Chapter 10

Ferreting Out Public Data Sources

In This Chapter

- ▶ Fathoming your sources
- ▶ Recognizing issues
- ▶ Surveying the public data landscape

When you need data that you don't already own, look for public sources first. Not only are these sources numerous and diverse, but in many cases, no commercial entity would be able to independently gather the same information. And public data is usually available free or at low cost.

Looking Over the Lay of the Land

Public data is primarily government data. Government agencies collect and share data about people, activities, and resources. In the U.S., you have the right to request information from any part of the federal government (though not all parts are equally responsive). Every state, county, and city collects and maintains data that may be available to you. Countries around the world have their own statistical agencies, as do a number of intergovernmental organizations.

Public data is a by-product of everyday government work; no government collects data for the purpose of sharing it with data miners. Obtaining relevant government data in a form you can use isn't always easy. You can develop a good sense of what to expect if you make an effort to understand why and how governments collect and share data:

- ✓ **Why governments collect data:** Every business keeps records of its activity, such as contracts, purchases, and sales; payments to employees and suppliers; and interactions with customers. These records are needed to support everyday business, because what we do today depends on what we did and agreed on yesterday, and keeping records ensures that we will have clear information about past activity when we need it. Governments have the same needs, so governments also keep records of their activities. In fact, record-keeping may be even more important in government than in industry, because every constituent has an interest in what the government does (or does not do). This type of data is also known as transactional data, and most government data is of this type.

Data miners (and other sorts of data analysts) are often more interested in data about people: who they are, what they do, and how they live. With that kind of data, you can discover behavior patterns and other aspects of human life that are relevant to your own business goals. Governments collect that kind of data too, because they also need to know about their constituents, how they live, and their needs for government services.

Perhaps you recall filling out a census survey form in the past. Your responses become government data. The census is only one of many surveys that the government uses to obtain data for analysis. This type of data is also known as statistical data. The purpose of these surveys, and other government research, is to provide data that government staff can analyze to provide information for lawmakers and other officials. Statistical data is a relatively small part of all government data, but it's important.

- **Why governments share data:** Governments sometimes share data to get a job done. If an initiative exists to encourage exercise, discussing survey data about current exercise patterns might help. They also share data to persuade. If you want funding to build a bridge, that case is more persuasive if you share data that indicates a need for a bridge, competitive costs for construction, and so on. And finally, governments share data because they must. Constituents expect it, and the law requires it.

  Not every bit of government data is available to the public. Some information is protected for security reasons, and some is kept secret to protect the privacy of citizens. When you respond to the census, your individual responses are not shared, but *aggregates,* information about groups of people, become public. (So, an individual's income is private, but the average income for a community is public.)

Exploring Public Data Sources

Public data resources are vast, but they have been tailored for specific government purposes, not your needs. You face certain challenges when you look to public sources for the data you need, so be prepared. Before you begin your search, make sure you first

- **Define your needs:** You'll need to know specifically what data you need, and in what form. You must be prepared to explain this, in writing or verbally, in the course of your search.
- **Find the right source:** You'll familiarize yourself with government agencies and find out which ones are likely sources for the data that you require.
- **Know how to obtain the data:** Sometimes getting the data you need will be simple; you'll just download it or obtain a report. But if the data you need is not already distributed through such simple channels, getting it can involve a complex and slow process.

Hundreds of governments and quasi-governmental organizations around the world collect, analyze, and share data. It may be shared as machine-readable data in files or via an application programming interface (API); it may be found in written reports, complete with sophisticated analysis; and it may even show up in small bits found in news reports.

Don't think of data only as fields and cases that you can work with in your data analysis software. Any form of data, raw or analyzed, may be useful to help you improve your understanding of the business issues that you face. You may even find that some agency or organization has already explored certain issues that affect you, collected data, analyzed it, and made a report available. If so, take advantage of what's already been done, discover, and move on to the next stage.

Data or statistics?

In this book, and in everyday situations, I talk about getting and using data. But, strictly speaking, that's not always the proper term. Often, what I'm really using are *statistics.* Most of the time, you can use any term you like and it won't matter, but at times, you'll need to know the difference.

