Journalist's Resource
Research on today's news topics

GOVERNMENT, REPORTING, RESEARCH

Major administrative datasets of the U.S. government — all in one place

Tags: data journalism, dataset digest

(sec.gov)

An immense number of U.S. government agencies play a central role in the collection of a wide array of public data — vital statistics on health, transportation, commerce, finance, agriculture, and more. Much of this information is gathered by the 13 principal statistical agencies, but smaller organizations — for example, the Consumer Financial Protection Bureau, the Army Corps of Engineers and USAID — also gather important information.

All this data gathering isn't inexpensive — the 13 agencies spend an estimated $3.7 billion annually on collection, processing and dissemination — but the benefits far outweigh the costs: In a 2014 report, the Commerce Department estimates that this information adds as much as $221 billion to the U.S. economy. Even better, journalists can use this wealth of data to deepen and broaden their reporting, anchoring it in facts and figures that can better inform their communities and the decisions they make.

Below are links to data sources and tools from a broad range of federal agencies, courtesy of Katherine R. Smith, executive director of the Council of Professional Associations on Federal Statistics (COPAFS). This post is part of our ongoing dataset digest series, which highlights important information sources for journalists. If you know of other hidden datasets that deserve wider exposure, please email or ping us on Twitter.

_____

U.S. Agency for International Development (USAID)

- Foreign aid from the United States: Data and Tools. Information gathered includes international food assistance programs, health surveys and "dollars to results" data for foreign aid to dozens of countries.

Department of Agriculture

- Census of Agriculture: Conducted every five years, most recently in 2012, the Census provides uniform, comprehensive data for all U.S. states and counties, including farms by size and type, inventory and values for crops and livestock and operator characteristics.

- National Agricultural Statistics Service: Cropland data.

- The Northern Research Station's Forest Inventory and Analysis page provides access to a range of materials, including data and tools, maps, and a data collection page. The NRS technically covers the U.S. from North Dakota in the west to Maine, and down to West Virginia and Missouri, but data is available on some other states as well.

- Natural Resource Conservation Service: Conservation Financial Assistance Programs' enrollment data.

- Risk Management Agency (RMA): program costs and outlays and actuarial data. The first provides government cost for the federal crop insurance program for the past 10 years, and the second is an Actuarial Information Browser with data from 2011 to 2015.

- Web-based Supply Chain Management Reports (WBSCM) data: Supports domestic and international food and nutrition programs by the USDA, Food and Nutrition Service, Farm Service Agency, Agriculture Marketing Service, Foreign Agricultural Service and USAID. The domestic programs include the National School Lunch Program, Emergency Food Assistance Program (TEFAP), and the Food Distribution Program on Indian Reservations.

- Economic Research Services: Supplemental Nutrition Assistance Program (SNAP) data. Provides time-series data on SNAP participation and benefits levels, population, the number of persons in poverty, and some socio-demographic characteristics.

- Food Safety Inspection Services: Recalls and quarterly enforcement reports.

U.S. Army

- Army Corps of Engineers: U.S. waterborne commerce data. The available information includes principal U.S. ports; locks by waterway and tons locked by commodity group; dredging abstracts; and commodity movements by region and state.

Department of Commerce (DoC)

- Bureau of Economic Analysis (BEA): Foreign direct investments in the United States. Data on the activities of minority- and majority-owned U.S. affiliates of multinational enterprises, including sales, assets and employment.

- BEA: U.S. National Income and Product Account (NIPA) data: GDP in current and

constant dollars, personal income and spending, corporate profits and government receipts and expenditures.

- Economic Development Administration (EDA): Program data. Includes annual reports from the EDA as well as those for Trade Adjustment Assistance for Firms (TAAF) and Community Trade Adjustment Assistance (CTAA).

- Census Bureau: Business register data and the longitudinal business database. Information covered includes business formation, growth and competition; labor market dynamics; business cycles; productivity growth; and credit markets and financing.

- Census Bureau: Longitudinal employer-household dynamics. The program's mission is to provide "dynamic information on workers, employers and jobs with state-of-the-art confidentiality protections and no additional data collection burden."

- Census Bureau: County and Zip code business patterns. The data include the number of businesses, employment and first-quarter and annual payroll. It is useful for studying the economic activity of small areas, analyzing economic changes over time, judging market potential, sales-program effectiveness and budgeting.

