

Federal Data Field Guide

federaldatafieldguide.us

April 2026 | V.1

*A product of the Executive Fellowship in Applied Technology
Policy, a joint program of UC Berkeley's School of Information
and Goldman School of Public Policy*

UC Berkeley





Datasets don't always fit neatly into a single category, which is why the semiaquatic, egg-laying mammal called the duck-billed platypus is this Field Guide's mascot.

Why a Field Guide for Federal Data?

The purpose of this guide is to provide a more complete context for federal data users and stakeholders that will inspire them to consider a broader range of data types in their research and advocacy; we also hope it will also inform national dialogues about the future of federal data.

What's included?

The Guide is organized into eight primary categories of federal data (described on the right), each representing distinct collection methods, policy frameworks, and use cases.

This field guide focuses primarily on publicly available datasets created, maintained, and published by executive branch agencies of the federal government. This Guide does not include sensitive or classified datasets, or derivative works such as reports or interactive web tools that use data.

Introduction



Statistical Data that measure population-level characteristics

Administrative Data generated through routine government operations



Geospatial Data capturing spatial and environmental information

Scientific Data advancing knowledge across disciplines



Accountability Data for primary purpose of transparency into government activities

Evaluation Data assessing program effectiveness



Navigation Data helping citizens access government services

Reference Data providing standardization across systems



Legal & Governance Landscape

Acronyms & Initialisms

Credits

Introduction

The U.S. federal government operates one of the world's largest and most complex data ecosystems; collecting, protecting, analyzing, and disseminating information that touches virtually every aspect of American life. This vast infrastructure encompasses everything from real-time satellite imagery tracking hurricanes to century-old immigration records preserved in the National Archives, from cutting-edge genomic research datasets at the National Institutes of Health to routine administrative records documenting government operations.

The federal government maintains hundreds of thousands of distinct datasets across hundreds of agencies, with some individual datasets containing billions of records spanning decades of continuous collection. The Census Bureau alone manages more than 130 different surveys, and NASA archives more than 32 petabytes of Earth observation data.



Why Do Federal Data Matter?

Federal data serve as essential infrastructure for democracy, economic activity, and scientific progress. For individual citizens, federal data can help cancer patients find clinical trials, provide vital safety information through weather forecasts, and offer transparency into where taxpayer dollars go. Parents rely on school performance data to make decisions about where to live. Roads are safer because autonomous vehicle companies rely on shared reference data about location and elevation. And veterans live longer because of research fueled by federal data.

Businesses use economic statistics to anticipate market conditions, while nonprofits leverage demographic and health data to identify community needs and measure program impact. State and local governments rely on federal data for everything from emergency preparedness to transportation planning. Academic institutions and think tanks use federal data to conduct studies that inform public policy and advance scientific knowledge.

Across the nation, people and organizations rely on the government's unique ability to collect comprehensive, nationally representative data that would be impossible for private entities to gather at scale.

Statistical Data

Statistical data measure **group-** and **population-level characteristics** of **people, businesses,** and the **economy**, and are often used to inform **planning, policymaking,** and **market research**, and to **allocate resources** and better **understand community needs**.

Coordinated by OMB's **U.S. Chief Statistician**, the **federal statistical system** includes:

- 16 OMB-recognized federal statistical agencies and units, 13 of which are considered principal statistical agencies (such as the Bureau of Economic Analysis, U.S. Census Bureau, and National Center for Health Statistics), whose primary purpose is the production of statistics;
- 24 statistical officials in agencies recognized by the CFO Act of 1990;
- 2 interagency advisory committees; and
- 100+ smaller programs that also collect, analyze, and publish statistics (such as the National Institute of Justice).



Statistical Data

Modes of Collection

- Primarily collected through surveys and interviews, which may be conducted in person, by telephone, through mail questionnaires, or with online instruments
- Some data are collected from third parties such as local law enforcement agencies, school districts, and public health entities
- Collections are often bolstered by [administrative data](#) (such as tax filings or program participation data) and private-sector data



Legal and Policy Context

A constellation of protections and governing authorities surround *statistical data*, including [Title 13](#) for the Census Bureau, [FERPA](#) for some educational data, and [HIPAA](#) for health data. OMB regulations and policies further establish guidelines and procedures to maintain the quality, timeliness, and relevance of statistical data.

Data collected under a pledge of confidentiality and used for statistical purposes only are protected by [CIPSEA](#). CIPSEA ensures that these data are used only for evidence-building and never for enforcement purposes.

Other relevant statutes include the [Privacy Act](#) which covers individuals' personal information, and the [Paperwork Reduction Act](#) which governs how the government collects data from the public.

Due to the confidential nature of many data assets in the federal statistical system, agencies often produce both restricted access and open access files.

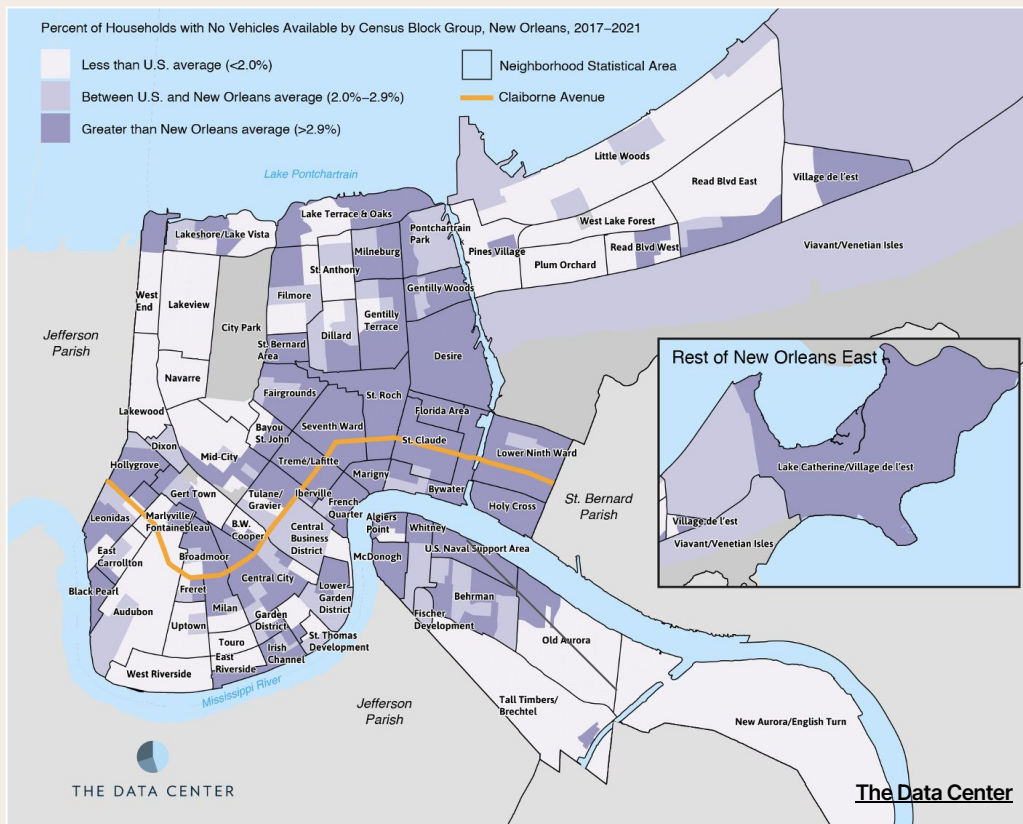
American Community Survey (ACS)

[LINK TO DATASET ↗](#)

U.S. Census Bureau, Department of Commerce

Local emergency planners use American Community Survey data on households without access to a vehicle to determine how many evacuation buses are required and the ideal locations for pickup.

This ensures there will be a seat on the bus for everyone who needs one when a hurricane approaches.



