

Employer Data Matching Workgroup
Data Inventory

| Department/ Agency | Component Agency or Office | Dataset Name | Dataset Type: Owned by Data Inventory Respondent, or Desired by Data Inventory Respondent? | Data Update Frequency and Lag | Contact E-mail Address | Link to Data Collection Tool | Data.gov Link | Other Relevant Link(s) | Purpose of the Data Collection | Access Restrictions | Coverage | Unit of Analysis | Definition of "Company" | Corporate Structure and Relationships | Quality of Fields for Matching | Unique Identifier Fields | Name Fields | Address Fields | Country Fields | FIPS Code Fields | Industry Code Fields | Geocoding | Other Fields | Additional Notes |
|---|--|--|--|---|---------------------------------------|--|--|--|---|--|--|---|--|--|---|--|--|---|--|---|--|--|---|--|
| U.S. Department of Commerce | U.S. Census Bureau | Business Register | Owned | Continuously update frequency varies by source; lag vary by the reference period of the sources. | | The Business Register is updated using information from several federal agencies, including the IRS, the Bureau of Labor Statistics (BLS), the Social Security Administration (SSA) and the Census Bureau. There are data collection tools linked to the sources. For example, The Company Organization Survey (COS), also known as the Report of Organization Survey, is an annual survey performed specifically for the purpose of maintaining BIR information about multi-unit establishments' current organization and operating information. https://www.census.gov/econ/view/mud700.html http://www.reginfo.gov/public/display/PRAViewDoc?ref_nbr=201406-1220-001 | Not Applicable | www.census.gov/econ/view/mud600.html | To provide a current and comprehensive database of U.S. business establishments and companies for statistical program use. | The Business Register information is confidential (Title 13 and Title 26, US Code). Access is restricted to persons specially sworn to uphold the confidentiality provisions of Title 13 and Title 26. | Establishments of all domestic businesses (except agriculture, forestry, fishing, hunting, rail transportation, the U.S. Postal Service, elementary and secondary schools, colleges and universities, labor organizations, political organizations, religious organizations, public administration, and private households) and organizational units of multi-establishment businesses. The Business Register (BR) covers more than 160,000 multi-establishment companies, representing 1.8 million affiliated establishments, 5 million single establishment companies, and nearly 21 million non-employer businesses. | Establishment-based. | An establishment is a single physical location where business transactions take place and for which payroll and employment records are kept. Groups of one or more establishments under common ownership or control are enterprises. A single-unit enterprise owns or operates only one establishment. A multi-unit enterprise owns or operates two or more establishments. | Multi-unit enterprises may use multiple EINs, and they may use different ones when reporting to different federal agencies, complicating the matching process. The Census Bureau has linked many, but not all of the EINs to enterprises. | The Business Register is one of the most complete, current, and consistent source of establishment-based information about U.S. businesses, and is essential to assuring full coverage and high quality in Federal economic statistics programs. Examples of quality considerations for this source include: - The Annual Company Organization Survey covers 80 percent of multi-unit companies and a small sample of firms that were single-unit firms in the most recent quinquennial Economic Census, so establishment openings and closings in the firms not covered may. The business register is divided into employer and non-employer business registers based on payroll employment. Some firms lease their employees from Professional Employer Organizations (PEOs) or use independent contractors. Such firms may appear in the non-employer business register despite having large revenues and many leased and/or contract employees, not reflected in the business register until after the next Economic Census (though the Census Bureau takes measures to address this). | A single-unit enterprise's primary identifier is its Employer Identification Number (EIN). A unique employer unit identification number identifies each establishment owned by a multi-unit enterprise on the Business Register. (https://www.census.gov/econ/usa/ue/theology.html) | Yes; Primary and secondary business names, permitting identification of both legal and trade designations for businesses that support them (e.g., franchises); character | Yes; Physical location, mailing address, post office (city or town) name, postal state abbreviation, ZIP+4; character | This field is not currently populated in the BR, but it may be possible to link in this information from the Company Organization Survey | Yes; County and State; character | 6-digit NAICS codes; (https://www.census.gov/econ/usa/ue/theology.html) | Yes; latitude and longitude to six digits after the decimal; numeric | Census Tract (character), Census Block (character), 4-digit SIC codes (character) | Please also reference the working paper, "Documenting the Business Register and Related Economic Business Data" papers.som.com/usd/papers.cfm?tab=act_id=2755723 |
| U.S. Department of Labor | U.S. Bureau of Labor Statistics- Office of Employment and Unemployment Statistics (OEUS) | Quarterly Census of Employment and Wages (QCEW) Business Register (BR) | Owned | The QCEW BR is updated quarterly, and the data become available 6 months after the reference cycle. | QCEWInfo@bls.gov BDMInfo@bls.gov | www.reginfo.gov/public/display/PRAViewDoc?ref_nbr=201406-1220-001 | catalog.data.gov/datasets/quarterly-census-employment-and-wages | www.bls.gov/nom www.bls.gov/nom | To provide a quarterly census of all establishments under State unemployment insurance programs, representing about 98 percent of employment on nonfarm payrolls. This database of U.S. business establishments serves as the basis for multiple statistical programs; Sampling frame & benchmarks:ICES, JOLTS, PPI, CES, LAUS, SON, and NCS, which includes the EO, EBS and OHS; labor market research, Business Employment Dynamics (BED) data. | All microdata are confidential subject to BLS non-disclosure standards. | Employment, wages, and administrative data (name, location, etc.) for over 9.5 million establishments covering approximately 98% of all employment. | Establishment-based | An economic unit that produces goods or services, usually at a single physical location, and engages in one or predominantly one activity - Potential lay synonyms: business, worksite, brick& mortar, site, storefront. Lay users may use "establishment" interchangeably with "firm" in the QCEW BR, however, there is a significant distinction between the two terms. | Multi-unit enterprises may use multiple EINs, and they may use different ones when reporting to different Federal agencies, complicating the matching process. The dataset indicates whether an establishment is a single or multi-establishment. | New census.gov/econ/view/mud600.html | EIN | Yes; Legal Name and Trade Name; character | Yes; Physical location and mailing address; | Yes; County and State; character | 6-digit NAICS codes; | Yes; latitude and longitude to six digits after the decimal; numeric | See also: www.bls.gov/ophub/nom/pdf/nomch5.pdf | | |
| U.S. Department of Labor | Wage and Hour Division | Wage and Hour Investigative Support and Reporting Database (WHISARD) | Owned | Continuously; Not Applicable. | brwn.branden@dol.gov | Not Applicable. | catalog.data.gov/dataset/wage-and-hour-division-compliance-action-data | www.dol.gov/eis/whisard | WHISARD is an automated data processing system that enables investigators, managers, and assistants in the Wage and Hour Division (WHD) to process complaints; assign and investigate cases; manage and close cases; assist with outreach; record and monitor investigator time; track case history through narratives and diary entries; process FOIA and publication requests; and report to management. (www.dol.gov/eis/whisard) | Publicly available data is available at: www.dol.gov/eis/whisard | WHISARD contains WHD compliance actions. Publicly available data show all compliance actions from FY 2005 - onward. | Compliance action. | WHISARD includes data on employers, following the definitions of "employee" within the statutes administered and enforced by WHD. The WHD is responsible for administering and enforcing some of our nation's most comprehensive labor laws, including: the minimum wage, overtime, and child labor provisions of the Fair Labor Standards Act (FLSA); the Family and Medical Leave Act (FMLA); the Migrant and Seasonal Agricultural Worker Protection Act (MSWPA); worker protections provided in several temporary visa programs; and the prevailing wage requirements of the Davis Bacon Act (DBA) and the Service Contract Act (SCA). | WHISARD has limited functionality for corporate and industry structures. It is possible to infer linkages from other fields. For example, Trade Name, Legal Name, or EIN. | "...WHISARD has been improved as an information system...Quality control procedures have been introduced that improve the accuracy and reliability of key fields. The ability to run analyses for specific employees, industries, or geographic areas has been enhanced." (https://www.dol.gov/whd/resources/strategicforcoment.pdf) | EIN | Employer Trade Name; Employer Legal Name; character | Employer Street Address, City Name and State Abbreviation, Zip Code; character | -- | NAICS Code; character | -- | -- | -- | |
