

Dataset Explainers

A Guide for Users of the US City Open Data Census

Dataset Explainers

- [Asset Disclosure](#)
- [Budget](#)
- [Business Listings](#)
- [Campaign Finance Contributions](#)
- [Code Enforcement Violations](#)
- [Construction Permits](#)
- [Crime](#)
- [Lobbyist Activity](#)
- [Parcels](#)
- [Procurement Contracts](#)
- [Property Assessment](#)
- [Property Deeds](#)
- [Public Buildings](#)
- [Restaurant Inspections](#)
- [Service Requests \(311\)](#)
- [Spending](#)
- [Transit](#)
- [Web Analytics](#)
- [Zoning \(GIS\)](#)

Notes added:

- *Also known as:*
- *Searching tips:*
- *Special considerations:*
- *Relevant standards or policy:*
- *Example dataset:*
- *Example data use:*
- *Further reading:*

Asset Disclosure: Asset disclosure data collection is mandated by state and local ethics commissions, boards, and code to make the financial background of top-level government officials transparent to the public. Open financial disclosure data is fundamental to hold government officials accountable and spot conflicts of interest. Asset disclosure data should include the following elements: name of top-level government officials, title, investment information, prior and current business relationships, real estate interests, and personal income (including gifts and travel or speaking payments).

- Also known as: *Asset Disclosure, Conflict of Interest Statements, Ethics disclosure, Ethics statements, Financial Disclosure, Financial Statements, Statements of Economic Interest.*
- Searching tips: If not on your local government website, it may be available on your state's Ethics Commission or Board website.
- Special considerations: Financial disclosure data should be released in real time to help empower public oversight of potent potential conflicts of interest as government decisions approach rather than after decisions have been made. Even monthly or quarterly updates to disclosure would be better than requiring only annual or semi-annual reports, which can grow out of date quickly as a person's assets change. [E-filing](#) would help ensure that more frequent reporting does not need to mean a repetitive process: filers could simply update existing forms with any relevant changes since the last reporting deadline.
- Example dataset: [Chicago](#)
- Example data use: There are examples from [Philadelphia](#), [Baltimore](#), and [Honolulu](#), in addition to [many other municipalities](#) that call out the importance of having timely access to this data.
- Further reading: The Sunlight Foundation's [page on Asset Disclosure and Transparency](#)

Budget: Open budget data is important to share to communicate how tax dollars are intended to be spent and to compare against policy plans and actual spending. For purposes of this Census, budget data should include the planned budget by unit of appropriation with a programmatic description of each unit of appropriation.

- Also known as: *Annual Budget, Budget and Fiscal Plan, Budget and Strategic Planning*

- Special considerations: Though often annual, the timing of planned budgets' fiscal years can vary by jurisdiction. Planned budgets should be posted online in open structured formats shortly after they are created.
- Relevant standards or policy: [Oakland's Open Budget Legislation](#)
- Example dataset: [Cook County](#) or [Austin](#)
- Example data use: [Open Budget Oakland](#)

Business Listings: A directory of all licensed businesses in the municipal area, including key information such as name, address, contact information, and business type.

- Example dataset: [San Francisco](#) or [South Bend](#)

Campaign Finance Contributions: Campaign finance disclosure is an essential public accountability mechanism. Providing access to open campaign finance data about the flow of money in politics empowers oversight and accountability in the government decision-making process. Campaign finance contribution data, for US City Open Data Census purposes, should be a structured dataset that details the amount contributed to each candidate and by whom.

- Searching tips: If not on your local government website, campaign finance data contributions may be available on your state's Ethics Commission or Board website, or on the websites of the state's [election commission](#) or Secretary of State.
- Special considerations: Public oversight of campaign finance activity would be further improved by releasing information in real time. Releasing campaign finance information every quarter can leave long gaps of time in which the public does not know which interests are working to influence candidates or elected officials. Even monthly disclosure would be better than requiring only quarterly or semi-annual reports as a way of keeping the public informed.
- Relevant standards or policy: DC's "[Campaign Finance Reform and Transparency Amendment Act of 2013](#)"
- Example dataset: [New York City](#)
- Example data use: Sharing complete and timely campaign finance data [empowers a range of reporting and apps](#) that can reveal the flow of money in politics.
- Further reading: The Sunlight Foundation's [page on municipal Campaign Finance and Transparency](#)

Code Enforcement Violations: Building code inspection data surfacing reports on particular properties from code enforcement officials.

- Relevant policy or standard: [HouseFacts](#)

- Example dataset: [San Francisco](#)
- Example data use: Read about how New York City is “[using data to fight fires](#),” or about [Chicago's Edifice app](#) that combines datasets to present information about Chicago buildings.