Examples of Statistical Data

FBI: [National Incident-Based Reporting System \(NIBRS\)](#) ↗

Annual data on crime collected from the nation's 18,000 law enforcement agencies

U.S. Census Bureau: [Decennial Census](#) ↗

Constitutionally-required count every ten years of all people living in the U.S. that determines how many seats in Congress each state gets, and informs federal funding and redistricting

Federal Reserve Board: [Survey of Consumer Finances \(SCF\)](#) ↗

Triennial survey of ~6,500 families on balance sheet, pension, income, use of banks, plus demographic characteristics

Federal Reserve Board: [Survey of Household Economics and Decisionmaking \(SHED\)](#) ↗

Annual survey of ~12,000 households on financial wellbeing, including savings, credit access, employment, and gig work

BJS: [National Crime Victimization Survey \(NCVS\)](#) ↗

Household survey on frequency, characteristics, and consequences of crimes reported and not reported to police

BJS: [Law Enforcement Management & Administrative Statistics \(LEMAS\)](#) ↗

Sample survey of law enforcement agencies every three to four years that covers staffing, policies, and budgets with supplements on timely topics

ED: [National Assessment of Educational Progress \(NAEP\)](#) ↗

The "Nation's Report Card," a congressionally-mandated assessment of performance across states and student groups

NSF: [National Survey of College Graduates \(NSCG\)](#) ↗

Biennial survey of education, career paths, occupation, salary, and demographics of science & engineering college graduates

BLS: [Consumer Price Index \(CPI\)](#) ↗

Monthly key economic indicator of prices urban consumers pay for a fixed basket of goods and services that is used as a measure of inflation

CDC: [Behavioral Risk Factor Surveillance System \(BRFSS\)](#) ↗

Phone survey of 400k+ adults that yields state-level data on behaviors linked to chronic/infectious diseases and injuries

Administrative Data

Administrative data mostly come from the **forms** Americans fill out when they engage with the federal government. Federal agencies use administrative data to run the business of government, such as **determining eligibility** for federal programs, **delivering services** to the public, **implementing policies**, and **enforcing laws** or **regulations**. Public uses for these data include **advocacy**, **accountability**, and **research**.

Depending on privacy and use restrictions, elements of administrative datasets are sometimes shared across programs to reduce paperwork burden and improve service delivery, or repurposed for secondary uses like **improving statistical products**, **managing program performance**, **detecting fraud**, or **conducting evaluations**.



Administrative Data

Modes of Collection

- Generally collected as part of routine program operations and service delivery, such as when Americans apply for benefits or file taxes
- Can also be collected as part of regulatory requirements, such as those associated with loan processing or environmental permitting
- Often drawn directly from paper or online forms, electronic records, or other internal databases
- In some cases, collected by third parties such as lending institutions or state governments



Legal and Policy Context

Elements in administrative data are sometimes statutorily required or part of regulations, but typically are driven by information needed for agencies to conduct business. OMB provides cross-agency guidance on how administrative data should be governed and managed.

In order to maintain public trust, agencies must maintain strong privacy protections for administrative data.

Administrative data that contain personally identifiable information primarily fall under [Privacy Act](#) protections, requiring SORNs, and limiting disclosure to authorized "routine uses."

These data are also subject to [FOIA](#) requests, though personal information is typically redacted. When published proactively, administrative data are typically aggregated to protect individual privacy, and must meet [OPEN Government Data Act](#) requirements.

Some administrative data are collected to comply with regulations, such as HMDA (for mortgages) and NEPA (for permitting). Sector-specific policies also apply, such as [HIPAA](#) for health and FERPA for education.

Form 990 data

[LINK TO DATASET ↗](#)

Internal Revenue Service (IRS)

Donors use IRS Form 990 data to research charities they are considering supporting. All nonprofit organizations in the U.S. must file an annual 990 form with the IRS. This administrative dataset's primary purpose is for the IRS to ensure an organization's continued eligibility for tax-exempt status and its compliance with tax laws.

The Taxpayer First Act of 2019 requires that the IRS release Form 990 data for free to the public in a machine-readable format. Opening access to this administrative data enabled the creation of third-party tools that make it easy for would-be donors to use the 990 data to vet charitable organizations.

These tools enable donors to look at trends in revenues and expenses over time, what percent of their budget goes toward fundraising expenses, whether the organization has diversified financial support, and if executive compensation is in line with the organization's size and mission.

Form **990-EZ** **Short Form** **Return of Organization Exempt From Income Tax** OMB No. 1545-0047
2024
 Under section 501(c), 527, or 4947(a)(1) of the Internal Revenue Code (except private foundations)
 Do not enter social security numbers on this form, as it may be made public.
 Go to www.irs.gov/Form990EZ for instructions and the latest information. **Open to Public Inspection**

Department of the Treasury
Internal Revenue Service

A For the **2024** calendar year, or tax year beginning **TaxPeriodBeginDt**, 2024, and ending **TaxPeriodEndDt**, 20

B Check if applicable:
 Address changed **AddressChangeInd**
 Name change **NameChangeInd**
 Initial return **InitialReturnInd**
 Final return **FinalReturnInd**
 Amended **AmendedReturnInd**
 Application pending **ApplicationPendingInd**

C Name of organization **BusinessName**
 Number and street (or P.O. box if mail is not delivered to street address) **InCareOfNm**
 Room/suite **RoomSuite**
USAddress **ForeignAddress**
 City or town, state or province, country **CityStateProvCountry**

D Employer identification number **EIN**
E Telephone number **PhoneNum** **ForeignPhoneNum**
F Group Exemption Number **GroupExemptionNum**

G Accounting Method **MethodOfAccountingCashInd** **MethodOfAccountingAccrualInd** (If **Y**): **MethodOfAccountingOtherDesc**

H Check if **ScheduleBNotRequiredInd** required to attach Schedule B Form 990.

I Website: **WebsiteAddressTxt**

J Tax-exempt status (check only one) Organization 501(c)3 Organization 501(c)29 Organization 501(c)28 Organization 501(c)27 Other: **TypeOfOrganizationOtherDesc**

K Form of organization: TypeOfOrganization Control TypeOfOrganization Trust TypeOfOrganization Assn Other: **TypeOfOrganizationOtherDesc**

L Add lines 5b, 6c, and 7b to line 9 to determine if gross receipts are \$200,000 or more, or if total assets (Part II, column (B)) are \$500,000 or more, file Form 990 instead of Form 990-EZ **GrossReceiptsAmt**

Part I Revenue, Expenses, and Changes in Net Assets or Fund Balances (see the instructions for Part I)
 Check if the organization used Schedule O to respond to any question in this Part I **InfoInScheduleOPartIInd**

1	Contributions, gifts, grants, and similar amounts received	1	Contributions/Gifts/Grants Etc Amt
2	Program service revenue including government fees and contracts	2	Program Service Revenue Amt



Photo by Holly Scheib

Examples of Administrative Data

USDA: [Characteristics of SNAP Households](#)

Annual data on demographics and financial status of households receiving SNAP food assistance, collected for quality control

NIH: [RePORTER \(Research Portfolio Online Reporting Tools\)](#)

Repository of all NIH-funded intra/extramural research (abstracts, budgets, PIs) + resulting publications and patents

DHS/USCIS: [H-1B Employer Data Hub](#)

H-1B visa petitions filed by employers, including approval/denial rates, geographic location, and industry code

CFPB: [Home Mortgage Disclosure Act \(HMDA\) Data](#)

Loan-level data on mortgage applications, originations, and purchases, including borrower demographics

DHS/FEMA: [Housing Assistance Program Data - Owners](#)

Detailed registration and eligibility data for homeowners seeking disaster financial assistance for needs and housing

VA: [Geographic Distribution of VA Expenditure \(GDX\)](#)

Annual report detailing VA spending (healthcare, compensation, construction) by state, county, and congressional district

DOD: [Defense Casualty Analysis System \(DCAS\)](#)

Official military records and demographics of military personnel deceased, wounded, ill, or injured from global conflicts

USDA: [Single Family Section 502 Direct Active Loans](#)

Aggregated data on direct home loans (including amount, leverage, and housing type) to low-income rural residents

ED: [FAFSA Completion by High School & Public School District](#)

Weekly tracking of submitted/completed FAFSA forms for high school seniors (aged 19 or younger) by school name and district

HHS/CDC: [National Vital Statistics System \(NVSS\) Birth Data](#)

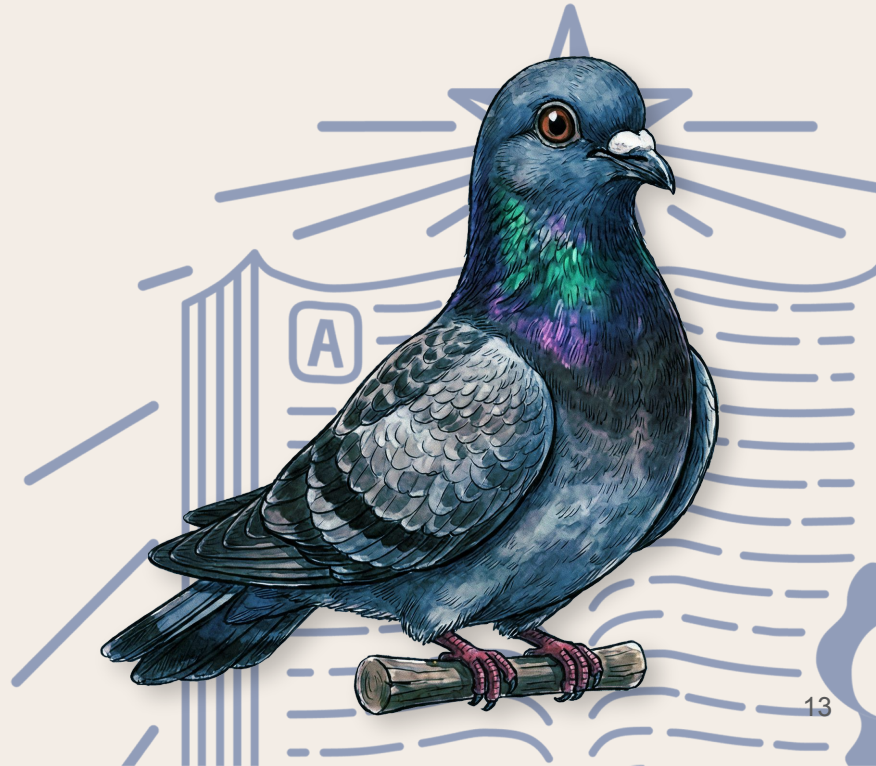
Annual births in U.S. and territories including demographic, health, and medical information collected from birth certificates

Geospatial Data

The defining characteristic of geospatial data is that they contain **specific location attributes** through coordinates like latitude/longitude, addresses, or other spatial references. Working with geospatial data often requires specialized GIS skills and software, as well as storage solutions for managing large volumes of data.