| U.S. Department of Commerce | U.S. Census Bureau | Longitudinal Employer Household Dynamics (LEHD) | Owned | Updated quarterly. | CES.Local.Employment.Dynamics@cms.gov | See QCEW entry. | catalog.data.gov/dataset/lehd-orig-destination-employment-statistics-tables catalog.data.gov/dataset/lehd-onthemap-for-employment-management | www.census.gov/ces/data/lehd/ www2.census.gov/cew/w/2014/CES-WP-14-26.pdf | The purpose is to provide high quality local labor market information and to improve the Census Bureau's economic and demographic data programs. | The data are confidential (Title 13 and Title 26, US Code). Access is restricted to persons specially sworn to uphold the confidentiality provisions of Title 13 and Title 26. | Employment, wages, and administrative data (name, location, etc.) for over 9.5 million establishments covering approximately 98% of all employment. | Job-level (worker-firm combinations) | An economic unit that produces goods or services, usually at a single physical location, and engages in one or predominantly one activity - Potential lay synonyms: business, worksite, brick & mortar, site, storefront. Lay users may use "establishment" interchangeably with "firm" in the QCEW, however, there is a significant distinction between the two terms. | The data are reported at the firm-state level and for reporting units (establishments) of the firm-state. The data have been linked to the Census Business Register to attach a firm identification number, firm age, and firm size. | The worker-level LEHD data are the earnings records used for the administration of the Unemployment Insurance program. The LEHD establishment-level data are Employment and Wages. In addition to benefiting from the data quality monitoring and improvements made by these upstream programs, the LEHD program also maintains a data quality staff to ensure the quality of the microdata. | Firm identifier is state employer identification number (SEIN), reporting unit ("establishment") identifier is SEINUNIT id. | Legal Name and Trade Name, character | Yes; Physical location and mailing address of establishments | Yes; County and State; character | 6-digit NAICS codes; | Yes; latitude and longitude to six digits after the decimal; numeric | Census Tract (character), Census Block (character), 4-digit SIC codes (character) | See also: lehd.ces.census.gov/ | |
| Equal Employment Opportunity Commission | Surveys Division | EEO-1 | Desired | Annually | benita.marsh@eoc.gov | See QCEW entry. | catalog.data.gov/dataset/eo1-orig-destination-employment-statistics-tables | www.eoc.gov/employers/eo1survey/index.cfm www.eoc.gov/employers/eo1survey/upload/eo1-2012.pdf | The EEO-1 Report is a compliance survey mandated by federal statute and regulations. The survey requires company employment data to be categorized by race/ethnicity, gender and job category. The EEO-1 Report is used by the EEO-1 Joint Reporting Committee comprised of the Equal Employment Opportunity Commission (EEOC) and the Department of Labor's Office of Federal Contract Compliance Programs (OFCCP) to collect data from private employers and government contractors about their women and minority workforce. The agencies also use the EEO-1 Report data to support civil rights enforcement and to analyze employment patterns, such as the representation of women and minorities within companies, industries or regions. | All reports and information from individual reports will be kept confidential, as required by Section 701(g) of Title VII of the Civil Rights Act of 1964, 42 U.S.C. 2000e-1(g). Only data aggregating information by industry or area, in such a way as not to reveal any particular employer's statistics, will be made public. | All Private employers who are (1) subject to Title VII of the Civil Rights Act of 1964, as amended with 100 or more employees; or (2) private employers subject to Title VII who have fewer than 100 employees if the company is owned or affiliated with another company, or there is centralized ownership, control or management (such as central control of personnel policies and labor relations) so that the group legally constitutes a single enterprise, and the entire enterprise employs a total of 100 or more employees. All Federal contractors (private employees), who (1) are not exempt as provided for by 41 C.F.R. § 60-1.5, (2) have fifty or more employees, and (a) have a federal contract or prime subcontract worth \$50,000 or more, or (b) act as depositories of federal funds in any amount, or (c) act as issuing and paying agents for U.S. savings bonds and notes. Single establishment employers submit only one EEO-1 report, while employers whose business was conducted at more than one location submitted a company wide consolidated report, a headquarters report, and individual reports for each establishment with fifty or more employees. | All at the firm level for all covered firms. Also at the establishment level for establishments with at least 50 employees. | "Employer" under Section 701(b), Title VII of the Civil Rights Act of 1964, as amended, means a person engaged in an industry affecting commerce who has fifteen or more employees for each working day in each of twenty or more calendar weeks in the current or preceding calendar year, and any agent of such a person, but such term does not include the United States, a corporation wholly owned by the government of the United States, an Indian tribe, or any department or agency of the District of Columbia subject by statute to procedures of the competitive service (as defined in section 2102 of Title 5 of the United States Code), or a bona fide labor membership club (other than a private organization) which is exempt from taxation under Section 501(c) of the Internal Revenue Code of 1954. The total number of employees indicated on the headquarters report, PLUS the list of establishments employing fewer than 50 employees, MUST equal the total number of employees shown on the Consolidated Report. OR any person or entity subject to Executive Order 11246 who is a federal government prime contractor or subcontractor at any tier (including a bank or other establishment serving as a depository of federal government funds), or an issuing and paying agent of U.S. Savings Bonds and Notes, or a holder of a federal government bill of lading) or a federally-assisted construction prime contractor or subcontractor at any tier. | Multi-establishment employers, i.e., employers doing business at more than one establishment, must complete online: (1) a report covering the principal or headquarters office; (2) a separate report for each establishment employing 50 or more persons; and (3) a separate report (Type 8 record) for each establishment employing fewer than 50 employees. OR an Establishment List (Type 6 record), showing the name, address, and total employment for each establishment employing fewer than 50 persons, including a Type 6 employment data grid that combines all employees working at establishments employing fewer than 50 employees by race, sex, and job category. For the EEO-1 online application, keyed employment data automatically transfers to the overall Consolidated Report. | EIN, Government contractors are also asked to provide their DUNS number. | Parent company name, establishment name. | Yes; number and street address, city, state, and zip code for parent company and establishment. Also county for establishment. | No. | Verbal description of major activity of the establishment. | Employment definition includes leased employees, unlike the Census Business Register. | | | | |
| Securities and Exchange Commission | Integrated Award Environment | SIC Company Filings | Desired | Continuously, daily updates | | www.sec.gov/info/edgar/s.html | www.sec.gov/dera/data/financial-statements-data-sets.html www.sec.gov/edgar/search/edgar/companysearch.html | www.sec.gov/dera/data/financial-statements-data-sets.html | EDGAR, the Electronic Data Gathering, Analysis, and Retrieval system, performs automated collection, validation, indexing, acceptance, and forwarding of submissions by companies and others who are required by law to file forms with the U.S. Securities and Exchange Commission (SEC). Its primary purpose is to increase the efficiency and fairness of the securities market for the benefit of investors, corporations, and the economy by accelerating the receipt, acceptance, dissemination, and analysis of time-sensitive corporate information filed with the agency. | Publicly available through EDGAR. | All public companies must file. Private companies also make filings when issuing unregistered securities. | | EIN, CIK | Business name | Business and mailing addresses. Includes number and street, city, state, 9-digit zip code. | | 4-digit SIC code | | | | | | | |