- International Trade Administration (ITA): U.S. exporting companies data. The information, available for 2012, includes number of exporters by size, company type and industry; export market by country, region or state; metropolitan statistical area (MSA) and export market; and ZIP codes.

- ITA: The Export-supported employment data estimates of jobs supported by state exports of all types of goods as well as those supported by the export of manufactured products.

- ITA: Visitor Arrivals Program (form I-94) data includes monthly and annual arrivals in the United States; the data is provided by the National Travel and Tourism Office (NTTO) in cooperation with the Department of Homeland Security (DHS) and Customs and Border Protection (CBP). For immigration data, see the DHS heading below.

- ITA: The International Air Travel Statistics (form I-92) program data covers arrival by air into the United States.

- National Climate Data Center: National climate and historic weather data. Information on the climate of the United States by location and time of year as well as extreme events, including heat waves, droughts, tornadoes, and hurricanes.

- National Marine Fisheries Service: Recreational fisheries statistics or Commercial fisheries statistics. Also includes information on marine animals, ecosystems, stock assessments and the economic aspects of fishing.

Commodities Futures Trading Commission (CFTC)

- Filings, transactions and other data. Includes lists of registered entities (DCMs, SEFs, DCOs and SDRs), applications, certifications and requests filed by registered entities concerning products and rules; public comments; requests for action and actions taken by the CFTC. You can search by organization, rules and amendments, products and

actions.

- Market report data. Information on futures and options markets as well as research and reports conducted by CFTC economists.

Consumer Financial Protection Bureau (CFPB)

- Credit card agreement database. A database of credit-card agreements from more than 300 issuers, searchable by the name of the issuer or by the text within the agreement.
- Consumer complaint database. Information about the subject and date of the financial-product complaint and the company's response; no personal data is shared. The entire dataset can be downloaded and an API is available.

Consumer Product Safety Commission (CPSC)

- Injury and death statistics. Figures on a wide range of consumer goods, including amusement rides, ATVs, fireworks, nursery products and children's outdoor equipment, home furnishings and fixtures, scooters, trampolines and toys. Data is also available on carbon monoxide deaths, drownings, lead poisoning, electrocutions, residential fires, senior hazards, emergency room injuries, sports injuries, and pediatric poisoning.

Department of Education (DoE)

- Civil rights data for public schools. Information on education access and equity for U.S. public schools, for single schools as well as school- and district-level summaries. With the data tables, you can compare across multiple schools and districts and also get state and national estimates.
- EDFacts data for K-12 educational programs. Centralized performance data supplied by K-12 state education agencies (SEAs), including financial grant information.
- National Center for Education Statistics: Common core of data on public schools. Includes fiscal and non-fiscal data on all public schools and districts as well as state education agencies.
- Federal student aid data. Application volume reports, Title IV Program volume by school, federal student loan portfolio and default rates.
- National reporting system data for adult education. Status of the state-administered grant program authorized under the Adult Education and Family Literacy Act (AEFLA), including state and national performance data and student characteristics.
- Nation's report card system data. Reported by the National Center for Education Statistics (NCES), the data is from the National Assessment of Educational Progress (NAEP). Subjects include the arts, civics, economics, geography, math, reading, science, history and writing.

Department of Energy (DoE)

- Energy Information Administration (EIA): Energy prices data, including crude oil and

natural gas prices, oil imports, gasoline and diesel prices, refiner prices of residual fuel oil, retail prices of electricity.

- EIA: Crude oil production and stocks data covers petroleum and other liquid fuels.
- EIA: Renewable energy market data, released monthly in PDF form, shows short- and long-term trends in production and consumption of renewable energy as compared to fossil fuels and nuclear, in both graphical and tabular form. Data is divided by source and sector.

Environmental Protection Agency (EPA)

- Air quality data. Information on air quality collected at outdoor monitors across the United States, both as reports and data visualizations. There's also an interactive map from which data can be downloaded.

- Enforcement dockets data contains information on administrative penalty cases filed by the EPA regions and headquarters. Dockets can be searched by region, year and statute.

- National Pollutant Discharge Elimination System (NPDES) permits and compliance data. Searchable database of information on facilities registered with the Federal Enforcement and Compliance (FE&C) and holding NPDES permits. Can be searched by facility name, permit number, location, industrial classification and chemicals. All states are searchable except Wyoming.