Geospatial data are useful for understanding, analyzing, and visualizing spatial or physical characteristics of places and natural systems. Uses include **environmental monitoring, natural resource management, transportation and urban planning, and disaster response**, as well as understanding differences between geographic areas, such as **urban** and **rural** places.

The **Federal Geographic Data Committee**, housed in the Department of the Interior, provides direction and oversight for federal geospatial data, establishing common frameworks and standards to enable integrating geospatial data across agencies and levels of government.



Geospatial Data

Modes of Collection

Data are primarily collected using:

- satellite and aerial imagery,
- sensors collecting data from the air, ground, and bodies of water,
- GPS-enabled or address-based surveys,
- data-sharing agreements with state and local partners, and
- commercial data products.

Geospatial datasets that change infrequently often serve as reference data, whereas the primary value in other geospatial data (such as weather from satellites) is that they are updated in near real-time.



Legal and Policy Context

- Federal geospatial data are subject to the [Geospatial Data Act](#) (GDA) requirements for standards, sharing, and coordination through the National Spatial Data Infrastructure (NSDI) framework.
- When geospatial datasets contain personally identifiable information (like individual GPS tracking), [Privacy Act](#) protections apply.
- Non-sensitive geospatial data must be made publicly accessible through platforms like GeoPlatform under GDA and [Evidence Act](#) requirements.
- Some geospatial data may have restricted access due to national security or public safety concerns, such as detailed infrastructure maps, sensitive facility locations, or locations of endangered species.

North American Bat Monitoring (NABat) Database

[LINK TO DATASET ↗](#)

US Geological Survey

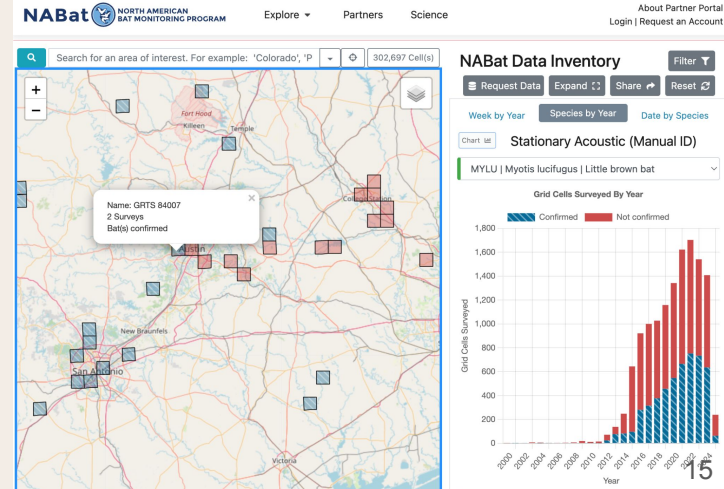
Bats are important, in part, because they provide America's farmers with billions of dollars of free services by eating insects that harm crops. If we want to protect the health of bats and the free services they provide, we need to protect the bats, and in order to protect the bats, we need to know where the bats are located.

That is where the North American Bat Monitoring Database comes in, providing a single source of truth in a standardized format about the location and characteristics of bat populations. The NABat dataset depends on local surveys of bat populations from local governments, nonprofits, research institutions, and other experts.

This dataset also makes the permitting process easier for projects that might impact bat populations, like wind farms, bridges, and underground mining operations. Developers can reference NABat data so they can mitigate harm to bat populations, instead of having to collect data on local bats themselves.



USGS
science for a changing world



Examples of Geospatial Data

USDA/NASS: [Cropland Data Layer](#) ↗

Annual, crop-specific land cover data since 1997 for the continental U.S., used for agricultural statistics, crop insurance, and resource management

NOAA: [Next Generation Weather Radar \(NEXRAD\)](#) ↗

Real-time and archived radar reflectivity and velocity data used for tracking weather, preventing airline bird strikes, etc.

DOT: [National Bridge Inventory \(NBI\)](#) ↗

Locations, structural condition ratings, and inspection dates for all publicly-owned bridges, used for infrastructure plans

DOE/EIA: [Power Plants in the U.S.](#) ↗

Locations of electric power plants, including capacity, energy source, and operational status

FAA: [Military Training Routes \(MTRs\)](#) ↗

Defined geographic corridors in the U.S. used for low-altitude, high-speed military flight training

USGS: [National Water Dashboard](#) ↗

Real-time and historical data on streamflow, groundwater, water quality, and precipitation from 13,500+ stations

EPA: [Air Quality System \(AQS\)](#) ↗

Repository of ambient air pollution and meteorological data from 4,000+ monitors nationwide

NASA/USGS: [Landsat](#) ↗

Continuous, high-res satellite imagery collected since 1972 enabling detection of land-use changes, agricultural trends

HUD: [Public Housing Developments](#) ↗

Locations of public housing and other assisted housing properties to aid planning and analyzing access to services

HHS: [Geospatial Health Systems](#) ↗

Open and secure access data to location information on health treatment and preparedness

Scientific Data

Scientific data are used to generate, validate, and replicate findings that **advance knowledge** across disciplines including **agriculture, climate, environment, health, energy, and technology**. Researchers use these data to **test hypotheses, evaluate clinical trials**, build **evidence** for **regulations** and **enforcement**, and develop new **theoretical models**.

Federal agencies generate scientific data in two ways.

1. Federal employees or contractors, such as the physician-scientists at the NIH Clinical Research Center, conduct activities **inside federal agencies**.
2. Agencies fund universities, small businesses, and other institutions outside of government to conduct **external activities** that advance science, such as the USDA's Agriculture & Food Research Initiative.

To increase the utility of the many scientific datasets generated through federal activities, agencies often index and store scientific data together in **designated repositories** so they can be found and reused by others.



Scientific Data

Modes of Collection

- Vary widely depending on study objectives, design, and methodology as well as agency resources, and include surveys, sensors and other scientific instruments, biological sampling, and direct observations
- May involve long-term and ongoing streams of observations (e.g., NASA's James Webb Space Telescope), consist of single measurements in time-limited cases (e.g., NOAA and the U.S. Navy's 2019 Dolphin Sounds data), or be drawn from routine operational monitoring systems (e.g., water quality monitoring systems)
- Should meet high standards of quality to satisfy peer-review



Legal and Policy Context

Scientific data fall under a [mosaic of policies](#), regulations, and laws that shape their management, access, and preservation. Many of these policies apply standards for reproducibility, openness, and peer-review while protecting intellectual property, human subjects, and national security interests.

Depending on the features and methods of a scientific data collection, activities may be covered by [geospatial](#) or [statistical](#) governance frameworks, or a combination. For example, CDC's National Health Interview Survey is a statistical dataset that also informs research on topics like the use of alternative medicines to manage pain.

Broad federal policies and statutes that govern scientific research data include the [Privacy Act](#), Federal Policy for the Protection of Human Subjects, OSTP and agency public access policies for federally-funded research, OMB [Evidence Act](#) and [Information Quality Act implementation](#) guidance, and myriad policies specific to agencies, domains, and projects.