Perhaps you measure the height of each child in a school. The individual measurements are data. And because they are just what you've measured and haven't been processed in any way, those measurements may also be called *raw data.*

You might calculate the average height of all the students in a classroom, or all students in the school; those averages (calculated values) are statistics. But people still call them data. And most of the time, that's not going to cause any confusion.

But maybe you're looking over a government agency website. You may see that the agency shares statistics. But it doesn't seem to share data. Or you may get into a conversation with an agency staff member and find yourself in a quibble over whether the agency releases data. The agency's staff may be much stricter than you in its choice of words. So remember that the proper term for what you want may be statistics, not data.

United States federal government

The U.S. government includes over 100 statistical agencies, agencies with a primary purpose of collecting and analyzing data for some government use. The result is a vast resource of professionally collected, managed, and analyzed data, much of which is available to you.

This section explains the purpose and data offerings of some major federal statistical agencies. It will also introduce you to Data.gov, a portal that helps you locate government data sources to match your needs.

Dozens of additional statistical agencies are not described individually here. You can find a list of them, with links and descriptions of the agencies and the kinds of data provided, at FedStats.gov, the portal to federal statistical agencies. FedStats (`www.fedstats.gov`) lets you find an agency by name or subject, find information by geography, and even find links to kids' pages on agency websites.

The federal data portal: Data.gov

If you're looking for data that the federal government might have, but you aren't sure which agency is involved, start your search on the federal data portal `www.data.gov`. There you will find a searchable catalog of data from all federal agencies. You can search for datasets by keywords and get information about what's available, the source for each dataset, the formats available, and where to find the data.

The data portal isn't a source for data, just information about what data is available and where to get it. And the portal doesn't cover every bit of government data available. So, if you find something that's useful to you on Data.gov, follow up by investigating the website of the agency that actually provides that data to search for additional information and data. If you need something you can't find, contact the agency directly. You may be able to speak with someone who can help you locate what you need, or at least find out why the data you want is unavailable.

While nothing is new about public data, the portal facilitates certain new initiatives. All newly generated federal government data is required to be made publicly available in open, machine-readable formats, while maintaining privacy and security. The key concept here is machine readability, providing data in formats that are appropriate for computing use, especially use in developing applications.

WARNING!

Top-ten hits on America's data portal

Over 100,000 datasets are available through the federal data portal. Of course, some are more widely used than others. A look at a few of the most popular examples will give you a sense of the variety of information that's available to you.

Don't assume that the most popular sources on Data.gov are necessarily the most popular overall. Many people obtain data directly from the agencies that produce it, or indirectly through third parties, such as news reports and data vendors.

So what's hot? Here's a list of the ten most popular datasets on Data.gov (at the time this was written).

Name	***Source***	***Format***	***Description***
Climate Data Online (CDO)	National Oceanic and Atmospheric Administration, Department of Commerce	HTML	Provides access to climate data products through a simple, searchable online web-mapping service.
Consumer Complaint Database	Consumer Financial Protection Bureau	CSV, JSON, XML	Complaints received about financial products and services.
NOAA National Weather Service - National Mosaic of Weather Radar	National Oceanic and Atmospheric Administration, Department of Commerce	N/A	National Weather Service's radar imagery allows interactivity with the display.
Federal Student Loan Program Data	Federal Student Aid, Department of Education	XLS	Quarterly recipient and disbursement information for the Direct Loan and Federal Family Education Loan Programs by postsecondary school.

(continued)

(continued)

Name	*Source*	*Format*	*Description*
State Education Data Profiles	National Center for Education Statistics, Department of Education	XLS	Searchable information in elementary/secondary education, postsecondary education, and selected demographics for all states in the United States.
Social Media Monitoring Metrics	U.S. Department of Health & Human Services	CSV, JSON, XML, RDF, XLS, XLSX	Basic social media metrics aggregated on a weekly basis. Metrics include SAMHSA's Facebook fans, comments, likes, and posts and Twitter followers and mentions.
Food Access Research Atlas	Department of Agriculture	HTML, JSON, XLS	A spatial overview of food access indicators for low-income and other census tracts.
U.S. International Trade in Goods and Services	U.S. Census Bureau, Department of Commerce	XLS, TXT	Monthly U.S. trade data, including imports, exports, and balance of payments for goods and services.
Campus Security Data	Office of Postsecondary Education, Department of Education	CSV, XLS	Rapid customized reports for public inquiries relating to campus crime data.
State Dropout and Completion Data	National Center for Education Statistics, Department of Education	CSV/TXT, SAS, XLS	The number of dropouts from each grade 9–12 and the relevant event dropout rates.