- Toxic Substances Control Act Chemical Substance Inventory (TSCACSI): The existing chemicals home page provides access to a wide range of information on risk assessment, toxic releases, pesticides, and other chemical-related questions. Also see the Agency for Substances and Disease Registry (ASDR) section below.

- Superfund sites (CERCLIS database). Information on current hazardous-waste sites, potential sites and remedial activities in the United States.

Equal Employment Opportunity Commission (EEOC)

- Enforcement and litigation statistics on employment discrimination. Tabular data on employment-discrimination charges and resolutions, from 2009 to 2013. The charges can be explored by issue — including race, equal pay, national origin, religion, gender — and statute.

Federal Court System

- Bankruptcy statistics, including personal and business filings as well as abuse protection and Consumer Protection Act figures.

Federal Deposit Insurance Corporation (FDIC)

- Financial-industry data, with information on specific banks, analyses on the banking industry and economic trends.
- Failed bank data. List of failed financial institutions since October 2000, summaries,

points of contact and historical statistics.

Federal Emergency Management Agency (FEMA)

- Assistance record data. Fields include applicant ID, application title, county, category of damage, disaster number, federal share, project amount and size, and state.

Federal Financial Institutions Examination Council (FFIEC)

- Financial and structural data for FDIC-insured institutions. Information on Call Reports and Uniform Bank Performance Reports (UBPR) are organized by subject, and include earnings, balance sheet, asset quality, liquidity and capital.
- Home mortgage loans data. Home mortgage lending activity in the United States, including flat files, maps and charts. You can also generate custom queries.

The Federal Reserve

- Consumer credit data. Yearly, quarterly and monthly figures on outstanding credit balances, flow and changes; terms of credit; and interest rates, maturity, loan-to-value ratios and finance amounts.
- Finance companies data. Aggregate information on debt held, flows and changes. Includes loans for new and used cars, real estate and equipment, as well as other receivables.
- Foreign-exchange rates. Weekly data on a range of major currencies as well as broad trends.
- Government receipts for expenditures and investments. Includes expenditures for consumption-related items, transfers, investments and social benefits.
- Money stock measures. Two years of monthly data on the supply of cash, deposits and checks outside of the private banking system (known as M1) and the sum of M1 plus savings and money-market accounts, retail money market mutual funds and CDs under $100,000 (known as M2).
- Treasury account series data. Daily, weekly and monthly data on Treasury bills, including four-week, three-month, and one-year. Both secondary-market and auction averages are available.
- The St. Louis Fed's FRED economic database offers more than 200,000 data series, including the consumer price index, real GDP and unemployment. FRED offers a wealth of interactive tools you can use to generate graphs of the data series of your choice.

Federal Trade Commission (FTC)

- Fraud and identity theft (Consumer Sentinel Network): The CSN report, in PDF form, contains aggregate data on more than 9 million complaints from 2009 through 2013.

Fish and Wildlife Services

- Wetlands data. Information, fact sheets and mapping services on U.S. wetlands.

General Services Administration (GSA)

- Federal procurement report data. Can be used for "geographical analysis, market analysis, and analysis of the impact of the congressional and presidential initiatives in socio-economic areas such as small business."

- Federal Subaward Reporting System (FSRS) data. Part of the Federal Funding Accountability and Transparency Act (FFATA), this system gathers data from prime awardees on the subcontracts they in turn award. USAspending also has information on prime awards and subawards, and can show trends over time.

- Small Business Goaling Report. Provided in PDF form, this report shows aggregate data on the amount, actions, and percentages for small-business contracts. Information includes amounts and percentages for businesses owned by veterans, the disabled and women, as well as those in HUBZones.

Department of Health and Human Services (HHS)

- Agency for Substances and Disease Registry (ASTDR): Environmental health webmap data, including toxicological Information, sites and facilities data, public health resource information and the Toxic Release Inventory (TRI).

- ASTDR: Hazardous Substances Emergency Events Surveillance Report.

- ASTDR: National Toxic Substance Incidents Program data.

- Centers for Disease Control and Prevention (CDC): Surveillance data on infectious and chronic disease, diabetes, disabilities, environmental health, risky behavior, HIV, sexually transmitted diseases, injury, maternal and child health, adolescent health, and occupational safety.