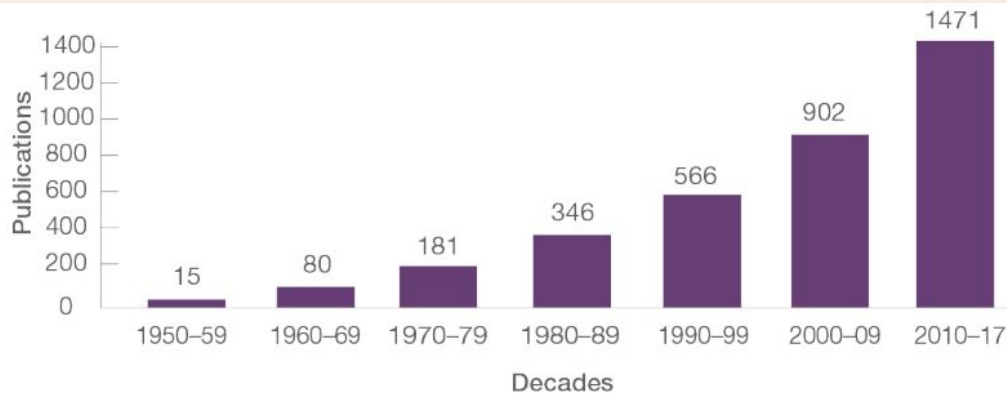
Framingham Heart Study (FHS)

[LINK TO DATASET ↗](#)

National Institutes of Health, Department of Health and Human Services

When the Framingham Heart Study began in 1948, heart disease was the leading cause of death in the U.S., and just a few years earlier had taken the life of President Franklin Delano Roosevelt.

Three generations and 15,000+ participants later, this scientific data gives patients and healthcare professionals insights into the major risk factors (like smoking, obesity, diabetes, and sedentary lifestyle), as well as factors such as genetics, blood lab results, and demographics.



Total articles published through November 2017 = 3,561

NIH



NIH History Office 19

Examples of Scientific Data

DOE: [Atmospheric Radiation Monitoring Dataset \(ARM\)](#)

30 years and 7 petabytes of continuous measurements of radiation, cloud microphysics, aerosol properties, and meteorological variables

NOAA/U.S. Navy: [Dolphin Sound Production](#)

Passive acoustic monitoring of dolphin whistles and clicks (2019–2021) from multiple U.S. National Marine Sanctuaries

NIH: [The database of Genotypes and Phenotypes \(dbGaP\)](#)

Controlled access to human genomic/phenotypic data from thousands of studies, linking genetic variation to health and disease outcomes

NIH: [All of Us](#)

Health information, biological samples, and electronic health records from 1M+ voluntary U.S. participants

EPA: [ScienceHub](#)

Research data supporting EPA's chemical, toxin, health and environmental studies and scientific publications.

USDA: [Dr. Duke's Phytochemical & Ethnobotanical Databases](#)

Details on chemical components of thousands of plant species, their reported medicinal uses, and biological activities

NSF: [National Ecological Observatory Network \(NEON\)](#)

Comprehensive ecological data in over 180 datasets collected from across the U.S.

NOAA/NCEI: [Local Climatological Data](#)

Hourly weather data from ~1,000 U.S. stations, including temperature, precipitation, wind, and snowfall since 2005

NASA: [Biological Data Repository](#)

500+ Genetic, molecular, and physical datasets from spaceflight experiments on microbes, plants, fruit flies, and rodents to advance human space exploration

USDA: [National Invasive Species Info Center Species Profiles](#)

Details on the biology, distribution, and management of invasive species U.S. to inform prevention and control

Accountability Data

Accountability data measure government functions such as **spending, policy implementation, and service delivery**. They are typically published on a schedule, either by individual agencies, or in government-wide collections:

- OMB often sets common data standards and reporting schedules through government-wide initiatives (e.g., the [DATA Act](#) that requires agencies to report spending data to [USAspending.gov](#)).
- Government-wide platforms managed by GSA, OMB, OPM, and Treasury collect and publish accountability data, whereas some government-wide accountability data are more domain-specific, such as DOE's Federal Building Energy and Water Use data.
- Agencies publish agency-level and programmatic data, such as Medicaid Enrollment Data (CMS) or SNAP Application Timeliness (USDA).

Accountability data are widely used by **Congress, watchdog groups, journalists, and agency leadership** to **monitor performance**. Federal officials use the data for **planning, improving processes, allocating staff, and deploying infrastructure**.



Accountability Data

Modes of Collection

- Primarily compiled from administrative records, financial reporting systems, operational software (such as customer service software), and compliance reports
- Also include data products such as performance metrics, Freedom of Information Act (FOIA) request logs, and other government records

Legal and Policy Context

Accountability data are often required by Congress as part of authorizing legislation for federal programs. Agencies must design collection and reporting systems that satisfy both the accountability mandate and existing privacy frameworks. Generally, accountability data become subject to FOIA access rights, Information Quality Act standards for accuracy, and Open Government Data Act requirements for proactive publication, creating a complex overlay where the same dataset may be governed by different rules.



Example datasets:

- **USASpending data** are subject to the DATA Act, requiring agencies to standardize public reporting on spending; they operate under FFATA transparency mandates, and have limited [Privacy Act](#) application as they focus on federal spending transparency not PII.
- **Freedom of Information Act Request logs** are governed by [FOIA](#) for the data collected in the request process itself, Privacy Act protections for personally identifiable information in requests, and Privacy Act exemptions that align with FOIA exemptions for personal information.
- **SNAP Application Processing Timeliness data** comply with the Food and Nutrition Act, establishing program requirements for this food assistance program, defining federal performance indicators, and requiring states to report to USDA on how long it takes to process applications. Data from individual applicants are protected through the Privacy Act.

SNAP Application Processing Timeliness

[LINK TO DATASET ↗](#)

U.S. Department of Agriculture

USDA Food and Nutrition Service
U.S. DEPARTMENT OF AGRICULTURE

SNAP Application Processing Timeliness

An application is considered processed timely if the household has an opportunity to participate within 30 days of the application date for regular processing or within 7 days of the application date for expedited service.

SNAP Application Processing Timeliness Rate by State

State	↕	Rate	↑
Alaska		38.98	
District of Columbia		48.13	
North Dakota		52.94	
New York		60.45	



A grandparent who recently took in their grandchild needs a little help putting food on the table for this growing teenager, so they apply for SNAP benefits (formerly known as *food stamps*).

If it takes more than a week, or especially more than a month, for that assistance to start, they'll be in bind. Congress recognized this, and that's why they require states to report on the percent of eligible applicants receiving benefits on time.

These accountability data enable local reporters to tell the story when bureaucracy is getting in the way of children getting the nutrition they need, and put state governments on notice that they need to deliver for their constituents.

Examples of Accountability Data

VA: [Geographic Distribution of VA Expenditures \(GDX\)](#)

Annual breakdown of VA spending (medical care, compensation, education) across states, counties, and congressional districts

OMB: [Approved Apportionments Data](#)

Statutorily-required transparency data on how OMB directs agencies to spend funds appropriated by Congress

DOJ: [Freedom of Information Act \(FOIA\) Requests](#)

Tracks agency processing of FOIA requests, including request volumes, outcomes, agency responsiveness, and backlogs

OMB: [Information Collection Review Data](#)

Data on federal agency proposals to collect information from the public (surveys, forms) under the Paperwork Reduction Act, tracking burden hours and costs for public oversight.

Treasury: [USASpending](#)

Tracks financial awards like contracts, grants, and loans across all federal agencies

DHS/USCIS: [Historical National Median Processing Time](#)

Processing times (in months) for 30+ immigration forms (I-90, I-485, N-400 etc), published annually by fiscal year

SSA: [National 800 Number Network Avg Speed to Answer](#)

Monthly performance metric since 2007 showing average time Social Security Administration call centers take to answer calls

CMS: [State Medicaid and CHIP Applications, Eligibility Determinations, and Enrollment Data](#)

State-by-state data tracking of Medicaid and Children's Health Insurance Program (CHIP) uptake and implementation

DOE: [Federal Government Energy and Water Use](#)

Annual energy consumption and water use by all federal agencies to support federal efficiency goals

DHS/CBP: [Assault and Use of Force](#)

Historical and current (with one month lag) reporting on the number of incidents and types of force used by CBP agents 24

Evaluation Data

Evaluation data are used to analyze the design, implementation, and impact of government initiatives to determine value, relevance, and areas for improvement. They inform **program improvement**, **resource allocation**, and **evidence-based policymaking**.

Evaluation data are often collected for the express purpose of evaluation as a one-time effort. Ongoing collection of administrative and accountability data, as well as statistical and geospatial data, can also help inform the evaluation of federal programs.

Evaluation data are often coordinated by dedicated evaluation offices within agencies, led by designated Evaluation Officers, as required by the Evidence Act. These offices ensure evaluation activities align with agency learning agendas and evidence-building plans.



Evaluation Data

Modes of Collection

- Usually collected through studies designed for a specific program or policy and often managed by the agency or a research partner leading the evaluation
- Typically use mixed methods, including surveys, interviews, experiments, and analysis of existing administrative or statistical data sources.