Agencies are also required to

- **Create a single agency data inventory:** They must document and track data assets as they do equipment, furniture, and other assets.
- **Publish a public data listing:** The listing must be posted on the agency's web pages, including all data assets that are public or that could be made public.
- **Develop new public feedback mechanisms:** They must provide ways for the public to provide feedback related to data-sharing priorities.

The federal data portal also allows local governments to add their datasets to the portal's catalog. This is not mandatory and not many cities are ready to participate, but you may come across some local data in the catalog, and you can expect to see more in the future.

WARNING!

While this portal can lead you to a large and diverse range of data, none of it was created specifically for data-mining use. All of it was originally collected for government use; sharing with the public is secondary. Privacy and security requirements prevent some data from being made public, and some data can only be shared in aggregate form. (For example, an individual's income may be private, while the average income of a group of people is public.) And open data initiatives are driven by programmers, not data miners, so the data may not be organized or formatted as you prefer.

REMEMBER

The data portal is a starting point, not a final destination, in your search for data. Not all government datasets are included in the catalog, and some that are may not be tagged with the keywords that you choose for your search. But Data.gov can guide you to many useful datasets and provide leads to agencies that may have more to offer. You may even discover some unexpected gems to enhance your data-mining work.

Bureau of Economic Analysis

The Bureau of Economic Analysis (BEA) (`www.bea.gov`) is a part of the United States Department of Commerce. The Commerce Department's job is to "help make American businesses more innovative at home and more competitive abroad." The Department is made up of 12 agencies that deal with matters as diverse as weather, communications, and patents. It's the BEA's job to "promote a better understanding of the U.S. economy by providing the most timely, relevant, and accurate economic accounts data in an objective and cost-effective manner."

BEA gathers economic data, conducts research and analysis, and makes the results available to the public. It provides information on matters such as economic growth, relationships among industries, and the nation's position in the world economy. It produces information on a national, international, and regional basis, and also for specific industries.

Here are some of the widely used types of data available through BEA:

- Balance of payments
- Foreign direct investment
- Gross domestic product (GDP)
- Gross domestic product by state
- Industry data
- International trade
- National income and product accounts (NIPAs)
- Personal income
- Personal income by state
- Gross domestic product by metropolitan area
- Gross domestic product by industry
- Personal income by county and metropolitan area

Bureau of Justice Statistics

The Bureau of Justice Statistics (BJS) (`www.bjs.gov`) is part of the Office of Justice Programs in the U.S. Department of Justice. The Justice Department's job is to "enforce the law and defend the interests of the United States according to the law; to ensure public safety against threats foreign and domestic; to provide federal leadership in preventing and controlling crime; to seek just punishment for those guilty of unlawful behavior; and to ensure fair and impartial administration of justice for all Americans." That's a lot! To do all that, the Justice Department has dozens of agencies, with wide-ranging responsibilities including antitrust matters; alcohol, tobacco, and firearms; civil rights; tribal justice; and a whole lot more.

BJS collects, analyzes, and shares information on crime, criminals, and victims, as well as the operation of the justice system. It also provides technical and financial assistance to state governments to develop their criminal justice statistics, criminal history records, and information systems.

BJS is the key source for data about

- Crime and victims
- Drugs and crime
- Criminal offenders
- Courts and sentencing
- Corrections
- Expenditure and employment
- Criminal record systems
- Firearms and crime
- Law enforcement

Bureau of Labor Statistics

The Bureau of Labor Statistics (`www.bls.gov`) is part of the U.S. Department of Labor. The Department of Labor's job is to "foster, promote, and develop the welfare of the wage earners, job seekers, and retirees of the United States; improve working conditions; advance opportunities for profitable employment; and assure work-related benefits and rights." It has more than two dozen agencies, whose responsibilities cover a range of issues including wages and benefits, occupational safety, occupational training, disability, and many others.