Construction Permits: Locations of issued construction permits.

- Also known as: *Building permits*
- Example dataset: [Louisville](#) (with information on [all permits](#) and [active permits](#), and a relevant [map](#))
- Example data use: Seattle’s interactive [Shaping Seattle](#) application, and [Chicago's Edifice](#) app.

Crime: Open crime data helps communities make informed judgements about public safety, as well as providing transparency into how local police power is being exercised. For US City Open Data Census purposes, crime report data should include at a minimum the following elements: date, time, location, incident type, and narrative information—best would be exact date, location, and type of crime, but per day per street or postal/zip code would be acceptable.

- Also known as: *Crime Reports, Incident Reports, Incident Tracking*
- Searching tips: Crime report data may be on your city’s GIS or Police website.
- Special considerations: Technology has enabled faster and more continual, real-time sharing of this vital information, and governments should use that opportunity to bolster up-to-date public access.
- Relevant standards or policy: [SpotCrime Working Standard](#)
- Example dataset: [Dallas](#) or [Philadelphia](#)
- Example data use: Knowing what crimes happened and where they happened, at a minimum, can help fuel [news stories and apps](#) that keep people updated on important public safety issues. Some cities, like [Chicago](#), now have apps that help track crimes over time and by area.
- Further reading: The Sunlight Foundation’s [page on local Crime and Transparency issues](#) and the Sunlight Foundation’s “[Hall of Justice](#)” [project](#). You can also read Sunlight’s analysis of the [need for more or better police data](#) and of how data could [improve policing](#) and [police-community relations](#); in addition, Sunlight’s [review of stop and frisk data](#) helps show how the ways data are collected and released can make a big difference for accountability. You can also look [here](#) for more general

information on criminal justice transparency, including on respecting privacy in crime data releases.

Lobbyist Activity: Lobbying might evoke images of power brokers in Washington, DC, but lobbyists are also influential at the local level. A complete suite of lobbying data should include information about registration and activity-- who is paying lobbyists and, ideally, how much and where money is spent.

- Also known as: *Registered Lobbyists / Lobbyist Registration, Lobbyist Activity Report, Lobbyist Search*
- Searching tips: If not on your local government website, it may be available on your state's Ethics Commission or Board website.
- Special considerations: Details about lobbying activity should be released as close to real time as possible. Releasing lobbying information every year or even twice per year leaves long gaps of time in which the public does not know which interests are working to influence their government. Even quarterly disclosure would be better than requiring only semi-annual or annual reports as a way of keeping the public informed.
- Example dataset: [New York City](#) or [Albuquerque](#) (data exists online, but isn't machine-readable); [Chicago](#) (data is comprehensive and machine-readable, and can be downloaded in bulk)
- Example data use: This data can help create apps that make the information easier to sift through and news stories that investigate how lobbying impacts the political process, [among other impacts](#). Chicago has a good example of an [app using open data](#) to illustrate local lobbying.
- Further reading: The Sunlight Foundation's [page on Lobbying and Transparency](#) or the [Municipal Lobbying Data Guidebook](#). You can also read through Sunlight blog posts on [lobbying in general](#) or [local lobbying in particular](#). If you're interested in more details about how lobbying works, you can look at Sunlight's federal lobbying trackers, like [Influence Explorer](#) and [Political Party Time](#).

Parcels: Parcel data is data on the geographic boundaries of property. Parcels are the most specific units of geodata that governments maintain. The data is mainly used by the County Assessor's office to assess property taxes, yet is also used to keep track of addresses, other type of taxes, and zoning information. Because of their granularity, parcel data can be used for very detailed maps, visualizations, and all kinds of applications. For example, Code for America

has used parcel data for work on its [House Facts](#) data standard and to find [earthquake-prone buildings](#) in Oakland.

- Also known as: *Land Base, Legal Lots, Land Tracts, Property Boundaries, Geolocation, Property Information*
- Searching tips: If parcel data isn't on an open data portal, check your county's GIS page, or County Assessor's page. Be sure to check if your county charges money for this data. Recent laws have been pushing to make parcel data freely available, yet many counties charge a premium for access.
- Special considerations: Appropriate file types include shapefile, KML, and GeoJSON; PDF maps are not useful for apps or automated processing. Also make sure that any parcel data you find has a unique ID (usually called an access point name or APN) and detailed geographic boundary coordinates.
- Relevant standards or policy: See Code for America's [House Facts specification for parcels](#).
- Example dataset: [Alameda County](#) and [Oakland](#) provide access to the same parcels dataset. The [Los Angeles County Assessor](#) sells parcel information; by contrast, the [city of Los Angeles](#) provides city parcel data for free.
- Example data use: AxisPhilly [created maps of Philadelphia](#) identifying factors like property taxes. Motor City Mapping [created interactive maps](#) to identify and respond to blight in Detroit. LOVELAND also compiled [parcel and other property data from every state](#).