Legal and Policy Context

Evaluation data fall primarily under the [Evidence Act](#) framework, which requires agencies to develop learning agendas and annual evaluation plans that guide systematic studies of program effectiveness and outcomes.

Evaluation datasets maintain privacy protections for participants while enabling evidence-based policymaking.

When evaluation data involves personal information from program participants, it requires [Privacy Act](#) protections including SORNs. Some [statistical](#) data may be used for evaluation and therefore also be protected by laws and policies governing federal statistics such as [CIPSEA](#). Published evaluation findings must meet [IQA](#) standards for accuracy and reliability.

Examples of the legal and policy context for evaluation data:

- HHS' Pathways to Work Evidence Clearinghouse has limited Privacy Act implications as it reviews existing studies rather than collecting PII; but the dataset follows PRISMA guidelines for systematic review standards.
- The NSF Innovation Corps Program maintains Privacy Act protections for participant personal information, follows NSF policies ensuring de-identification of public-use files, and operates under [PRA](#) to minimize respondent burden.

Weatherization Assistance Program (WAP) Retrospective Evaluation

[LINK TO DATA SET ↗](#)

U.S. Department of Energy, Oak Ridge National Laboratory (ORNL)



A family in Kansas is no longer being poisoned by carbon monoxide from their hot water tank because an evaluation of DOE's Weatherization Assistance Program (WAP) spurred it to improve its health and safety assessments when going into homes.

Data collected through ORNL's evaluation of WAP found that although the program was successfully delivering energy savings, there were missed opportunities for weatherization workers to make homes even more healthy and safe. DOE standardized its protocols to drive more consistent evaluations for health and safety and improved the training for this federally-funded workforce. Now, ~35,000 low-income households each year not only benefit from lower energy costs, but also from safer and healthier homes with improvements like carbon monoxide alarms, checks that heating systems are working safely, and indoor air quality strategies like kitchen exhaust fans.

Examples of Evaluation Data

USDA: [WIC Tribal Organizations & U.S. Territories Study](#)

Qualitative study assessing Women, Infants, and Children programs in Tribal Organizations and U.S. Territories to inform improvements

HHS/HRSA: [Telehealth Resource Center Performance Measurement](#)

Data collected across health systems on millions of telehealth appointments to evaluate access, quality, and utilization

DOL/SSA: [RETAIN \(Retaining Employment & Talent After Injury/Illness Network\) Evaluation](#) Assesses how stay-at-work/return-to-work strategies for injured/ill workers impact retention and reliance on disability benefits

ED: [Pathways to Partnerships Program Evaluation](#) Evaluates state approaches to employment and education transition services for children/youth with disabilities through interagency collaboration

NSF: [Innovation Corps \(I-Corps\) Program dataset](#)

Tracks immediate and long-term outcomes (business creation, funding, career changes) for scientists trained to commercialize federally funded research

USAID: [Malaria Behavioral Survey \(MBS\) Data](#)

Household survey data collected across African countries to measure use/effectiveness of malaria prevention tactics like bed nets, antimalarials, and health-seeking behaviors

Evidence and Evaluation Clearinghouses

- HHS/ACF:
 - [Pathways to Work Evidence Clearinghouse](#)
 - [Home Visiting Evidence of Effectiveness \(HomVEE\)](#)
 - [Title IV-E Prevention Services Clearinghouse](#)
- DOJ/NIJ: [CrimeSolutions](#)
- SAMHSA: [National Registry of Evidence-Based Programs and Practices \(NREPP\)](#)
- Education/IES: [What Works Clearinghouse](#)

Navigation Data

Navigation data characterize federal and other resources in a structured way, in order to help people more efficiently and effectively identify relevant **benefits, services, information, and programs**. Agencies often use navigation data to build a public-facing website that allows non-technical users to access the data.

Non-governmental entities can add value to federal navigation data by using application programming interfaces (APIs) to build digital tools that meet the specific needs of a particular audience, such as **local governments seeking infrastructure funding** or **cancer patients seeking clinical trials**.

Some navigation datasets only include federal resources, such as the Medicare Provider Data Catalog. Data.gov, the U.S. government's official open data portal, includes metadata about federal and local datasets. Others provide a trusted place for people to find data about governmental *and* non-governmental resources, such as NIH's PubMed free database of the biomedical literature.

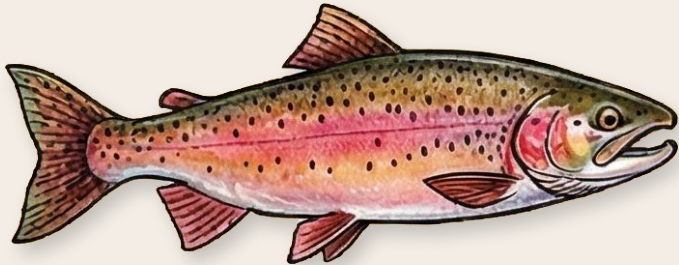


Navigation Data

Modes of Collection

What distinguishes navigation data are the structured variables that enable data users to search for information that meets their need, whether they are looking for a clinic, a grant opportunity, or allowed exceptions for environmental permitting requirements.

- Navigation data are assembled and maintained through a combination of centralized inventories, [administrative](#) and [reference](#) data, and interagency information-sharing.
- Input from program administrators, ongoing public feedback, and direct usability testing inform the continuous refinement and expansion of variables in navigation datasets.



Legal and Policy Context

Navigation data falls primarily under the [OGDA](#), which requires this type of non-sensitive information to be published in machine-readable formats to enable third-party tools. Because navigation data typically don't contain personal information, [Privacy Act](#) protections do not apply, though any user data collected through interactive systems requires appropriate safeguards.

Navigation data must also meet [IQA](#) standards for accuracy and reliability because incorrect navigation information could prevent eligible individuals from accessing crucial services, information, and benefits.

Examples of additional legal/policy contexts for specific data:

- **Grants.gov data:** OMB Grant Memoranda and Uniform Guidance cover grant information-sharing, and FFATA requires standardized disclosure of funding opportunities.
- **ClinicalTrials.gov data:** The FDA Amendments Act and NIH Policy on Dissemination of NIH-funded Clinical Trial Information require clinical trials to register with [clinicaltrials.gov](#). The WHO and International Committee of Medical Journal Editors Policy both have clinical trial registration policies that invoke [clinicaltrials.gov](#).

Health Care Access & Quality Information

[LINK TO DATASET ↗](#)

U.S. Department of Veterans Affairs

Veterans and their caregivers can find a nearby VA medical center or nursing home that best meets their needs with data about what types of care are offered, how long it takes to get an appointment, what the quality of care is, and how satisfied patients are with the care they receive.



Compare VHA Facilities

VA data may not be available if services are not offered, technical limitations prevent data collection or sample sizes are too small.

- Effective Care
- Safe Care
- Veteran-Centered Care
- Wait Times

[Add Facility](#)

Biloxi, MS

Jackson, MS

Average wait times (A)

New and established patient wait times

[Established](#)
[New](#)

Service ↑	New (A) ↑↓	New (A) ↑↓
AUDIOLOGY	29 days	14 days
CARDIOLOGY	87 days	25 days
DENTAL	64 days	N/A
DERMATOLOGY	43 days	63 days
ENDOCRINOLOGY	25 days	37 days
GASTROENTEROLOGY	62 days	57 days
MENTAL HEALTH GROUP	50 days	N/A
MENTAL HEALTH INDIVIDUAL	46 days	19 days

Examples of Navigation Data

NIH: [PubMed](#)

Free database of 29M+ biomedical journal articles and books serving researchers, doctors, and the public for 30 years

CEQ: [Permitting Innovation Center Categorical Exclusions](#)

One-stop shop for agency rules on pre-approved minor actions that are exempt from NEPA environmental studies

USPTO: [PatentsView](#)

AI-enhanced database of 40 years of U.S. patent data designed to improve access, utility, and findability

GSA: [Data.Gov](#)

Statutorily required, government-wide catalog for metadata of federal agency datasets to help users find the data they need

USDA: [SNAP Retailer Locator](#)

Locations of grocery stores and farmers markets authorized to accept SNAP benefits (food stamps) from recipients

CMS: [Medicare Provider Data Catalog](#)

List of Medicare-enrolled physicians, group practices, hospitals, and clinics managed by CMS

GSA: [FAS AAS Opportunities Dashboard](#)

Current and past federal contract opportunities offered through GSA's Assisted Acquisition Service programs

NIH: [ClinicalTrials.gov](#)

Public registry of human clinical trials (protocol, eligibility, results) with ~40k new studies posted each year globally

HHS: [Grants.gov](#)

Consolidates funding opportunities from federal agencies so it is easier to find, apply for, and manage grants

USITC: [Harmonized Tariff Schedule](#)

Official source for all U.S. merchandise imports, with current tariff rates and product classification numbers for entering goods

Reference Data

Reference data provide **structure, standardization, and consistency** across other data sources.