BLS is responsible for measuring and tracking the labor market, price changes, and working conditions. It collects, analyzes, and shares information on these and related matters.

BLS provides data such as

- Compensation
- Consumer expenditures
- Consumer price index
- Contingent workers
- Displaced workers
- Employee benefits
- Employer-provided training
- Employment
- Employment cost trends
- Employment projections
- Foreign labor
- Import-export prices
- Industry employment
- Job injuries
- Labor force
- Locality pay

- Longitudinal surveys
- Occupational projections
- Producer price index
- Productivity
- Real earnings
- State and area employment
- Unemployment
- Union members
- Wages
- Weekly earnings

Bureau of Transportation Statistics

The Bureau of Transportation Statistics (BTS) (`www.rita.dot.gov/bts`) is a part of the Research and Innovative Technology Administration (RITA). RITA has four agencies that deal with matters of transportation issues pertaining to safety, intermodalism, cost-effective regulation, compliance, training, and research.

It's BTS's job to "to create, manage, and share transportation statistical knowledge with public and private transportation communities and the Nation," to help advance the strategic goals of the Department of Transportation.

BTS shares transportation-related data on topics such as

- Airlines: On-time performance and financials
- Economics and finance
- Commodity Flow Survey
- Freight
- Household travel
- International travel and transportation
- Transportation snapshot
- Publication: National Transportation Statistics
- Ferry operators

Census Bureau

The United States Census Bureau (`www.census.gov`) is a part of the United States Department of Commerce. The Census Bureau's job is to "serve as the leading source of quality data about the nation's people and economy." This may sound something like the role of BEA, but while BEA focuses on whole industries and regional economies, the Census Bureau focuses on the characteristics and well-being of people and businesses.

If you use, or are even aware of, any government data, it's probably data from the Census Bureau. This is the agency that reports on how many Americans exist, who we are, and where and how we live. It tells us about the number and health of businesses. It tells us what's being built and what's being made in the United States.

The Census Bureau provides information on matters including

- Business ownership
- Construction
- Governments
- International trade
- Income and poverty
- Manufacturing
- Population estimates
- Population projections
- Social and economic characteristics
- Retail and wholesale trade

Economic Research Service

The Economic Research Service (ERS) (`www.ers.usda.gov`) is a part of the United States Department of Agriculture (USDA). The USDA's job is to "provide leadership on food, agriculture, natural resources, rural development, nutrition, and related issues based on sound public policy, the best available science, and efficient management." It is made up of more than 20 agencies and offices that deal with matters such as marketing of U.S. agricultural products, ensuring the health and care of animals and plants, and agricultural policy.

ERS "communicates research results and socioeconomic indicators via briefings, analyses for policymakers and their staffs, market analysis updates, and major reports."

Examining the American Community Survey

The Census Bureau's American Community Survey is one of the most widely used sources of public data. If you hear a news report that mentions demographics of your local region, that data came from the American Community Survey. If your congressman posts district facts and figures on his website, the data came from the American Community Survey. If a commercial source offers you data with income estimates and related information about individuals, it's based on data from the American Community Survey. Data miners, especially those involved in marketing and social sciences, depend on the American Community Survey every day, yet most are unaware of the data's origins.

The American Community Survey (ACS) is an annual survey conducted by the U.S. Census Bureau. The primary purpose of the survey is to provide communities with information they need to plan services and investments. Community governments need to know about the number of people in the community, who they are, and what they need, so each year, a sample of Americans are asked about themselves, their families, and how they live. The survey includes questions about the respondents' age, sex, and race; their family, education, income, and benefits; getting to work; veteran status; disabilities; and cost of living.

Government decisions at every level are made based on information obtained through the American Community Survey. And because data from the survey is available to the public, it is also the basis of business decisions. (Privacy concerns prevent sharing information about individual people, so the shared data is always aggregated. You can't get the survey data of an individual, but you can get information about groups based on geography or other factors.) This data is so widely used, analyzed, and integrated into reports and other information sources that users often are not aware of the primary source.