- CDC: National Program of Cancer Registries data. Federal statistics on cancer incidence and mortality from 2007 to 2011, produced by the CDC and the National Cancer Institute (NCI). The site provides graphical information as well as map.

- CDC: Community water fluoridation statistics, including the total and percentage of the U.S. population covered, number of community water systems (CWS).

- Center for Medicare and Medicaid Services (CMS): Medicare claims data, anonomized and available in CSV format. Additional information is available through the Medicare Data Communications Network (MDCN).

- CMS: National health expenditures data. Information on personal health care (PHC) spending by type of good or service and source of funding in five age groups and for males and females. You can also get figures for state health expenditures and business, household and government spending on health care.

- CMS: Provider of service data. Contains data on the characteristics of hospitals and other health care facilities, including their location and type of Medicare services provided.

- National Center for Health Statistics vital statistics data, including births, deaths, marriages and divorces.
- Temporary Assistance to Needy Families (TANF) administrative records, including active and closed cases.

Department of Homeland Security (DHS)

- Immigration statistics. Reports for fiscal year 2013 on the number and characteristics of those admitted to the United States as refugees or granted asylum; information on the apprehension, detention and return of non-residents; and non-immigrant admissions.

Department of Housing and Urban Development (HUD)

- Community Development Block Grants (CDBG) expenditures data.
- Family data on public and Indian housing and microdata.
- Fair-market rents data, covering the years 2000 to 2015.
- Government-sponsored enterprise data. Data sets can be ordered providing information on single-family and multifamily mortgage purchases by Fannie Mae and Freddie Mac (Government-Sponsored Enterprises, or "GSEs").
- Metropolitan area quarterly residential and business vacancy report data. Residential and business vacancies, based on data from the US Postal Service, aggregated to the census-tract level. Designed to supplement vacancy data from the American Community Survey (ACS), Housing Vacancy Survey (HVS) and American Housing Survey (AHS).
- National Low-Income Housing Tax Credit (LIHTC) database. The LIHTC program allows the issuing of tax credits for the acquisition, rehabilitation, or construction of rental housing for lower-income households. The database includes the project year, address, number of units, bedrooms, and other characteristics.
- Neighborhood Stabilization Program (NSP) data. The Neighborhood Stabilization Program provides assistance to state and local governments to acquire and redevelop foreclosed properties that might otherwise be abandoned.
- Program Income Limits Data.

Department of Interior

- U.S. Geological Survey (USGS): Biodiversity and species data. Interactive map with continuously updated records of the occurrence of plants and animals in the United States. You can search by both common and scientific name, and zoom in to specific locations.
- USGS: Land-cover and land-use data. Extensive collection of links to sites with land-related data, both in the United States and around the world.
- USGS: Water Resources data. The mission of the site is to "collect and disseminate reliable, impartial, and timely information that is needed to understand the Nation's water

resources." Users can access real-time data on streamflow conditions, groundwater quality, precipitation and more.

- USGS: Water quality data. Brings together a range of USGS sites related to water, including the water quality, aquatic bioassessment, hydrologic benchmarking and deposition of chemicals through precipitation.

International Trade Commission (ITC)

- U.S. tariff and trade data. Provides data on thousands of goods, from agricultural products to manufactured items. Includes annual value, tariff treatment and rate, and country-specific exceptions.

Department of Justice (DoJ)

- Bureau of Prisons (BoP): Inmate, population and staff statistics. Interactive graphics on key federal criminal-justice statistics, including inmate ethnicity, gender, offense and sentence.
- Bureau of Justice Statistics (BJS): Court statistics project data. Details the operation of state court systems, with information on court structure, jurisdiction, caseload volume and trends.
- BJS: Federal Justice Statistics Program data, including the annual workload, activities and outcomes of federal criminal cases
- BJS: Law Enforcement Management and Administrative Statistics (LEMAS): Data from more than 3,000 state and local law-enforcement agencies; information includes agency responsibilities, expenditures, job functions; salaries, special pay and demographics of officers; weapons and armor policies; education and training requirements; computers and information systems; and vehicles, special units, and community policing activities.
- BJS: National Corrections Reporting Program data: Annual data on prison admissions and releases, and year-end custody populations and on parole entries and discharges.
- BJS: National Incident-Based Reporting System data. Information on crimes known to authorities, including homicides and manslaughter, robberies, assault and arson.
- BJS: National Prisoner Statistics Program data. Annual national and state-level data on the number of prisoners in state and federal prison facilities. Reports are released in PDF and text-file format, and data in CSV format.
- Federal Bureau of Investigation: Uniform Crime Reports data. Wide range of crime-related data, publications and resources. Source data comes from more than 18,000 federal, state, county, city, tribal and university/college agencies participating in the program.