Federal reference datasets:

- Take the form of **unique classification** and **geographic codes, geospatial boundaries, official names** of geographic areas or categories, and **administrative boundaries**;
- Serve as the backbone for **standardization, integration, and interoperability** across a wide variety of government and private-sector datasets;
- Enable **linking or joining** of datasets that can deliver more **sophisticated insights**, and **reduce redundant data collections**; and
- Provide an **authoritative foundation** for **data innovations** in the private and public sectors, and across all domains, including technology, healthcare, and public safety.



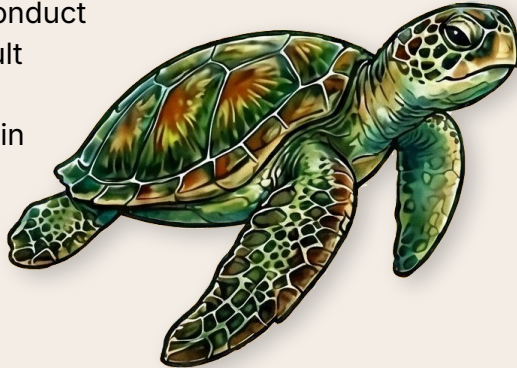
Reference Data

Modes of Collection

Reference data can be collected through:

- Administrative processes that establish official codes, standards, definitions, and boundaries;
- Statistical processes that apply statistical data (such as population) to standardized, objective definitions;
- Legislative and regulatory mechanisms or agency directives;
- Geospatial surveying or analysis, including automated processes such as those used to determine land use categories from satellite images.

Agencies often gather input from other governmental units, use public records, conduct periodic reviews, and consult subject-matter experts to ensure reference data remain accurate and aligned with public needs.



Legal and Policy Context

All federal reference data must be included in data inventories required by the [Evidence Act](#) because they are foundational for other datasets, and non-sensitive reference data should be proactively published under the [OGDA](#) in machine-readable formats to enable easy standardization across government and third-party applications.

Geographic reference data (such as administrative boundaries and geographic codes) are subject to the [Geospatial Data Act](#), which requires adherence to Federal Geographic Data Committee standards to ensure consistency across agencies.

Because reference data underpins other datasets and analyses, it must meet [IQA](#) standards for accuracy and reliability, as errors in foundational reference data can cascade across systems and analyses.

National Spatial Reference System (NSRS)

[LINK TO DATASET ↗](#)

NOAA/National Geodetic Survey

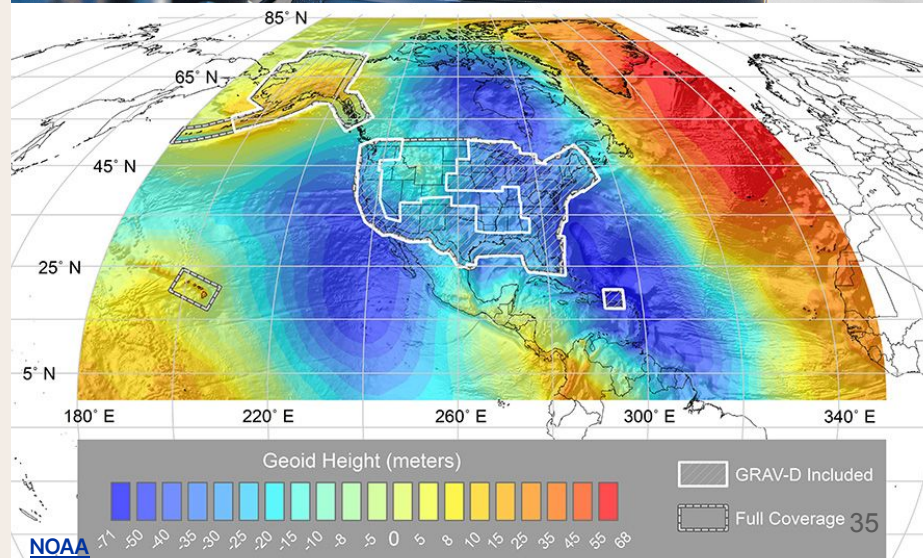
The Earth's shape, size, gravitational field, and orientation in space are changing all the time, which has major implications for things we take for granted on the ground, like latitude, longitude, and elevation.

Self-driving cars are able to stay safely in their lanes because they rely on the official geographic location and elevation data from the National Spatial Reference System.

NOAA's National Spatial Reference System provides a single source of truth for positioning, meaning that every self-driving car company doesn't have to create its own version of positional reality.



iStock: Manhattan001



Examples of Reference Data

DOI/National Park Service: [Administrative Boundaries of National Park System Units](#)

Defines where national parks officially begin and end, ensuring trails, campgrounds, and facilities are correctly placed on maps

NOAA: [U.S. Maritime Limits & Boundaries](#)

Official map of ocean area boundaries where the U.S. has legal control, such as territorial waters and exclusive economic zones

OMB: [Metropolitan & Micropolitan Statistical Area Delineations](#)

Geographic groupings of counties that are economically and socially integrated around a population core, used for statistical purposes and informing federal program eligibility

OMB: [North American Industry Classification System \(NAICS\)](#)

Standard classifications of businesses (e.g., retail trade, utilities, and manufacturing,) key to understanding the U.S. economy

Census Bureau: [Census Designated Places](#)

Boundaries and names of towns or places where people live close together, with no official city government or legal border

Census Bureau: [TIGER/Line Shapefiles](#)

Depicts legal and statistical boundaries, roads, rivers, and railroads for the U.S. for mapping and data tabulation

FDA: [Orange Book \(Approved Drug Products with Therapeutic Equivalence Evaluations\)](#)

Official list used by pharmacists and states to determine if a generic drug can be safely substituted for a brand-name drug

NIST: [National Vulnerability Database](#)

Official, uniquely-identified list of known software/hardware security flaws, including severity of each flaw and its impacts

Education: [Federal School Code Lists](#)

6-digit identifiers for U.S. colleges, universities, and other post-secondary institutions to enable financial aid, other programs

Legal & Governance Landscape

The legal and governance landscape undergirding the federal data apparatus is fundamental to understanding the different types of federal data, why the data are produced, how they must be protected, and under what conditions they can be shared or made public.

Federal data policies have evolved and layered over decades in attempts to balance sometimes competing priorities: protecting privacy and security, ensuring transparency, promoting use of data and evidence in policymaking, and enabling legitimate research and oversight. The policies described in the following pages overlap with each other in some places, and leave gaps in others.

This is by no means a complete list, and is not intended to be a legal reference.

Privacy & Security

Security and privacy are among the most widely discussed aspects of federal datasets, addressing two distinct but related concerns. **Security** refers to the protection of data from unauthorized access, modification, or destruction. **Privacy** refers to the individual's right to restrict the disclosure of their personal information, as well as to access and amend their own records. It also requires agencies to be transparent about their record-keeping systems.

When the government collects sensitive information, from tax returns to health records to census responses, **any perceived failure in privacy or security safeguards reduces public trust in government and can, in turn, reduce data quality.**



FISMA (FIS-muh)

[FISMA](#) addresses the confidentiality, integrity, and availability of federal information systems. Focused on protecting data from external and internal threats, it applies to executive branch agencies, federal grantees, and contractors. Highlights include:

- Standards set by Commerce's National Institute of Standards and Technology (NIST);
- Requirements for security protections commensurate with magnitude of harm that would result from a breach; and
- Annually updated implementation guidance from OMB to keep up with ever-changing cybersecurity risks.

OMB Circular A-130

[A-130](#) describes how federal agencies should manage information as a strategic asset. It establishes the government-wide policy for planning, budgeting, governance, acquisition, and management of federal information, including personnel, equipment, funds, IT resources, and supporting infrastructure. A-130 also guides agencies on securing systems that collect personal information and confidential data.

Privacy Act

The [Privacy Act of 1974](#) is the cornerstone of federal data privacy protection, establishing the fundamental rights for individuals whose information is held by government agencies. Passed in response to concerns about government surveillance and data misuse, it governs how federal agencies handle personally identifiable information (PII).

The Privacy Act was enacted at the dawn of the computer age in response to growing public fears about the government's ability to maintain massive, secret databases, enabled by new technology. An outgrowth of more than a decade of discussions within the federal government about establishing a national data center for storing and linking information on American citizens, combined with high-profile illegal surveillance scandals like Watergate and COINTELPRO (Nixon's FBI-led political surveillance program), the Privacy Act ensures that PII is collected, stored, and used by the government in transparent ways.