The American Community Survey enhances general census data (the survey that reaches out to every American once every ten years) to provide greater depth and more timely information that is vital to support government and business decisions.

With so many users and decisions depending on this survey, you'd think its future was assured, but it isn't. Open data initiatives oblige agencies to share the data they have in certain ways, but not to collect data. The budgets of statistical agencies are at risk. And the American Community Survey has opponents. Some lawmakers oppose the survey, citing budget and privacy concerns. (One public data user told a story of reaching out to a congressman who opposed ACS, and noticed that he had data from ACS posted on his website. Apparently, the congressman was unaware that even he depended on ACS data to do his work.)

Data made available through ERS includes

- Agribusiness/industry concentration
- Biotechnology
- Chemicals and production technology
- Crops
- Diet, consumption, and health
- Farm financial and risk management
- Farm structure, income, and performance
- Farm/rural finance and tax
- Food and nutrition assistance programs
- Food market structures
- Food prices, spreads, and margins
- Food safety
- International agriculture
- Livestock, dairy, poultry, aquaculture
- Macroeconomics in the agricultural and food economy
- Natural resources, environment, and conservation
- Policy topics
- R&D and productivity
- Rural America
- Trade
- U.S. state fact sheets with information on population, income, education, employment, federal funds, organic agriculture, farm characteristics, farm financial indicators, top commodities, and exports

Energy Information Administration

The Energy Information Administration (`www.eia.gov`) is a part of the United States Department of Energy (DOE). The DOE's job is to "ensure America's security and prosperity by addressing its energy, environmental, and nuclear challenges through transformative science and technology solutions."

EIA's job is to collect, analyze, and share "independent and impartial energy information to promote sound policymaking, efficient markets, and public understanding of energy and its interaction with the economy and the environment." By law, the data and analysis produced by EIA is independent; it's not subject to approval by any other employee or officer of the government.

EIA provides data on topics like these:

- Coal
- Coal supply and disposition
- Commercial
- Consumption
- Diesel prices
- Electricity
- Energy, forecasts
- Energy, statistical overview
- Environmental data
- Forecasts
- Gasoline prices
- Fuel economy
- Natural gas
- Nuclear
- Oxygenates
- Petroleum
- Power plants
- Prices, monthly all sources
- Refinery capacity
- Renewables
- Residential
- State energy profiles

Environmental Protection Agency

The Environmental Protection Agency (EPA) (www.epa.gov) is an independent agency whose job is to protect human health and the environment.

EPA provides data on environmental pollution. EPA's Envirofacts online database (www.epa.gov/enviro) is a central starting point for EPA data.

Office of Research, Analysis and Statistics

The Office of Research, Analysis and Statistics (RAS) (www.irs.gov/uac/Tax-Stats-2) is a part of the Internal Revenue Service (IRS). The IRS is "the nation's tax collection agency."

The job of the RAS is to provide "leading research, analytical, and technology services" to support the IRS.

RAS provides data on taxation, such as

- Corporation tax
- Estate and gift taxes
- Excise taxes
- Exempt organizations and bond tax

- Individual tax
- International or foreign-related tax
- Partnership economic data
- Sole proprietorship tax

National Agricultural Statistics Service

The National Agricultural Statistics Service (NASS) (`www.nass.usda.gov`) is a part of the United States Department of Agriculture (USDA). The NASS's job is to provide "timely, accurate, and useful statistics in service to U.S. agriculture."

NASS offers data on an extensive range of agricultural matters, including

- Crop progress and condition
- Dairy production
- Field crops
- Fruits, nuts, and vegetables
- Poultry production

National Center for Education Statistics

The National Center for Education Statistics (NCES) (`http://nces.ed.gov`) is a part of the United States Department of Education (ED). It's ED's job to "promote student achievement and preparation for global competitiveness by fostering educational excellence and ensuring equal access."

NCES collects, analyzes, and shares data about education in the United States and around the world.

NCES provides data on educational matters such as

- Early childhood education
- Education assessment
- Elementary and secondary education
- Finance
- International comparisons
- Libraries
- Other education subjects
- Postsecondary education

National Center for Health Statistics

The National Center for Health Statistics (NCHS) (www.cdc.gov/nchs) is a part of the Centers for Disease Control and Prevention (CDC). It's CDC's job to "protect America from health, safety, and security threats, both foreign and in the U.S."