| Department of Health and Human Services | Administration for Children and Families, Office of Child Support Enforcement | National Directory of New Hires (NDNH) | Desired | Quarterly | hinda.boyer@acf.hhs.gov | | | www.acf.hhs.gov/ce/neo/urce/a-guide-to-the-national-directory-of-new-hires | The NDNH is a national repository of employment, earnings, and unemployment insurance (UI) information to help state child support agencies locate noncustodial parents, establish child support orders, and enforce child support orders. | Researchers can only receive access to data without personal identifiers, and only for research found by the Secretary of Health and Human Services to be likely to contribute to achieving the purposes of Part A or Part D of the Social Security Act. There has been discussion about new legislation to include statistical purposes in the list of authorized uses. | Similar to QCEW? | Establishment | | | EIN | Employer name | Business address, including number, street name, city, state, and zip code (5-digit, and second field for 4-digit extension if provided) | Foreign country code | | | | | | |
| General Services Administration | Integrated Award Environment | System of Award Management (SAM) | Desired | Continuously, in real time | | www.sam.gov/portals/SAM/ | www.sam.gov | www.sam.gov | Registration is required to apply for a federal gov. contract or grant. | Searchable database and recent filings publicly available on sam.gov. | All firms doing business with the Federal government must file. | Establishment | In accordance with the Federal Acquisition Regulation | Additional information available for authorized users | Data are matched to authoritative sources. | DUNS number, CAGE code | Business name | Business and mailing street address, city, state, 5-digit zip code, and 4-digit zip code extension | Country abbreviation. | NAICS code | company website | | | |

Employer Data Matching Group
Data Inventory

| Department/ Agency | Component Agency or Office | Dataset Name | Dataset Type: Owned by Data Inventory Respondent, or Desired by Data Inventory Respondent? | Data Update Frequency and Lag | Contact E-mail Address | Link to Data Collection Tool | Data.gov Link | Other Relevant Link(s) | Purpose of the Data Collection | Access Restrictions | Coverage | Unit of Analysis | Definition of "Company" | Corporate Structure and Relationships | Quality of Fields for Matching | Unique Identifier Fields | Name Fields | Address Fields | Country Fields | FIPS Code Fields | Industry Code Fields | Geocoding | Other Fields | Additional Notes |
|------------------------------|---|---|--|---|--------------------------|---|--|--|--|---|--|--|--|---|--|--|--|---|--|----------------------------------|--|--|---|---|
| U.S. Department of Commerce | U.S. Census Bureau | Governments Integrated Directory (GID). A migration from the GID to the Governments Master Address File (GMAF) is planned for October 2017. | Owned | The GID universe is updated continuously, although a formal and comprehensive update occurs once every five years during the Census of Governments. There are also annual surveys, including the Boundary and Annexation Survey, the Annual Finance Survey and the Annual Survey of Personnel and Payroll, which provide periodic updates. From the annual survey's collection period, the data are reviewed and edited as necessary, and the GID universe is updated 1.5-2 years following initial collection cycle. | esmd.gov.pfcb@census.gov | www.census.gov/gov/cog/ | N/A | www.meps.ahrq.gov for published statistics from this dataset | In order to measure government activity, various surveys collect statistics for state and local governments and their dependent agencies. A complete list of local governments is needed to sample for these surveys. This list of governments is found in the Governments Integrated Directory (GID). | Public Sector Statistics are considered public information, which is available online through AFF. We also provide additional statistics, not available in our publications, upon data user request in the form of a Special Tabulation file. | The GID includes all governments in the United States, along with their dependent agencies. This includes all states and dependent agencies, counties and townships, municipalities, special districts, authorities, boards and commissions. | Government Unit | A governmental unit is defined as an organized entity which, in addition to having governmental character, has sufficient discretion in the management of its own affairs to distinguish it as separate from the administrative structure of any other governmental unit. An entity must possess all three of the critical attributes to be regarded as a government: Existence as an Organized Entity, Governmental Character, and Substantial Autonomy. | Both parent and dependent agencies as well as independent agencies. | | Each unit has a 14 digit GID ID: SS T CCC UIUU SUP SB where: SS = State. Numeric sequence assigned to the states alphabetically - Alabama is 01, Alaska 02, and so forth, until Wyoming at 51. Note the District of Columbia is assigned code 09 and the U.S. Federal Government is assigned code 00. T = type of government. As defined by the Census Bureau, Government type codes are: 0 = state governments 1 = county governments 2 = municipal governments 3 = township governments 4 = special district governments 5 = school district governments 6 = Federal Government CCC = county area within the state. The county area codes, are assigned alphabetically within each state area, although there are some exceptions where county consolidations have taken place or new counties have been created. UIUU = unit code within the county area. The unit code identifies a specific government within the state, county, and type of government sequence. They, too, are generally alphabetical within the county area and type of government combination, but exceptions exist (especially for special districts). The unit code for a county government is usually the same as its county area code. SUP = supplemental unit code, for dependent agencies of the parent unit. The supplemental code identifies different dependent agencies of the same parent state or local government. Every supplement must have a parent government. The first nine digits of a supplement and parent code always are the same. SB = subsidiary code for dependent agencies that require separate data collection / canvassing. The subsidiary code is used for component units of dependent agencies that require a separate data collection. A good example is found in state dependent university systems, where separate campuses exist for a dependent university, each campus having its own subsidiary code. | Yes - complete governmental unit name | Yes physical location and mailing address | N/A | N/A | N/A | No | | Although this information is specific to the current GID, most of the information also applies to the GMAF with the exception of the ID format. |
| U.S. Census Bureau | Sponsored by the Agency for Healthcare Research and Quality | Medical Expenditure Panel Survey - Insurance Component - Public Sector Employer Sample | Owned | Fielded annually. | Karen.Davis@ahrq.hhs.gov | www.meps.ahrq.gov has information on survey questionnaires and methodology reports. | N/A | www.meps.ahrq.gov for published statistics from this dataset | To collect information about employer-sponsored health insurance. The MEPS-IC collects information on whether an employer offers coverage, details of plans and information on establishment and workforce characteristics. For the government sample, information is collected on all plans that are offered to active employees. | Public Sector Statistics are considered public information, which is available online through AFF (American Fact Finder: https://factfinder.census.gov/). We also provide additional statistics, not available in our publications, upon data user requests in the form of a Special Tabulation file. | A sample drawn from the Governments Integrated Directory (GID) of 1,000 government units | Government Unit | A governmental unit is defined as an organized entity which, in addition to having governmental character, has sufficient discretion in the management of its own affairs to distinguish it as separate from the administrative structure of any other governmental unit. An entity must possess all three of the critical attributes to be regarded as a government: Existence as an Organized Entity, Governmental Character, and Substantial Autonomy. | Both parent and dependent agencies as well as independent agencies. | | Each unit has a 14 digit GID ID: SS T CCC UIUU SUP SB where: SS = State. Numeric sequence assigned to the states alphabetically - Alabama is 01, Alaska 02, and so forth, until Wyoming at 51. Note the District of Columbia is assigned code 09 and the U.S. Federal Government is assigned code 00. T = type of government. As defined by the Census Bureau, Government type codes are: 0 = state governments 1 = county governments 2 = municipal governments 3 = township governments 4 = special district governments 5 = school district governments 6 = Federal Government CCC = county area within the state. The county area codes, are assigned alphabetically within each state area, although there are some exceptions where county consolidations have taken place or new counties have been created. UIUU = unit code within the county area. The unit code identifies a specific government within the state, county, and type of government sequence. They, too, are generally alphabetical within the county area and type of government combination, but exceptions exist (especially for special districts). The unit code for a county government is usually the same as its county area code. SUP = supplemental unit code, for dependent agencies of the parent unit. The supplemental code identifies different dependent agencies of the same parent state or local government. Every supplement must have a parent government. The first nine digits of a supplement and parent code always are the same. SB = subsidiary code for dependent agencies that require separate data collection / canvassing. The subsidiary code is used for component units of dependent agencies that require a separate data collection. A good example is found in state dependent university systems, where separate campuses exist for a dependent university, each campus having its own subsidiary code. | Yes - complete governmental unit name | Yes physical location and mailing address | N/A | N/A | Yes, latitude and longitude to six digits after the decimal, numeric | | Although this information is specific to the current GID, most of the information also applies to the GMAF with the exception of the ID format. | |