OMB Circular [A-108](#) provides federal agencies with guidance on implementing reporting requirements under the Privacy Act, and operates in concert with Both the [OMB FISMA Guidance](#) and [OMB Circular A-130](#).



Core requirements include:

Systems of Records Notices (SORNs): Agencies must publish detailed public notices that include what data are collected, how they are used, whom they are shared with, and how individuals can access their own data.

Individual Rights: Citizens have the right to access their own records and request corrections to inaccurate information.

Disclosure Limitations: Personal information can only be disclosed for authorized "routine uses" or with individual consent.

Data Matching Rules: Special procedures govern computer matching of records across different agency systems, including between federal and state and local government systems.

Private Right of Action: Individuals have a private right of action to sue federal agencies in federal district court for specific violations of the Privacy Act.

Confidential Information Protection and Statistical Efficiency Act (CIPSEA; SIP-sea)

Generating statistics for policy and operational decision requires its own distinct legal framework. For example, when BLS collects employment data from employers, businesses share sensitive information about revenues, employment levels, and operational details. These data power critical economic indicators like GDP calculations and productivity measures, yet participating businesses need assurance their responses won't be shared with competitors, regulators, or tax authorities.

[CIPSEA](#) addresses this challenge of maintaining the highest levels of privacy while also maximizing public benefit public benefit.



Key protections include:

Use Restrictions: Data collected for statistical purposes can only be used for describing or analyzing group characteristics, never for enforcement, regulation, or identifying individuals.

Confidentiality Pledges: All staff accessing protected statistical data must sign legally binding confidentiality agreements; violations constitute Class E felonies.

Limited Data Sharing: Statistical agencies can access data from other federal agencies for limited purposes only.

Immigration/Law Enforcement Shield:

CIPSEA-compliant data are explicitly protected from use in immigration enforcement or other law enforcement activities

Secure Data Handling: Agencies must implement robust access controls, privacy-preserving techniques, and secure systems to store the data

Title 13

[Title 13](#) regulates how Census collects, protects, and disseminates data. The law's primary mandate ensures that private data remains strictly confidential; personal identifiers such as names, addresses, Social Security Numbers, and telephone numbers are never published or shared.

Like CIPSEA, Title 13 provides for significant penalties to anyone who violates this law. The Census Bureau is restricted to using data it collects solely for statistical purposes. Most importantly, the law prohibits any government agency, including law enforcement or the court system, from using a respondent's personal information against them. These protections are further reinforced under Title 13 as it requires Census staff and contractors who work with data to commit to protect Census data. This commitment persists even after their employment ends. Violating these confidentiality requirements is a major federal crime, punishable by up to five years in prison, a fine of up to \$250,000, or both.



Title 26

[Title 26](#) represents a large set of laws that is often referred to as the Internal Revenue Code. In data policy, section 6103 of Title 26 provides the special conditions under which the Internal Revenue Service may share individual tax data with other agencies.

For instance, the Census Bureau and Bureau of Economic Analysis are authorized to receive tax data for statistical purposes. Other recipients may include the Social Security Administration for benefits management and state tax authorities for tax administration. In legal matters, law enforcement access is generally restricted to cases involving a federal court order.

Any agency receiving tax data must comply with rigorous safeguarding requirements, including physical security and regular audits. Under Title 26 even knowledge about whether an individual filed taxes or not is treated confidentially. Unauthorized sharing of tax data is a felony, providing a powerful deterrent against its misuse.

Family Educational Rights (FERPA; FUR-puh)

FERPA applies specifically to educational agencies and institutions that receive funding from the U.S. Department of Education. It is designed to protect the privacy of student education records while ensuring access for parents and government officials, as well as students once they turn 18 or attend post-secondary school at any age.

The primary impact of publicly accessible federal data about students collected under FERPA is that education records may only be shared in a way that students cannot be identified.

Health Insurance Portability & Accountability Act (HIPAA; HI-puh;)

HIPAA addresses the privacy and security of sensitive medical information for “covered entities,” which includes health plans, healthcare clearinghouses, and healthcare providers. It addresses:

- **Privacy**, with standards that require patient authorization before disclosing personal health information; and
- **Security**, requiring implementation of technical, administrative, and physical safeguards for electronic health data.

Federal agencies that handle patient health records (like the VA hospitals, or OPM's Federal Employee Health Benefits program) must comply with HIPAA as any other covered entity. The HIPAA “Safe Harbor” rule regulates how covered entities can use and disclose personal Health information. In general, personal identifiers must be removed from data products derived from HIPAA-protected patient records.



Transparency and Evidence-Building

Public access to federal data is necessary for running a modern society and accountable government, and is in constant tension with privacy.

FOIA (FOY-uh; Freedom of Information Act)

The [Freedom of Information Act \(FOIA\)](#) creates public access to government records while protecting individual privacy. FOIA provides the primary mechanism for the public to request access to federal agency records, with specific exemptions that align with the protections afforded by the Privacy Act. Following a revision to the Act in 2016, agencies are required to “make available for public inspection in an electronic format, records that have been requested 3 or more times.”



OPEN Government Data Act (OGDA)

The [OPEN Government Data Act of 2018 \(OGDA\)](#) represents a shift toward “open by default,” moving beyond reactive FOIA responses toward proactive data publication. The OGDA is part of the broader Evidence Act, with a focus on:

- **Default Openness:** All non-sensitive federal data must be made available in open, machine-readable formats unless specifically prohibited by law.
- **Data Inventories:** Agencies must maintain comprehensive catalogs of their data assets with appropriate metadata, and publish their data at data.gov.
- **Public Input:** CDOs should engage public stakeholders around what datasets to prioritize for release and how the agency can improve its data offerings.

Just like FOIA, OGDA explicitly recognizes existing privacy protections such that data covered by the [Privacy Act](#), [Title 13](#), [Title 26](#), and confidentiality laws remain protected while maximizing access to non-sensitive information.

Evidence Act

The [Evidence Act](#) bridges the gap between privacy protection and data accessibility, creating new structures and requirements that touch nearly every aspect of how federal agencies collect, share, and use data.

It requires agencies to develop multi-year research plans called *learning agendas* that identify key policy questions and inform what data agencies should prioritize collecting. It also includes the [Open Government Data Act \(OGDA\)](#) and its provisions around comprehensive data inventories and metadata standards so that federal datasets are more discoverable and usable across agencies.

The Evidence Act creates three key positions at every major federal agency:

- **Chief Data Officers** lead agency-wide data governance, data inventories, and both internal and public data access.
- **Evaluation Officers** oversee evidence-building and ensure evaluation results inform policy decisions).
- **Statistical Officials** ensure data quality, appropriate methodology, and compliance with statistical standards.

[CIPSEA](#), which is Title III of the Evidence Act:

- reinforces the separation between statistical and operational government functions by strengthening protections for survey respondents and statistical data subjects;
- prohibits the use of CIPSEA-implicated statistical data for law enforcement purposes; and
- enables statistical agencies to share data with each other while maintaining strict confidentiality protections.

The Evidence Act's emphasis on data sharing and coordination is critical given [Paperwork Reduction Act](#) constraints that limit agencies' ability to collect new data from the public.



Data Quality

The quality of federal data is dependent on many factors, including whether: people trust the government to keep their data secure and private; collection instruments are burdensome; measurement devices are properly maintained and functioning; and administrative mechanisms ensure data are objective, useful, and secure from manipulation.

Paperwork Reduction Act (PRA)

With the intent to minimize paperwork burden and improve participation rates, the PRA constrains how agencies collect information from the public, requiring OMB approval for instruments including surveys, forms, or interviews involving 10+ people. Most applicable to [statistical](#) and [administrative](#) data, the PRA requires agencies to justify the need for the data, provide time and cost estimates for respondents, and describe how the data will be used.

Substantive changes to data require public notice and comment—one of the few formal triggers for stakeholder engagement within the federal data apparatus.



Information Quality Act (IQA)

OMB's IQA guidelines support three aspects of federal data quality at federal agencies:

1. **Objectivity:** substantively complete, accurate, and unbiased, using reliable methodology and sound analytic techniques
2. **Utility:** useful to the public and for policymaking
3. **Integrity:** secure from corruption, falsification, or unauthorized access

For "influential" data that will have a clear and substantial impact on public policies or private-sector decisions, the data and methods must also be transparent enough to be reproduced by qualified third parties and maybe subject to peer-review. In general, the more influential a dataset is, the higher its quality should be.

One unique feature of the IQA is that it includes an administrative mechanism for "correction," where individuals, companies, or advocacy groups can file a request for correction if they believe a given influential dataset is of poor quality.

Specialized Data Governance

Some categories of federal data present unique technical, security, or coordination challenges that require specialized approaches to governance, building upon broader principles of utility, privacy, and transparency.