NCHS's job is to "provide statistical information that will guide actions and policies to improve the health of the American people."

NCHS provides data on health-related topics, such as

- Asthma
- Births/Natality
- Child and infant health
- Deaths/Mortality
- Diabetes
- Disabilities/Impairments
- Divorces
- Health insurance coverage
- Heart disease
- Home health/Hospice care
- Hospital utilization
- Hypertension
- Influenza
- Leading causes of death
- Life expectancy
- Mammography/Breast cancer
- Marriages
- Men's health
- Nursing home care
- Occupational health
- Overweight prevalence
- Prenatal care
- Prescription drugs
- Sexually transmitted diseases
- Smoking
- Teen pregnancy
- Women's health

National Science Foundation, Science Resources Statistics

The National Science Foundation (www.nsf.gov/statistics) is charged with promoting the progress of science, advancing the national health, prosperity, and welfare, and securing national defense. The National Science Foundation provides data on topics such as

- Science and engineering education
- Science and engineering workforce
- Research and development

Office of Management and Budget

The Office of Management and Budget (OMB) (`www.whitehouse.gov/omb`) is part of the Executive Office of the President of the United States.

It's OMB's job to "serve the President of the United States in implementing his vision across the Executive Branch." This is the agency that manages the nation's budget and oversees other federal government agencies.

OMB shares the current and historical information about

- Budget of the U.S. government
- Fact sheets on government and social issues

Office of Retirement and Disability Policy

The Office of Retirement and Disability Policy (`www.ssa.gov/policy`) is a part of the Social Security Administration (SSA). It's SSA's job to administer retirement and disability benefits for Americans. It is the principal advisor to the Commissioner of Social Security on major policy.

Research and policy analysis for SSA is done by three parts of ORDP: the Office of Research, Demonstration, and Employment Support; the Office of Research, Evaluation, and Statistics; and the Office of Retirement Policy.

ORDP provides data on matters such as

- Income of the aged
- Social Security (Old-Age, Survivors, and Disability Insurance, OASDI) beneficiaries and benefits
- Supplemental Security Income (SSI) beneficiaries and benefits
- Trends in Social Security and disability programs
- Income of the Population 55 or Older
- Workers covered under Social Security and Medicare
- Congressional statistics

Governments around the world

The United States is only one of many governments that share data with the public. While you won't find exactly the same range or types of data from every country, you will find that most nations have some data to share. This section also includes some intergovernmental and nonprofit organizations that offer international data resources.

OFFSTATS

University of Auckland's OFFSTATS database (www.offstats.auckland.ac.nz) is a portal to statistical agency sources around the world, like an international version of the United States' FedStats portal. It has links organized by country, region, and subject. (International agencies do business in the local languages, and many don't have English-language versions.)

Organisation for Economic Co-operation and Development

The Organisation for Economic Co-operation and Development (OECD) (accessible through the statistics portal at www.oecd.org/statistics) aims to promote policies to improve the well-being of the world's people. The OECD measures productivity, global trade, and investment. It analyzes data on trade and everyday life.

The OECD also offers resources targeted for use by statisticians at www.oecd.org/statistics/statisticalresources.htm. These are also valuable for data miners. And the OECD has a portal to data sources around the world at http://stats.oecd.org/source.

U.S. open government portal

You can find lists of open data portals for international sources (as well as U.S. states, counties, and cities) on Data.gov at www.data.gov/open-gov.

United Nations

The United Nations (UN) is the world's most influential intergovernmental organization. The UN offers a portal to its statistical sources at http://data.un.org.

European Union

The European Union, which includes most of the Western European nations, has a portal to its statistical sources at http://europa.eu/publications/statistics/index_en.htm.

TIP

The Open Data Institute (http://theodi.org) promotes sharing and use of open data around the world. It's a key source for news on open data, as well as a center for research and education.

United States state and local governments

Finding the data you need from state and local governments can be very challenging. Some states are more interested in sharing data than others. You can't count on every state or local government to have an open data portal, or on finding someone in the local government to help you find what you need or address your questions.