Procurement Contracts: The full text of municipal contracts with vendors, including the amount, awardee (name, address), and date awarded. The data should include if there have been changes to a contract, with information on the final payments as well as the initial contract award amount. Contracts should also be categorized according to the type of purchase and the purchasing department or agency.

- Also known as: *Purchasing, Purchase Orders, Contracting, Vendor Services, Request for Proposals*
- Relevant standards or policy: the Open Contracting Partnership's [Open Contracting Data Standard](#)
- Example dataset: [Miami](#) (data exists online, but not machine-readable or available in bulk), or [Baton Rouge](#) (data exists online, and is comprehensive, machine-readable, and available in bulk)

- Example data use: Information about the kinds of city purchases could be used to see where city contracts could be merged to get better pricing, where there is more or less competition among vendors (and which vendors have provided less reliable goods or services), and how (or why) the city has spent certain funds in the past, which could then encourage other vendors to place bids for city contracts. The [Open Contracting Data Standard](#) goes further in-depth on how open contract data can help and improve government. Bad or corrupt contracting has created serious problems for city governments, including (not too long ago) in [Chicago](#) and [Washington, DC](#).
- Further reading: The Sunlight Foundation's [Open Procurement Data Guidelines](#), or Code for America and Sunlight's analysis of "[The State of Local Procurement](#)." For a very comprehensive overview to open contract data, read the Open Contracting Partnership's online [Open Contracting Guide](#) book, as well as its [map](#) of contract data worldwide.

Property Assessment: Data about assessed property values.

- Also known as: *Property Value, Property Appraisal, Real Estate Assessment, Assessor's Information*; may also be combined with parcel data ([see above](#)).
- Searching tips: Property Assessment values and shapefiles may be on your city's or county's Planning, Finance, GIS, or Property Assessment website(s).
- Example dataset: [New York City's PLUTO](#) data
- Example data use: Philadelphia released an online property tax calculator and AxisPhilly, though no longer active, created maps of city property values, leading to [debates and changes to city property taxes](#). LOVELAND also compiled [parcel and other property data from every state](#).
- Further reading: If you're interested in how land-related data compares around the world, look at Cadasta's review of [challenges and effects of opening land data](#).

Property Deeds: The recording of property sales, mortgages, and foreclosures.

- Searching tips: See your local Registry/Recorder of Deeds.
- Example dataset: [New York City](#)
- Further reading: If you're interested in how land-related data compares around the world, look at Cadasta's review of [challenges and effects of opening land data](#).

Public Buildings: Locations of city-owned buildings.

- Also known as: *City (Owned) Property, Public Facilities*
- Searching tips: See also your local Registry/Recorder of Deeds.

- Example dataset: [Boston](#) or [Baton Rouge](#)
- Example data use: [OpportunitySpace](#) has helped governments sell or redevelop government-owned properties.
- Further reading: [Transparency in property held by governments](#).

Restaurant Inspections: Outcomes of food safety inspections of restaurants and other similar providers of food to the public.

- Also known as: *Health Department inspections, Food Establishment inspections, Food Safety and Sanitation, Food Facility inspections*
- Relevant standard or policy: [Local Inspector Value-entry Specification](#) (LIVES)
- Example dataset: [San Francisco](#), [Chicago](#), or [King County](#)
- Example data use: When the City of Los Angeles required that restaurants display hygiene grade cards on their entrances, studies found it led to a [13 percent decrease in hospitalizations](#) due to foodborne illness. The Colorado Springs *Gazette* compiled restaurant inspection data to create an [interactive database](#); CU-CitizenAccess put data on [failed inspections on a map](#).
- Further Reading: The [Transparency Policy Project](#) analyzes some of the background of releasing restaurant inspection reports, and some of the problems with inspections that greater transparency has helped reveal.

Service Requests (311): Non-emergency service requests (which some cities facilitate by dialing 3-1-1), covering issues such as graffiti, broken traffic lights, noise complaints, parking law enforcement, and potholes. Open service request data provides transparency about what types of requests are being requested and where, as well as how quickly requests are resolved. For Census purposes, service request data should at a minimum include: request type, responsible city department, location, and when the service was requested and fulfilled.