| U.S. Census Bureau | Sponsored by the Agency for Healthcare Research and Quality | Medical Expenditure Panel Survey - Insurance Component - Private Sector Employer Sample | Owned | Fielded annually. | Karen.Davis@ahrq.hhs.gov | www.meps.ahrq.gov has information on survey questionnaires and methodology reports. | N/A | www.meps.ahrq.gov for published statistics from this dataset | To collect information about employer-sponsored health insurance. The MEPS-IC collects information on whether an employer offers coverage, details of plans and information on establishment and workforce characteristics. For the private-sector sample, information is collected on up to 4 plans that are offered to active employees. | The Business Register information is confidential (Title 13 and Title 26, US Code). Access is restricted to persons specially sworn to uphold the confidentiality provisions of Title 13 and Title 26. Micro-data can be used in a Census Research Data Center. | A sample drawn from the Business Register of approximately 40,000 establishments each year. | Private-sector establishments | An establishment is a single physical location where business transactions take place and for which payroll and employment records are kept. Groups of one or more establishments under common ownership or control are enterprises. A single-unit enterprise owns or operates only one establishment. A multi-unit enterprise owns or operates two or more establishments. The treatment of establishments on the Business Register differs according to whether the establishment is part of a single-unit or multi-unit enterprise. (https://www.census.gov/econ/subs/methology.html) | Multi-unit enterprises may use multiple EINs, and they may use different ones when reporting to different federal agencies, complicating the matching process. The Census Bureau has linked many, but not all of the EINs to enterprises. | The Business Register provides the most complete, current, and consistent source of establishment-based information about U.S. businesses, and is essential to assuring full coverage and high quality in Federal economic statistics programs. Examples of quality considerations for this source include: - The annual Company Organization Survey covers 30 percent of multi-unit companies and a small sample of firms that were single-unit firms in the most recent quinquennial Economic Census, so establishment openings and closings in the firms not covered may not be reflected in the business register until after the next Economic Census. - The business register is divided into employer and nonemployer business registers based on payroll employment. Some firms leave their employees from Professional Employer Organizations (PEOs) or use independent contractors. Such firms may appear in the nonemployer business register despite having large revenues and many leased and/or contract employees. www.census.gov/econ/overview/mu0606.html | A single-unit enterprise's primary identifier is its Employer Identification Number (EIN). A unique employer unit identification number identifies each establishment (owned by a multi-unit enterprise on the Business Register. (https://www.census.gov/econ/subs/methology.html) | Yes, Primary and secondary business names, permitting identification of both legal and trade designations for businesses that support them (e.g., franchises), character | Yes, Physical location, mailing address, post office (city or town) name, postal state abbreviation, ZIP+4, character | This field is not currently populated in the BR, but it may be possible to link in this information from the Company Organization Survey | Yes, County and State, character | 6 digit NAICS codes | Yes, latitude and longitude to six digits after the decimal, numeric | Census Tract (character), Census Block (character), 4-digit SIC codes (character) | |
| U.S. Department of Education | National Center for Education Statistics | Common Core of Data (CCD) | Owned | Annual Collection. LAG: Data on 2014-15 school year are set to be released in September | mark.glander@ed.gov | ednet.ed.gov/EDENPorta/ | catalog.data.gov/datasets/effects-general-information-201112 | nces.ed.gov/ccd | CCD is a program of the U.S. Department of Education's National Center for Education Statistics that annually collects fiscal and non-fiscal data about all public schools, public school districts and state education agencies in the United States. The data are supplied by state education agency officials and include information that describes schools and school districts, including name, address, and phone number; descriptive information about students and staff, including demographics; and fiscal data, including revenues and current expenditures. | Publicly available data is available through nces.ed.gov/ccd/cddata.asp Otherwise, restricted use datasets can be requested by licensees of the Institute for Education Sciences' Restricted Use Data License | Public, K-12 local education agencies. Universe collection performed from all state education agencies each year. | State, Local Education Agency and School | N/A. Collection is focused upon state and local education agencies | All schools are reported as members of local education agencies/school districts. NCEES assigns each school district a unique identifier (LEAD) and each school it's own unique identifier (SCHNO). These IDs are combined into a 12-digit identifier (NCESSCH) which is unique for each school | LEAD, NCESSCH, PISP, SCHNO, STD, SEASCH | LEANN, SCHNAM | MSTREE, MCITY, MSTATE, MDIP, MZPA, LSTREE, LCITY, LSTATE, LZIP, LZPA | N/A | FPST | N/A | For individual schools only: LATCOD, LONCOD | CONUM (County Number), CONAME (County Name), CDODC (Congressional District code) | | |

Employer Data Matching Workgroup
Data Inventory

| Department/Agency | Component Agency or Office | Dataset Name | Dataset Type: Owned by Data Inventory Respondent, or Desired by Data Inventory Respondent? | Data Update Frequency and Lag | Contact E-mail Address | Link to Data Collection Tool | Data.gov Link | Other Relevant Link(s) | Purpose of the Data Collection | Access Restrictions | Coverage | Unit of Analysis | Definition of "Company" | Corporate Structure and Relationships | Quality of Fields for Matching | Unique Identifier Fields | Name Fields | Address Fields | Country Fields | RPS Code Fields | Industry Code Fields | Geocoding | Other Fields | Additional Notes |
|---|--|--|--|---|---|--|---|---|--|--|--|---|---|--|--|--|---|--|--|--|---|--|--|------------------|
| U.S. Department of Education | National Center for Education Statistics | Integrated Postsecondary Education Data System (IPEDS) | Owned | Annual Collection. LAG: Institutional Characteristics data on 2015-16 was released in July 2016 | ipeds@ipeds.org | nces.ed.gov/ipeds | catalog.data.gov/dataset/01314-integrated-postsecondary-education-data-system | | IPEDS provides basic data needed to describe and analyze trends in postsecondary education in the United States, in terms of the numbers of students enrolled, staff employed, dollars expended, and degrees earned. Congress, federal agencies, state governments, education providers, professional associations, private businesses, media, students and parents, and others rely on IPEDS data for this basic information on postsecondary institutions. IPEDS forms the institutional sampling frame for other NCS postsecondary surveys, such as the National Postsecondary Student Aid Study and the National Survey of Postsecondary Education. | Public data is available through: nces.ed.gov/ipeds/Home/UseTheData | Postsecondary Institutions. Universe collection performed every year | Postsecondary Institutions | N/A. Collection is focused upon postsecondary institutions | Several institutions are recognized as "multi-institution" or "multi-campus" organizations, otherwise each institution is reported as an independent entity | | UNTIID (Unique identification number of the institution), EIN (Employer Identification Number), DPID (Office of Postsecondary Education (OPE) ID Number), NEWID (UNTIID for merged schools), FISSCDD (ID number of multi-institution or multi-campus organization) | INSTNM (Institution [entity] name); FYSYSNAM (Name of multi-institution or multi-campus organization) | ADDR, CITY, STABBR, ZIP | N/A | RPS | N/A | LONGITUDE, LATITUDE | | |