U.S. states

Start your search for state-specific data the easy way with an online search for your state's data portal. If that turns up nothing, check the federal data portal (`www.data.gov/states/page/states-data`) and look for a link for your state.

If you find no sign of a portal, that doesn't mean that your state doesn't have data to share. You'll just have to work harder to find what you need. Check the state's website for information about your state agencies. If you don't find what you need online, start calling agencies by phone. Be prepared to explain what you need.

Your state may have a librarian who can help you understand how to locate information. Librarians at your local library may also be able to advise you on navigating government information sources. Be polite and persistent; state agencies are not always responsive about data requests.

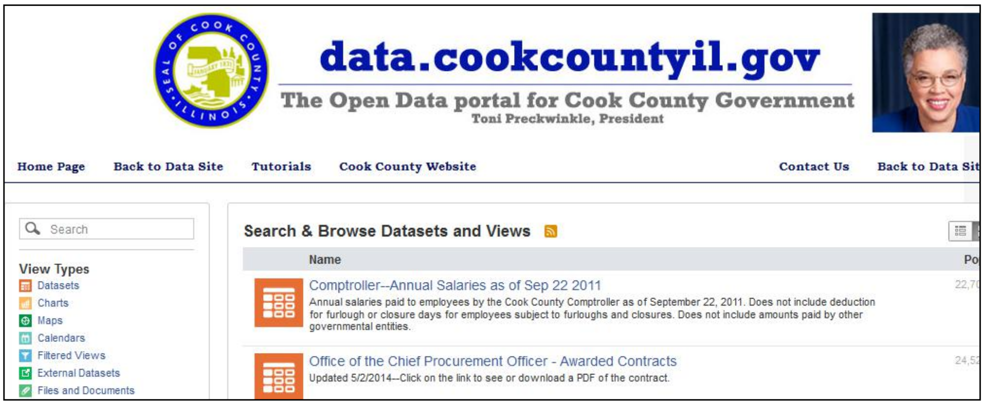
Every U.S. state, as well as the District of Columbia, has an open records law similar to the federal Freedom of Information Act. If you're having trouble getting data that you know (or have good reason to believe) exists, you can request the information through the rights guaranteed to you by these laws.

Pew Charitable Trusts

The Pew Charitable Trusts (`www.pewstates.org/states`) is a good nonprofit source for research and other information about U.S. states. It conducts research and reporting across all states and the District of Columbia.

U.S. counties

Most counties don't yet have centralized open data portals, but some, like Illinois' Cook County (see Figure 10-1), do. Try an online search for one when you start looking for data. The next place to check is the county open data portal list on Data.gov at `www.data.gov/counties`.



data.cookcountyil.gov

The Open Data portal for Cook County Government

Toni Preckwinkle, President

Search & Browse Datasets and Views

Comptroller--Annual Salaries as of Sep 22 2011 22,70

Annual salaries paid to employees by the Cook County Comptroller as of September 22, 2011. Does not include deduction for furlough or closure days for employees subject to furloughs and closures. Does not include amounts paid by other governmental entities.

Office of the Chief Procurement Officer - Awarded Contracts 24,52

Updated 5/2/2014--Click on the link to see or download a PDF of the contract.

Figure 10-1: Data portal for Cook County, Illinois (United States)

Chances are, your county won't make it easy for you, so it may take some time to locate the data you seek. Call county offices and explain your needs. Don't be surprised if you have trouble reaching someone who seems to understand your request, but keep asking around. Try speaking with a local librarian for advice on obtaining local government data.

TIP

If all else fails, talk with the staff at the office of your local government representatives (local to the source of the data you need). Many of them deal with similar challenges obtaining data all the time and can offer advice or other help with the process. Remember that these offices exist to serve constituents, so if you don't live in the area, mention a little about how your work helps the locals. If your work might lead to new businesses, employment, or any economic benefit for the area, be sure to say so!

U.S. cities

Many cities are establishing open data portals now. You may find yours easily in an online search, especially if you are interested in a large city. You can also find a list of city portals on Data.gov at `www.data.gov/cities`.