- Also known as: *311, Citizen Services, City Services, Non-emergency Service Requests, Service Requests*
- Searching tips: Service request data may be on your city's 311 or Services webpage.
- Special considerations: Many 311 systems provide a broad range of information and services, but for Census purposes, focus on service request data by citizens for non-emergency services.
- Relevant standards or policy: [Open311 GeoReport v2](#), [Open311 Inquiry v1](#)
- Example dataset: [San Francisco](#)

- Example data use: Open311 [Applications](#), in addition to Chicago's [Vacant and Abandoned Building Finder](#) (releasing 911 data has also had important effects, such as with [Los Angeles rescue response times](#))
- Further reading: [Open311.org](#). You can also [listen to a discussion](#) on the release of 311 data in Raleigh and Kansas City.

Spending: Open spending data informs citizens, journalists, and government officials alike as to just how public funds are being spent. Spending data should include a complete list of city expenditures at a detailed transactional level (including tax breaks, loans, contracts, grants, and operational spending). In other words, records should be fairly granular, covering, for instance, month to month expenditures on specific items costing in the thousands rather than the millions. This data category refers to detailed ongoing data on *actual* expenditures; a database of contracts awarded is *not* considered sufficient.

- Also known as: *Government Spending*, *Government Expenditure*, *Checkbook*, or *Government Spending on Goods and Services*
- Example dataset: [New York City](#) (with [downloads here](#)), or [Boston](#)

Transit: Open transit data provides transparency and legibility to when and where public transit is running. Timetables (schedules), locations of stops, and real-time location information of all municipally run or commissioned transit services (buses, subway, rail, tram, etc.) should be easily and openly accessible.

- Searching tips: Public transit is often managed by larger authorities, so you may have to look further than your city's website. Check out the American Public Transportation Association's [resources](#) for more information.
- Relevant policy or standard: [GTFS](#), [GTFS real time](#), [SIRI](#), [TCIP](#), [NextBus](#)
- Example dataset: [Hartford](#), [Asheville](#), or [New York](#)
- Further reading: ["Pioneering Open Data Standards: The GTFS Story"](#)

Web Analytics: Web analytics are one of the most useful datasets for understanding direct interaction between local governments and citizens. Data collected using programs like Google Analytics, KISSMetrics, or other similar programs can allow cities to make better decisions about web development priorities based on citizen needs. In the hands of civic hackers, web analytics can also be used to advocate for better accessibility for all residents and to demonstrate challenges that both the city and the civic hacking community should address. As an example, Code for America created a [real-time analytics viewer](#) (in alpha) based on GDS's analytics dashboard. While it requires access to real-time data, which many cities do not yet

release, it provides powerful information about actual usage for city staff who publish on the web.

- Also known as: *Search Terms, Page Visits, Google Analytics, Website Analytics, Website Visit Data, Site Statistics, Webstats*
- Relevant policy or standard: The federal Digital Analytics Program's [Digital Metrics Guidance](#) (in particular, see the section on "What to Collect")
- Example dataset: Oakland's [search terms](#) and [page visits](#), or Anchorage's very comprehensive [analytics page](#). The federal government has a thorough [analytics site](#) for executive branch pages, which [Philadelphia](#) also adapted for its own analytics site.
- Example data use: 18F's [short summary of federal web analytics data](#), and the White House blog's [summary of potential uses](#)
- Further reading: The federal Digital Analytics Program's [Digital Metrics Guidance](#) gives many details on how web analytics can be collected and analyzed. 18F also provides [information and code](#) for how the federal government's analytics site was built, and describes how ["even a junior developer"](#) can create a local web analytics site based on the federal version. You can read more about Philadelphia's implementation of an analytics site [in this article](#).

Zoning (GIS): Zoning impacts the most physical elements of our communities and daily lives — from the location of your local supermarket (and what you can purchase there) to the size and height of your home. Having access to the mapped zone (GIS) shapefiles of designated permitted land use where you live can help empower a better understanding of the forces shaping your community.

- Searching tips: Zoning shapefiles may be on your city's GIS or City Planning website.
- Special considerations: Zoning data should be released in a timely manner. Updating it as quickly as possible when changes are made gives people a chance to be aware of and react to modifications that might impact them.
- Relevant standards or policy: [Cook County's Open GIS Policy \(14-0076\) for free access to all GIS zoning](#); ["Let's Build an Open Zoning Data Standard"](#) and [follow up discussion](#)
- Example dataset: [Chicago](#) or [Anchorage](#)
- Example data use: [Apps using zoning data](#) in places like [Chicago](#) have shown how access to this data makes it easier to understand local land use laws.
- Further reading: [Sunlight Foundation's Zoning and Transparency](#)

Dataset Explainers updated July 5, 2016.