| U.S. Department of Education | National Center for Education Statistics | Private School Universe Survey (PSS) | Owned | Biennial (most recent publicly available data: 2011-12) | stephen.broughman@ed.gov | N/A | catalog.data.gov/dataset/01516-private-school-universe-survey | | The purpose of this data collection activity are: a) to generate biennial data on the total number of private schools, teachers, and students; and b) to build an accurate and complete list of private schools to serve as a sampling frame for NCS surveys of private schools. The PSS began with the 1989-90 school year and has been conducted every two years since. | Public data available at: nces.ed.gov/survey/pss/ssdata.asp | All private schools | Individual private schools | N/A | N/A | PNIN (Permanent Identification Number) | PNST (School Name) | PADDRS, PCITY, PSTABR, PZIP, PZIP4, PCNTY, PCNTNM, PL_ADD, PL_CITY, PL_STABR, PL_ZIP, PL_ZIP4 | N/A | PSTANSI (ANSI State Code) | N/A | LATITUDE, LONGITUDE# (where # = year of collection) | REGION (Census region) | | |
| U.S. Department of Education | Office for Civil Rights / National Center for Education Statistics | Civil Rights Data Collection | Owned | Biennial Collection. LAG: Data on 2013-14 school year were released in Spring 2016 | stephanie.miller@ed.gov (NCS); mary.schiffer@ed.gov (OCR) | Not currently online (informational link: www2.ed.gov/about/offices/olc/ocr/crda.html) | catalog.data.gov/dataset/civil-rights-data-collection-2013-14 | ocrdata.ed.gov/ | Since 1968, the U.S. Department of Education (ED) has conducted the Civil Rights Data Collection (CRDC) to collect data on key education and civil rights issues in our nation's public schools. The collection was formerly administered as the Elementary and Secondary School Survey (EAS Survey). The CRDC collects a variety of information including student enrollment and educational programs and services, most of which is disaggregated by race/ethnicity, sex, limited English proficiency and disability. The CRDC is a longstanding and important aspect of the ED Office for Civil Rights (OCR) overall strategy for administering and enforcing the civil rights statutes for which it is responsible. | Publicly available data is available through ocrdata.ed.gov Otherwise, restricted use datasets can be requested by licensees of the Institute for Education Science's Restricted Use Data License | Public, K-12 local education agencies. Universe collection from all agencies every two years | Primary unit of analysis: individual public schools. Secondary unit of analysis: individual school districts/local education agencies | N/A. Collection is focused upon school districts | There are three main identifiers (IDs) used in the data files: one to identify the LEA, one to identify the school (in school-level files), and a combined LEA/school ID: LEAID is the 7-digit district identification code. SCHID is the 7-digit school identification code. COMBKEY is the 7-digit LEAD + the 5-digit SCHD. Districts or schools that had an existing LEAD or SCHD kept that ID. In most cases, the LEAID and SCHID match the National Center of Education Data (NCS) ID used in EDData, although there are some cases that differ because of differences in definitions and procedures between EDData and the CRDC. Districts that did not have an existing ID were issued an NCS ID if there was one, or a new 7-digit LEAID, which consisted of the 2-digit state RPS code followed by a 5-digit ID assigned during collection. Schools that did not have an existing ID were issued an NCS ID if there was one, or a new ID which consisted of the LEAID and a 6-digit school code assigned during collection. | LEAID, SCHID, COMBKEY | LEA_NAME | LEA_ADDRESS, LEA_CITY, LEA_ZIP | N/A | NOTE: The first two digits of LEAID are the RPS Code for that district, however the RPS code is not separately reported within its own field | N/A | N/A | | | |
| U.S. Department of Commerce - Bureau of Economic Analysis | Direct Investment Division (DID) | BEA surveys on U.S. direct investment abroad | Owned | Quarterly, annual, and every five years depending on the data item and on the size of the reporting unit | ricardo.limes@bea.gov | www.bea.gov/surveys/dias/surv.htm | Not Applicable | Statistics: http://www.bea.gov/international/direct_investment_multinational_companies_comprehensive_data.htm | BEA produces comprehensive statistics on U.S. direct investment abroad that are essential to the compilation of the U.S. economic accounts and for the analysis of multinational enterprises (MNEs). | Aggregates are public; microdata are not public | Statistics on U.S. multinational enterprises (MNEs). U.S. direct investment abroad (DIDA) is defined as ownership by a U.S. investor of at least 10 percent of a foreign business. The direct investor is known as a U.S. parent, and the U.S.-owned foreign business is known as a foreign affiliate. The combined global operations of a U.S. parent company and its foreign affiliates constitute a U.S. MNE. BEA produces two broad sets of statistics on U.S. direct investment abroad: (1) Direct investment transactions and positions statistics (which are used in the U.S. International Transaction and Investment Position Accounts), and (2) statistics on the activities of U.S. MNEs. | Enterprise level; data are available on U.S. parents and each of their foreign affiliates. For U.S. parents, data are available for the fully consolidated U.S. domestic business enterprise. | The U.S. reporter is the U.S. person that has direct investment in a foreign business enterprise, including a branch. If the U.S. parent is a business enterprise, the U.S. parent is the fully consolidated U.S. domestic business enterprise (see corporate structure and relationships for consolidation rules). The foreign affiliate is a business enterprise located in one country which is directly or indirectly owned or controlled by a U.S. person to the extent of 10 percent or more. A business enterprise is any organization, association, branch, or venture which exists for profit making purposes or to otherwise secure economic advantage, and any ownership of any real estate. | The data are collected and reviewed according to documented procedures and best practice standards and ongoing review by the appropriate supervisor. The data are validated using a battery of edit checks to detect potential errors and to otherwise ensure that the data are accurate, reliable, and relevant for the estimates being made. Data are routinely revised as more complete source data become available. | ENIs are collected for U.S. parents. | Name of U.S. parent, name of contact person for U.S. parent, and name of each foreign affiliate (all character) | Address of U.S. parent and address of contact person for U.S. parent. (Character) | Each foreign affiliate record includes a field indicating which country it is located in. All U.S. parents are located in the United States. | No | International Surveys Industry (ISI) classification system, which is based on 4-digit NAICS codes (http://www.bea.gov/surveys/itrcmat.htm). Each reporting unit is classified into a single industry for publication, but BEA collects data for up to 10 industries in which they have revenue. This more detailed industry breakdown approximates the industry of the establishments that comprise the enterprise. | No | Telephone and fax numbers for contact person and certifying official. Email address of contact person. | | |
| U.S. Department of Commerce - Bureau of Economic Analysis | Direct Investment Division (DID) | BEA surveys on foreign direct investment in the United States | Owned | Quarterly, annual, and every five years depending on the data item and on the size of the reporting unit | ricardo.limes@bea.gov | www.bea.gov/surveys/di/surv.htm | Not Applicable | Statistics: http://www.bea.gov/international/direct_investment_multinational_companies_comprehensive_data.htm | BEA produces comprehensive statistics on foreign direct investment in the United States that are essential to the compilation of the U.S. economic accounts and for the analysis of multinational enterprises (MNEs). | Aggregates are public; microdata are not public | Statistics on U.S. affiliates of foreign multinational enterprises (MNEs). Foreign direct investment in the United States is defined as the ownership by a foreign investor of 10 percent or more of a U.S. business. The direct investor is known as a foreign parent, and the foreign-owned U.S. business is known as a U.S. affiliate. BEA produces three broad sets of statistics on foreign direct investment in the United States: (1) Direct investment transactions and positions statistics (which are used in the U.S. International Transaction and Investment Position Accounts), (2) statistics on the activities of U.S. affiliates of foreign MNEs, and (3) new foreign direct investment statistics. | Enterprise level; data are available for the consolidated