Geospatial Data Governance

The Geospatial Data Act (GDA) of 2018 addresses the unique characteristics of location-based information, with a National Spatial Data Infrastructure that includes the “technology, policies, criteria, standards, and employees necessary to promote geospatial data sharing” throughout all levels of government and the private sector.

The GDA recognizes that geospatial data are most valuable when layered, combined, and analyzed across different sources and jurisdictions.

A flood response, for example, requires NOAA weather data, census data, private utility data, and local zoning, spatially aligned and interoperable.

By establishing the Federal Geographic Data Committee (FGDC), the GDA provides leadership and coordination for federal geospatial activities, while the GeoPlatform serves as a public portal for accessing geospatial datasets and metadata. Agencies must:

- use standards for data quality and discoverability;
- regularly audit how they collect, use, and share geospatial data to identify gaps, and reduce duplication; and
- provide open access to non-sensitive geospatial data while protecting sensitive location information that could pose security risks.



Scientific Data Governance

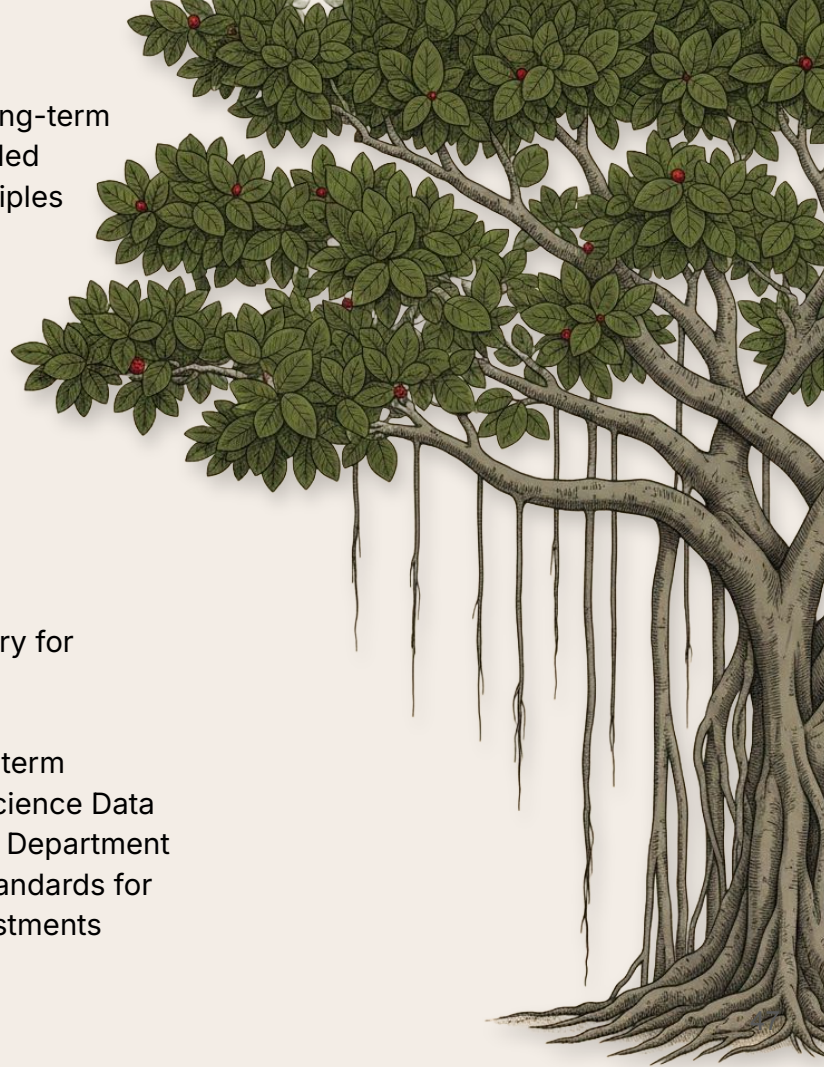
Scientific data must meet standards for reproducibility, peer review, and long-term preservation to maintain value for future research and validation. This has led to specialized governance frameworks that emphasize open science principles while protecting intellectual property and national security interests.

For example, federal scientific agencies increasingly require that research data be findable, accessible, interoperable, and reusable (FAIR). This means datasets must:

- include comprehensive metadata,
- use standard formats and vocabularies,
- be archived in persistent repositories, and
- remain available for verification and reuse by other researchers.

Agencies like the NIH, NSF, NASA, NOAA, and EPA have implemented data management and sharing requirements that make FAIR principles mandatory for federally-funded research.

Federal agencies maintain specialized data repositories designed for long-term scientific data preservation and access. Examples include NASA's Earth Science Data Centers, NIH's research databases, NOAA's climate data archives, and the Department of Energy's scientific data repositories. These systems must meet strict standards for data integrity, searchability, and preservation to ensure that scientific investments remain accessible for decades or centuries.



Glossary of Acronyms and Initialisms

API	Application Programming Interface	FERPA	Family Educational Rights & Privacy Act	NIH	National Institutes of Health
BJS	Bureau of Justice Statistics	FFATA	Federal Funding Accountability & Transparency Act	NIST	National Institute of Standards & Technology
BLS	Bureau of Labor Statistics	FGDC	Federal Geographic Data Committee	NOAA	National Oceanic & Atmospheric Administration
CDC	Centers for Disease Control	FISMA	Federal Information Security Modernization Act	NPS	National Parks Service
CFO	Chief Financial Officers Act of 1990	FOIA	Freedom of Information Act	NSF	National Science Foundation
CIPSEA	Confidential Information Protection & Statistical Efficiency Act	GAO	Government Accountability Office	OGDA	OPEN Government Data Act
CMS	Centers for Medicare and Medicaid Services	GDA	Geospatial Data Act	OMB	Office of Management and Budget
DHS	Department of Homeland Security	GDP	Gross Domestic Product	OPM	Office of Personnel Management
DOC	Department of Commerce	GIS	Geographic Information Systems	OSTP	White House Office of Science & Technology Policy
DOD	Department of Defense	GPS	Global Positioning System	PII	Personally Identifiable Information
DOE	Department of Energy	GSA	General Services Administration	PRA	Paperwork Reduction Act
DOI	Department of the Interior	HHS	Department of Health & Human Services	SNAP	Supplemental Nutrition Assistance Program
DOJ	Department of Justice	HIPAA	Health Insurance Portability and Accountability Act	SORN	System of Records Notice
DOL	Department of Labor	HMDA	Home Mortgage Disclosure Act	SSA	Social Security Administration
DOT	Department of Transportation	HUD	Department of Housing and Urban Development	UCR	Uniform Crime Reporting
ED	Department of Education	IQA	Information Quality Act	USAID	U.S. Agency for International Development
EPA	Environmental Protection Agency	IRS	Internal Revenue Service	USDA	U.S. Department of Agriculture
FAFSA	Free Application for Federal Student Aid	NASA	National Aeronautics & Space Administration	USPS	U.S. Postal Service
FAIR	Findable, Accessible, Interoperable, and Reusable	NEPA	National Environmental Policy Act	VA	Department of Veterans Affairs
FBI	Federal Bureau of Investigation			WHO	World Health Organization
FDA	Food and Drug Administration				

Credits

Authors

Denice W. Ross is a Berkeley Executive Fellow in Applied Technology Policy and served as the nation's second U.S. Chief Data Scientist

Christopher Steven Marcum, Ph.D. is an expert on statistical and scientific data who served as Assistant Director for Open Science and Data Policy at the White House OSTP

Student Researchers

Saanvi Arora, Bachelor's Degree, Computer Science & Legal Studies (2026 expected)

Chelsea Chen, Master of Information Management Studies (2026 expected)

Isabelle Qian, Bachelor's Degree, Computer Science, Philosophy, & German Studies (2027 expected)

Designers

Audrey Denson of [Denson Design](#), **Denice W. Ross**, and Google's **Gemini Nano Banana** (Fall 2025 version)

About the Executive Fellowship in Applied Technology Policy

The Executive Fellowship in Applied Technology Policy is designed to empower senior leaders at the intersection of technology and public policy. Jointly led by UC Berkeley's School of Information and Goldman School of Public Policy, this fellowship brings together accomplished public-sector technology leaders and distinguished academic thought leaders to shape the future of digital governance and policy.

Advisors

Deirdre Mulligan, Professor, School of Information, UC Berkeley
Meghan Maury, former Chief of Staff, Office of the Undersecretary for Economic Affairs, Department of Commerce

Recommended Citation

Ross, Denice W. and Marcum, Christopher Steven. 2026. "Federal Data Field Guide." Version 1. University of California, Berkeley. DOI: <https://doi.org/10.15779/J2P043>

Website: federaldatafieldguide.us

Contact: authors@federaldatafieldguide.us