Department of Labor

- Bureau of Labor Statistics: Quarterly census of employment and wages.

- Foreign Labor Certification Office: H-1B Data. The H-1B non-immigrant visa allows U.S. employers to temporarily employ foreign workers in specific occupations. The data, based on the U.S. fiscal year, is available in Access and CSV format files.

- Labor retirement and welfare benefit plan data set. The Form 5500 Annual Report provides information about the operation, funding and investments of approximately 800,000 retirement and welfare-benefit plans. Related data is available from the IRS (see below).

- Occupational Safety and Health Administration (OSHA): Work-related injury or illness data. Searchable database and text-format data on a representative sample of workplace incidents, from 1996 through 2011.

- OSHA: Enforcement Data (Inspection Data). Originally an in-house resource for OSHA staff and state agencies, it includes data on OSHA interventions and enforcement activity at specific work sites.

- OSHA: Worker fatalities/catastrophes report (FAT/CAT). List of workplace incidents that result in casualties or hospitalizations.

National Aeronautics and Space Administration (NASA)

- Urban Landsat. The Cities from Space dataset contains images of 66 urban areas and raw data for 28.

Patent and Trademark Office

- U.S. Patent and Trademark Office patent data. Organized by geographic origin, organization, technology and inventor. Historical and year-end statistics are also available.

Department of Transportation

- Bureau of Transportation (BTS): Air-carrier statistics. This table contains data on domestic and international flights by U.S. carriers, including route, aircraft type and hours, capacity and load factors and departures scheduled and performed.

- BTS: Intermodal Passenger Connectivity Database. Nationwide data table of thousands of rail, air, bus and ferry terminals in the United States.

- Maritime Administration: Maritime travel and transportation statistics.

Department of Treasury

- Bureau of Fiscal Service: Public debt reports, including debt position and activity, Treasury receipts, outlays of the U.S. Government, schedules of Federal debt, and Treasury securities.

- Financial Crime Enforcement Network: Mortgage and real estate fraud data. Searchable by state, urban area and county; includes resources for the public and institutions.

- Interest Rate Statistics, including daily Treasury yield curve rates as well as Treasury bill

rates and long-term averages.

- Internal Revenue Service (IRS): Corporate Tax Statistics (Form 1120).
- IRS: Employee benefit plans (Form 5500). Related data is available from the Department of Labor (see above).
- IRS: Individual tax statistics (Form 1040).
- IRS: Quarterly payroll taxes (Form 941).

Securities and Exchange Commission (SEC)

- SEC filings. The SEC's EDGAR database allows you to search filings by company, address, state, country, and date, and well as mutual funds and insurance products; advanced search allows the use of Boolean operators. (See the related Journalist's Resource post, "How to Use SEC filings to Cover Companies," on how six common SEC filings can help you better report on U.S. businesses.)
- Mutual fund fees and expenses. Includes interactive data and mutual fund risk/return summaries.
- Program and market data. Eighteen datasets on everything from administrative law decisions to summary metrics.
- Short sale volume data.

Small Business Administration (SBA)

- Small business lender and loan data. Information on the 100 most-active 7(a) lenders as well as additional resources.

Social Security Administration (SSA)

- Social Security programs data. The SSA's Annual Statistical Supplement includes hundreds of statistical tables with data on beneficiary counts and benefit amounts, and the status of the program's trust funds.
- Earnings and employment data for workers covered under Social Security and Medicare. Available in PDF and XLS form, the report and data cover the year 2011 and were released in 2014.

Department of Veterans Affairs (VA)

- Veterans Benefits Administration reports. Status reports on key VA statistics, including claims inventory, backlog and accuracy.

Keywords: data journalism, big data, data visualization

Writer: Leighton Walter Kille | Last updated: January 6, 2015

We welcome feedback. Please contact us here.