U.S. affiliate. | A U.S. affiliate is a business enterprise located in the United States which is directly or indirectly owned or controlled by a person of another country to the extent of 10 percent or more. A business enterprise means any organization, association, branch, or venture which exists for profit making purposes or to otherwise secure economic advantage, and any ownership of any real estate. | The consolidated U.S. affiliate consists of the U.S. business enterprise which is directly owned by a foreign person and all U.S. business enterprises proceeding down each ownership chain whose voting securities are more than 50 percent owned by the U.S. business enterprise above. The fully consolidated entity is considered one U.S. affiliate. Data on the industry of sales for U.S. affiliates are classified under the International Surveys Industry classification system, which is based on 4-digit NAICS codes (http://www.bea.gov/surveys/itrcmat.htm). | The data are collected and reviewed according to documented procedures and best practice standards and ongoing review by the appropriate supervisor. The data are validated using a battery of edit checks to detect potential errors and to otherwise ensure that the data are accurate, reliable, and relevant for the estimates being made. Data are routinely revised as more complete source data become available. | ENIs are collected for the U.S. business enterprise which is foreign-owned and for each U.S. business enterprise consolidated with it. | Name of U.S. affiliate, name of contact person for U.S. affiliate, name of certifying official, and name of each U.S. business enterprise consolidated. (all character) | Address of U.S. affiliate and address of contact person for U.S. affiliate. (Character) | Each U.S. affiliate is located in the United States. Each U.S. affiliate record includes a field indicating the country of the foreign parent and a field with the country of the ultimate beneficial owner (which may be a foreign or U.S. person or entity). | No | International Surveys Industry (ISI) classification system, which is based on 4-digit NAICS codes (http://www.bea.gov/surveys/itrcmat.htm). The consolidated affiliate is classified into a single industry for publication, but BEA collects data for up to 10 industries in which they have revenue. This more detailed industry breakdown approximates the industry of the establishments that comprise the enterprise. | No | Telephone and fax numbers for contact person and certifying official. Email address of contact person. Number of employees associated with each industry. Number of employees, number of manufacturing employees, gross property, plant, and equipment (PP&E), and commercial property by state. | |
| U.S. Department of Commerce - Bureau of Economic Analysis | Balance of Payments Division | BEA surveys on U.S. international trade in services | Owned | Quarterly and annually depending on the level of component detail; more granular data by country and by type of service are published annually than quarterly | christopher.stein@bea.gov | www.bea.gov/surveys/us/surv.htm | Not Applicable | Statistics: www.bea.gov/tables/tables.htm?reqid=62&steps=2&sur=186210-1 | BEA combines data from the trade in services surveys with data from other sources to compile the official statistics on trade in services, a component of the U.S. international transactions accounts. Other government agencies use these statistics to formulate U.S. trade policy, to analyze the impact of that policy and to support trade promotion activities. The trade statistics are also used by companies involved in international trade, researchers, international organizations, and the general public. | Aggregates are public; microdata are not public | BEA's international trade in services surveys collect data on U.S. exports and imports of transportation services, financial services, insurance services, other business services, and charges for the use of intellectual property. | Respondents include U.S. persons engaged in international trade in services who person means any individual, branch, partnership, associated group, association, estate, trust, corporation, or other organization (whether or not organized under the laws of any State), and any government (including a foreign government, the United States Government, a State or local government, and any agency, corporation, financial institution, or other entity or instrumentality thereof, including a government sponsored agency). If the U.S. person is a business enterprise, the U.S. parent is the fully consolidated U.S. domestic business enterprise (see corporate structure and relationships for consolidation rules). A business enterprise is any organization, association, branch, or venture which exists for profit making purposes or to otherwise secure economic advantage, and any ownership of any real estate. | The fully consolidated U.S. domestic business enterprise is defined as: 1) The U.S. business enterprise whose voting securities are not owned more than 50 percent by another U.S. business enterprise, and 2) proceeding down each ownership chain from that U.S. business enterprise, any U.S. business enterprise (including Foreign Sales Corporations located in the United States) whose voting securities are more than 50 percent owned by the U.S. business enterprise above it. | The industry codes reported reflect BEA's International Surveys Industry classification system, which is based on 4-digit NAICS codes (http://www.bea.gov/surveys/itrcmat.htm). | The data are collected and reviewed according to documented procedures and best practice standards and ongoing review by the appropriate supervisor. The data are validated using a battery of edit checks to detect potential errors and to otherwise ensure that the data are accurate, reliable, and relevant for the estimates being made. Data are routinely revised as more complete source data become available. | ENIs of survey respondents are collected. | Name of U.S. "person" responding to the survey, name of contact person, and name of certifying official; all character fields. | Address of the U.S. "person" responding to the survey and address of contact person; both character fields. | All survey respondents are U.S. "persons." | No | BEA's International Surveys Industry (ISI) classification system, which is based on 4-digit NAICS codes (http://www.bea.gov/surveys/itrcmat.htm). | No | Telephone and fax numbers for contact person and certifying official. Email address of contact person. | |
| U.S. Department of Labor | Occupational Safety & Health Administration | OSHA Information System (OIS) | Owned | Continuously; Not Applicable. | Moore, Joshua (Moore.Joshua@dol.gov) | Not Applicable. | catalog.data.gov/dataset/publisher-U.S. Department of Labor Occupational Safety and Health Administration | enforcedata.dol.gov/view/data_catalog.php | To track, document and analyze OSHA's Enforcement and Consultation programs. | Only users in the OSHA family have access to the system. Some data is available via osha.gov and the data.gov | OIS contains all Enforcement and Consultation information starting in FY 11 in a staggered roll out. All Federal enforcement was on the system starting in FY13. | Compliance actions, Outreach and Consultation services | "Employer" means a person engaged in business affecting commerce who has employees, but does not include the United States or any State or political subdivision of a State. OSHA does not issue citations with penalties. OSHA collects data on the establishment (business location) level; employers are the "parents" of establishments. | OIS does not have this functionality at this time. | All data is user entered. We have edit checks on the number of characters on fields such as the FEIN but we have no way to verify if the numbers entered are accurate. | Federal EIN/TIN, DUNS, DUNS plus4 and CAGE Code. All of these are optional in the system. | Establishment Name and Doing Business As Name | Full address of Mailing Business and Site Address | Business and Mailing address contain Country, International City, International Division and International Postal Code | Yes; County and State; character | 6 digit NAICS | No | | |
| U.S. Department of Labor | Occupational Safety & Health Administration | OSHA Legacy Data (OLD) replaces the Legacy IMS system (Aug 22, 2016) | Owned | Continuously; Not Applicable. | Legasse, Gizaw (Legasse.Gizaw@DOL.GOV) | Not Applicable. | catalog.data.gov/dataset/publisher-U.S. Department of Labor Occupational Safety and Health Administration | enforcedata.dol.gov/view/data_catalog.php | To track, document and analyze OSHA's Enforcement and Consultation programs. NO new case data is being collected but maintained for historical records. (New case data is being collected in OIS) | Only users in the OSHA family have access to the system. Some data is available via osha.gov and the data.gov | The Integrated Management Information System (IMIS) contained all Enforcement and Consultation information starting in FY 2010. The IMIS data was migrated over a new database called OSHA Legacy Data (OLD) | Compliance actions, Outreach and Consultation services | "Employer" means a person engaged in business affecting commerce who has employees, but does not include the United States or any State or political subdivision of a State. OSHA does not issue citations with penalties. OSHA collects data on the establishment (business location) level; employers are the "parents" of establishments. | OLD does not have this functionality at this time. | All data is user entered. We have edit checks on the number of characters on fields such as the FEIN but we have no way to verify if the numbers entered are accurate. | Federal EIN/TIN, DUNS, All of these are optional in the system. | Establishment Name and Doing Business As Name | Full address of Mailing Business and Site Address | Business and Mailing address contain Country, International City, International Division and International Postal Code | Yes; County and State; character | 6 digit NAICS | No | | |