Here are some of the established big-city data portals:

- **Chicago:** `https://data.cityofchicago.org`
- **New York:** `https://nycopendata.socrata.com`
- **Boston:** `https://data.cityofboston.gov`
- **Seattle:** `https://data.seattle.gov`
- **San Francisco:** `https://data.sfgov.org`

WARNING!

County and city governments collect a lot of transactional data — records about government activity such as building permits, property transfers, licenses, and tax payments. But they don't usually gather and share much information about people and how they live. If you are interested in demographics, cost of living, lifestyle, and so on, you may get more relevant data from the federal government or a commercial data supplier.

Getting user support for public data sources: An interview with Becky Sweger

Becky Sweger uses public data every day. She's director of Data and Technology at National Priorities Project (NPP), a nonprofit, nonpartisan research organization dedicated to making complex federal budget information transparent and accessible. NPP aims to help people prioritize and influence how their tax dollars are spent. It's Becky's responsibility to create research products that put the federal budget in context for partners, the media, and novice budget users.

Q: Tell us about a situation where you weren't able to get the information you needed. What you were trying to accomplish?

Sweger: I was trying to get a state-by-state breakdown of federal spending on a handful of programs — specifically, the Children's Health Insurance Program (CHIP) and Low Income Home Energy Assistance Program (LIHEAP). Both are administered by Health and Human Services.

Q: What agency was involved?

Sweger: Each federal agency is responsible for supplying information to USASpending.gov. I was working with data specific to Health and Human Services (HHS).

Q: What problem led you to ask for help? What kind of help did you need?

Sweger: For both CHIP and LIHEAP, I found records that were missing a field called "Place of Performance State." This field drives the maps displayed on USASpending.gov. As you might expect, it's a required field according to the data dictionary and data requirements published by the Office of Management and Budget (OMB).

Q: What kind of response did you get?

Sweger: The USASpending.gov help desk personnel did answer my email for this issue, saying that while they're not responsible for data issues that originate from the agencies, they would try to find out why the HHS data was able to get into the database without having all the required fields. Here's the answer, directly from the email sent by the USASpending.gov help desk:

"Health and Human Services is one of the agencies that does not require a Place of Performance Code. We are unaware of the policy behind this. You will need to contact someone from HHS for the policy explanation or to ask that they correct any existing awards that you believe should have a code. We do not have an HHS point of contact for you."

Q: So, you never got the data or a satisfactory explanation.

Sweger: It's clear that neither OMB nor some of the agencies have made it a priority to provide complete, accurate data to USASpending.gov, despite the 2006 law that requires them to do so, the Federal Funding Accountability and Transparency Act (FFATA). Ultimately, this is a people/process problem, not a data problem — I've found that to be the case most of the time.

Q: But you've also had good experiences. Tell us about one of those.

Sweger: Yes. Here's an example: I was trying to find the best source of state-level poverty data.

(continued)

(continued)

I was looking at several United States Census Bureau products that measure poverty (there are about six ways to get the number).

Q: What problem led you to ask for help? What kind of help did you need?

Sweger: As a novice user of this data, I needed help understanding the differences between the various poverty numbers to make sure that I chose the right one for our purposes.

Q: How did it go?

Sweger: The Census Bureau provides a description of each product and when it should be used. I learned that the poverty measure I'd been looking at was not, in fact, appropriate for doing year-over-year comparisons at the state level.

And there's a bonus from Bureau employees: If you still have questions after reading their documentation, the Census Bureau is happy to talk with you on the phone. They understand that the job entails more than just publishing data — it's also their job to make sure that people can use it appropriately.

Q: Any final advice for newbie public data users?

Sweger: Data published by entities that are solely in the data business (for example, Census, Bureau of Labor Statistics, or Bureau of Economic Analysis) is always accompanied by documentation. Read it. This will not only make you smarter than most of the other people who use that data, but it will also help you be smarter about using "exhaust" data (in other words, data that's generated as a by-product of doing business, like the stats published by the IRS). Entities that publish exhaust data aren't necessarily in the business of creating information that can be used for rigorous analysis.

When assessing a source of public data, find out who else uses it and for what. Don't be afraid to reach out — most people are happy to discuss their work.