect facts, ideas, systems, or methods of operation, although it may protect the way these things are expressed

Circular 1



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)



Creative Expression

Under the Copyright Act, a **compilation** is defined as a "collection and assembling of preexisting materials or of data that are selected in such a way that **the resulting work as a whole constitutes an original work of authorship.**" [17.](#)

[U.S.C. 101](#)

The collection receives protection so long as the collection is both **original and fixed**, and "**extends only to the material contributed by the author of such work**, as distinguished from the preexisting material employed in the work, and does not imply any exclusive right in the preexisting material." [17 U.S.C. 103](#)



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

Works Produced for the U.S. Government: Lifecycle of Data

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.

Other **Data Scientists** 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 will act as **Publisher**, putting the synthesized resources on its website for all: primarily for **decision makers** but also for **citizens**, to read.

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

› REMEMBER: Public domain data must be attributed.



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 Archives](#)

File here [FOIAonline](#)

Help with [FOIA.gov](#)

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

OPEN Government Act of 2007

Citizen Journalist

20 day response

Watchdogs

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

[Reporters Committee](#) for the
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*, Massachusetts*, and Minnesota*,

*Check with issuing agency

› 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

- Free data flow
- Law
- Better than Open Access
- Government products
- Data at any stage can be retrieved by FOIA
- Not for some sub-nationals

> Open Access

- Free data flow
- Because of ownership of copyright
- Principles and license
- Reuse and redistribution of the data
- Allows derivative works as Open only
- No restrictions on who can access and use
- 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](#)

Funding for Federal Data Collection

				FY18		FY19	
	FY15	FY16	FY17	Final	Change from FY17	Request	Change from FY18
Research Agency (amounts in millions of dollars)							
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%
Statistical Agency (amounts in millions of dollars)							
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 (permission pending)

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 Stat. (NSF)

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

SOI- Statistics of Income (IRS)

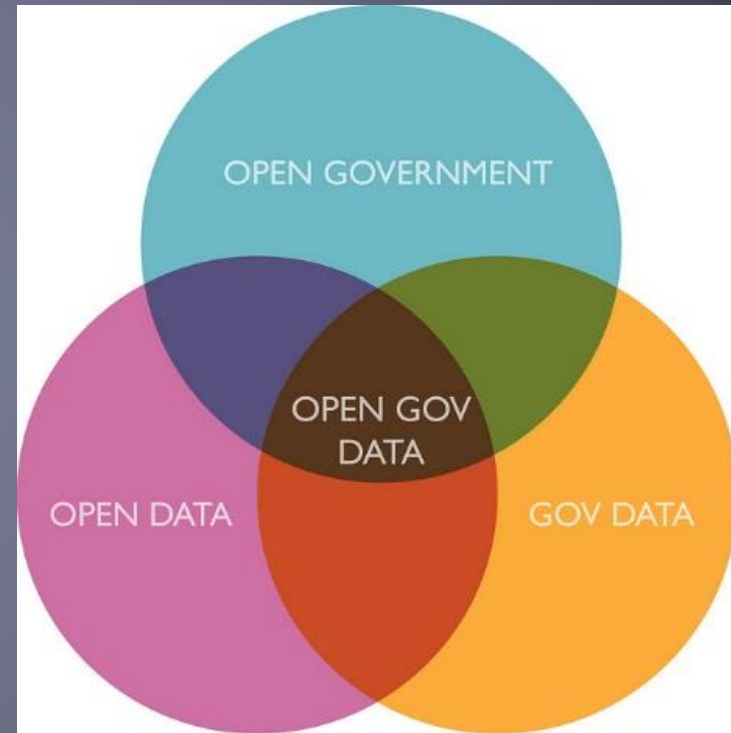


Major Sources of Social Science Data in the US Government

- › 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>



Data Policy in the Federal Government

- › Federal government policy
 - Passed by Congress
 - Implemented by Executive Branch
 - Refined by Courts
- › Policy on data collection (priorities)
 - Mandated by law (in CFR) HR 2810 Sec 6012
 - 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
 - 1st 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
- › Previous Administrations
 - Creation of Data.gov (2009), from G.W. Bush Administration
 - OMB Revised Circular A-130: Managing Information as a Strategic Resource
 - Memorandum for the Heads of Executive Departments and Agencies M-13-13
 - “enables the data to be fully discoverable and usable by end users”



Role of OMB

- Produce and disseminate relevant and timely information
- Conduct credible and accurate statistical activities
- Conduct objective statistical activities
- Protect the trust of information providers

strategy.data.gov

- 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



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
- › Compatibility
 - Standardization
 - Combining two datasets even from same source **might not be possible**
 - Combining two different sources must **look at methodology**



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.



Open Government

US FEDERAL

- › [DATA.GOV](#)
- › [Performance.gov](#)
- › [Center for Effective Government](#)
(formerly OMB Watch)
- › [Government Accountability Project](#)
- › [Open Government Foundation](#)

INTERNATIONAL

- › [Center for Data Innovation](#)
- › [Data Transparency Coalition](#)
- › [Open Government Partnership](#)
- › [Open Knowledge](#)
- › [Sunlight Labs](#)

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+\)
- › 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- Organisation
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



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](#)

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

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)
[Innovation in Federal Statistics](#) (National Academics, 2017)

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

Licensing Data

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

Let's discuss

boettcher@georgetown.edu

202 687-7495

@jenny.wombat

[PLEASE read](#) and comment on
[Federal Register Notice](#),

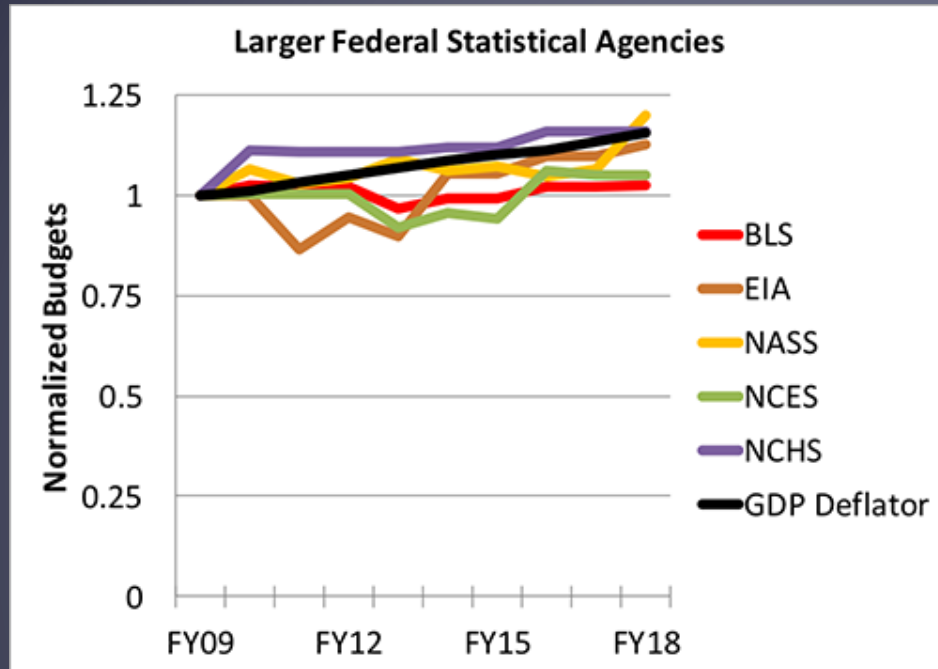
by Nov 16



© [Bill Waterhouse](#), with permission

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

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