Employer Data Matching Workgroup
Data Inventory

| Department/Agency | Component Agency or Office | Dataset Name | Dataset Type: Owned by Data Inventory Respondent, or Desired by Data Inventory Respondent? | Data Update Frequency and Lag | Contact E-mail Address | Link to Data Collection Tool | Data.gov Link | Other Relevant Link(s) | Purpose of the Data Collection | Access Restrictions | Coverage | Unit of Analysis | Definition of "Company" | Corporate Structure and Relationships | Quality of Fields for Matching | Unique Identifier Fields | Name Fields | Address Fields | Country Fields | FIPS Code Fields | Industry Code Fields | Geocoding | Other Fields | Additional Notes | | |
|------------------------------------|---|--|--|--|---|--|---|---|---|---|--|---|---|--|---|---|--|--|--|--|---|--|--|--|--|--|
| U.S. Department of Labor | Occupational Safety & Health Administration | Whistleblower Program Retaliation Data | Owned | Continuously, Not Applicable. | Garrahan, MaryAnn - OSHA <Garrahan.MaryAnn@doh.gov> | www.osha.gov/whistleblower/WBComplaint.html | Not Applicable | Not Applicable | To track and administer the OSHA whistleblower protection provisions that include more than twenty whistleblower protection statutes, including Section 116(g) of the Occupational Safety and Health (OSH) Act, which prohibits any person from discharging or in any manner retaliating against any employee because the employee has complained about unsafe or unhealthful conditions or exercised other rights under the Act. Whistleblower protection provisions administered by OSHA also protect employees from retaliation for reporting violations of various airline, commercial motor carrier, motor vehicle safety, consumer product, environmental, consumer finance, food safety, health insurance reform, nuclear, pipeline, public transportation agency, railroad, maritime and securities laws. | Only pertinent OSHA OWPP employees | The Integrated Management Information System (IMIS) contains Whistleblower Complaint and Investigation information starting in FY1982. | Whistleblower complaints | "Employee" is statutorily defined by the numerous statutes which contain the 22 different whistleblower provisions administered by OSHA. Under the whistleblower provisions, "employee" may include public and private sector employees, including the U.S. Postal Service, as well as other federal, state, and tribal entities. OSHA collects data on the establishment (Business location) level; employees are the "parents" of establishments. | Whistleblower database does not have this functionality | All data is user entered. We have edit checks on the number of characters on fields such as the FEIN but we have no way to verify if the numbers entered are accurate. | Federal EIN/TIN, DUNS, All of these are optional in the system. | Establishment/Employer Name and Doing Business As Name | Full address of Mailing Business and Site Address | Business and Mailing address contain Country, International City, International Division and International Postal Code | Yes, County and State, character | 6 digit NAICS | No | No | No | | |
| U.S. Small Business Administration | Office of Entrepreneurial Development | Entrepreneurial Development Management Information System (EDMIS) | Owned | Updated Quarterly, no lag | edmis@sba.gov | N/A | N/A | www.sba.gov/edmis | Data are collected for program administration, performance, and evaluation purposes. | | Client (entrepreneur or small business owner) level data is available for all those who have sought services from OED Resource Partner or online Learning Center course. There are approximately 1 M records annually across all programs. | Client (clients may or may not own a business) | The client self-identifies if they have an active business | N/A | There is no client level PI in the system | None | Client names are not required | Client address is not required | Not required | No | Industry Category (not NAICS) | No | No | N/A | | |
| U.S. Department of Commerce | US Census Bureau | Survey of Business Owners | Owned | Annually | esd.survey.of.business.owners@census.gov | www.census.gov/program/s-surveys/sbo/technical-documentation/questionnaires.html | www.census.gov/econ/overview/m2200.html | | The Survey of Business Owners (SBO) provides the only comprehensive, regularly collected source of information on selected economic and demographic characteristics for businesses and business owners by gender, ethnicity, race, and veteran status. | The data are confidential (Title 13 and Title 26, US Code). Access is restricted to persons specially sworn to uphold the confidentiality provisions of Title 13 and Title 26. | Included are all nonfarm businesses filing Internal Revenue Service tax forms as individual proprietors, partnerships, or any type of corporation, and with receipts of \$1,000 or more. The SBO covers both firms with paid employees and firms with no paid employees. The SBO is conducted on a company or firm basis rather than an establishment basis. A company or firm is a business consisting of one or more domestic establishments that the reporting firm specified under its ownership or control. | Firm, usually single unit | A firm is a business organization or entity consisting of one domestic establishment (location) or more under common ownership or control. All establishments are included as part of the owning or controlling firm. For the economic census, the terms "firm" and "company" are synonymous. | NA | Outside data must first be matched to the Business Register and then linked to the SBO using the BR's business identifiers. | Internal Census identifiers such as the LBDNUM | NA (available through Business Register link) | NA (available through Business Register link) | Yes, County and State, character | 6-digit NAICS | NA (available through Business Register link) | | | | | |
| U.S. Department of Commerce | US Census Bureau | Longitudinal Business Database | Owned | Annually | esr.contacts@census.gov | The LBD is built by linking consecutive years of the Census Bureau's Business Register | www.census.gov/evs/data/products/datasets/lbd.html | | The Business Register information is confidential (Title 13 and Title 26, US Code). Access is restricted to persons specially sworn to uphold the confidentiality provisions of Title 13 and Title 26. | The data are compiled by combining data | Establishment-based. | An establishment is a single physical location where business transactions take place and for which payroll and employment records are kept. Groups of one or more establishments under common ownership or control are enterprises. A single-unit enterprise owns or operates only one establishment. A multi-unit enterprise owns or operates two or more establishments. | Multi-unit enterprises may use multiple EINs, and they may use different ones when reporting to different federal agencies, complicating the matching process. The Census Bureau has linked many, but not all of the EINs to enterprises. | Outside data must first be matched to the Business Register and then linked to the LBD using the BR's business identifiers. | A Census constructed "LBDNUM" field links establishments over time. | NA | NA | Yes, county, state, FIPS, character. | 6-digit NAICS codes; | Yes; latitude and longitude to six digits after the decimal; numeric | | | | | | |
| U.S. Department of Transportation | NHTSA | Manufacturer Information Database (MID) | Owned | Continuously | manufacturerinfo@dot.gov | vpic.nhtsa.dot.gov/mid | catalog.data.gov/dataset/hvsa-product-information-catalog-and-vehicle-licensing-vpic-mid-8ee79 | vpic.nhtsa.dot.gov/ | Per CFR 49 Part 566, CFR 49 Part 565, CFR 49 Part 574 | Publicly available data is available at: vpic.nhtsa.dot.gov/mid | Manufacturer information database contains the following items: - World Manufacturer Identifiers (WMIs) assigned to a manufacturer assigned by SAJ and referenced in VPIIC. - Motor vehicle manufacturer information (Address, Products Manufactured, Makes, etc.). - Motor vehicle identification information (vehicle VIN) - Vehicle equipment plant information | establishment | Motor vehicle modifiers for handicapped | MID is designed to store parent/child relationship, but most of the fields are not populated. | When data are collected, there is a group of staff reviewing the information before they are accepted into the system. Email are sent out to the entities to confirm the contact information submitted. For vehicle identification information, quality check run are regularly performed to compare with the results from other databases. | manufacturerID (System generated) | Common Name, Previous Legal Name URL, All Company DBAs, Trade/Brand Names | Full company mailing address | Contains country fields for non-US companies | no | no | | | | | |
| U.S. Department of Transportation | NHTSA | Manufacturer Information Database (MID) - Modifiers | Owned | Continuously | manufacturerinfo@dot.gov | vpic.nhtsa.dot.gov/mid/home/ModifierSearch | TBD | vpic.nhtsa.dot.gov/ | CFR 49 Part 595 | Publicly available data is available at: vpic.nhtsa.dot.gov/mid/home/ModifiersSearch | Manufacturer information database contains the following items: - Vehicle modifier (for handicapped) registration information | establishment | Motor vehicle industry manufacturers and alterers of vehicles, motor vehicle equipment plants, motor vehicle modifiers for handicapped | MID is designed to store parent/child relationship, but most of the fields are not populated. | When data are collected, there is a group of staff reviewing the information before they are accepted into the system. Email are sent out to the entities to confirm the contact information submitted. For vehicle identification information, quality check run are regularly performed to compare with the results from other databases. | manufacturerID (System generated) | Common Name, Previous Legal Name URL, All Company DBAs, Trade/Brand Names | Full company mailing address | Contains country fields for non-US companies | no | NO | | | | | |
| U.S. Department of Transportation | Pipeline and Hazardous Materials Administration (PHMSA) Office of Pipeline Safety | pipeline operator registry data is contained in PHMSA Pipeline DataMart(PDM) | Owned | Continuous update by operator for registration and notification changes. | william.mccarthy@dot.gov | Not publicly available | catalog.data.gov/dataset/hvsa-product-information-catalog-and-vehicle-licensing-vpic-mid-8ee79 | primis.phmsa.dot.gov/com/for/for.html | The pipeline national operator registry is a data collection tool PHMSA developed to provide pipeline operators an automated portal in which they can register and update changes to their pipelines. This data is combined with various regulatory applications to perform inspections and enforcements of all interstate gas and liquid pipelines. | This data is collected and conformed with inspections, enforcements, incident and annual reporting in a PHMSA data warehouse that is available to PHMSA and state regulatory agencies | registration based | Companies may be either an owner of a pipeline or an operator for another owner. | Corporate pipeline operator may structure themselves as a single corporate entity or sub corporate entities. | an operator identification number is assigned by PHMSA upon operator registration in the PHMSA application portal. | Company name | Physical address | headquarters country | none | none | none | company contacts(name, phone, email, address) | | | | | |
| U.S. Department of Transportation | Pipeline and Hazardous Materials Administration (PHMSA) Office of Hazmat Safety | Hazardous Materials Information Portal (HIP) | Owned | Updated daily by various PHMSA and FMCSA sources | adrian.carter@dot.gov | Not publicly available | Special Permits: catalog.data.gov/dataset/hvsa-special-permits-data-mining-tool-07acf Approvals: catalog.data.gov/dataset/hvsa-approval-data-mining-tool-9e63b Hazmat incident reports: catalog.data.gov/dataset/hvsa-hazmat-incident-reports-data-mining-tool-e8800 Hazmat Registration: catalog.data.gov/dataset/hvsa-hazmat-registration-statistics-2011-f2e80 | www.phmsa.dot.gov/hazmat-services | The purpose of gathering this data is to help regulate those that manufacture, transport, ship, test and package hazardous materials. | This information is restricted to US DOT hazmat personnel | All records are transactional | Complains include hazmat carriers, shippers, manufacturers, packages and testers. | Corporate relationship exists between locations that share a legal liability | Company information is maintained in the HIP master company hub (MCH). The fields include company name, address, city, state zip and D&B number. Any company that does business with PHMSA through the various transactions gets validated by D&B and is added to the MCH if they don't already exist. | The D&B number is used as the unique identifier in the master company hub. If no D&B number exists, the system generates a unique number. | Company name | Physical address | physical country | FIPS country code | DOT Mode of transportation, US DOT #, D&B number | N/A | N/A | N/A | | | |
| U.S. Department of Transportation | Federal Aviation Administration | Air Operators | Owned | Weekly | Tim.Perez@faa.gov | av-info.faa.gov/data/AirOperators/Ab/AnoOpera.txt | av-info.faa.gov/data/AirOperators/Ab/AnoOpera.txt | av-info.faa.gov/data/AirOperators/Ab/AnoOpera.txt | The FAA regulates the Air Operators. | The dataset is publicly available. | Contains Air Operator name, certificate number, and types of certificates. Contains 5210 records. | Establishment based | The company is an Air Operator. This is a designator of an Air Operator or Air Agency by a governmental registration authority. They are certificate holders. The designator is an identifying code that identifies an organization as holding one or more certificates for conducting aviation operations. Air Operator is often referred to as an "air carrier" which is a organization that undertakes directly by lease, or other arrangement, to engage in air transportation. | | Air Operator Designation Code (Design_Code) | Air Operator Name | None | No, does not contain this field. | No, does not contain this field. | None | None | None | None | The file layout is: av-info.faa.gov/data/AirOperators/AirOpera.doc | | |
| U.S. Department of Transportation | Federal Aviation Administration | Contractors for Air Operators | Owned | Weekly | Tim.Perez@faa.gov | av-info.faa.gov/data/AirOperators/Ab/AnoOpera.txt | av-info.faa.gov/data/AirOperators/Ab/AnoOpera.txt | av-info.faa.gov/data/AirOperators/Ab/AnoOpera.txt | The FAA regulates the Air Operators. | The dataset is publicly available. | Contains Air Operator name, address, and type. Contains 3075 records | Establishment based | The company is an employer. The company provides contracting services to Air Operators. | | Contractor Designation Code (Contractor_Design_Code) | Air Operator Contractor Name | Employer Street Address, City Name and State Abbreviation, Zip Code. Contains the full address. | Yes, the dataset does include a Country field and contains data of countries outside of the United States, such as Canada. | No, does not contain this field. | None | None | Air Operator Designation Code (Design_Code) can be used to tie this Contractor record to an associated Air Operator record in the various datasets containing Air Operator. | The file layout is: av-info.faa.gov/data/AIROPERATORS/Contractor.doc | | | |
| U.S. Department of Transportation | Federal Aviation Administration | Air Operator CEO Name | Owned | Weekly | Tim.Perez@faa.gov | av-info.faa.gov/data/AirOperators/Ab/AnoOpera.txt | av-info.faa.gov/data/AirOperators/Ab/AnoOpera.txt | av-info.faa.gov/data/AirOperators/Ab/AnoOpera.txt | The FAA regulates the Air Operators. | The dataset is publicly available. | Contains Air Operator CEO's name, address, phone. Contains 5910 records | | The company is an employer. This is a designator of an Air Operator or Air Agency by a governmental registration authority. They are certificate holders. The designator is an identifying code that identifies an organization as holding one or more certificates for conducting aviation operations. Air Operator is often referred to as an "air carrier" which is a organization that undertakes directly by lease, or other arrangement, to engage in air transportation. | | Air Operator Designation Code (Design_Code) | Name of the Chief Executive Officer of Air Operator | Chief Executive Officer Street Address, City Name and State Abbreviation, Zip Code. Contains the full address. | Yes, the dataset does include a Country field and contains data of countries outside of the United States, such as Canada. | No, does not contain this field. | None | Air Operator Designation Code (Design_Code) can be used to tie this Contractor record to an associated Air Operator record in the various datasets containing Air Operator. | The file layout is: av-info.faa.gov/data/AIROPERATORS/OperC.doc | | | | |
| U.S. Department of Transportation | Federal Aviation Administration | Air Operator Doing Business As Name | Owned | Weekly | Tim.Perez@faa.gov | av-info.faa.gov/data/AirOperators/Ab/AnoOpera.txt | av-info.faa.gov/data/AirOperators/Ab/AnoOpera.txt | av-info.faa.gov/data/AirOperators/Ab/AnoOpera.txt | The FAA regulates the Air Operators. | The dataset is publicly available. | Contains Air Operator Doing Business As Authorized name. Contains 1784 records | Establishment-based | The company is an employer. This is a designator of an Air Operator or Air Agency by a governmental registration authority. They are certificate holders. The designator is an identifying code that identifies an organization as holding one or more certificates for conducting aviation operations. Air Operator is often referred to as an "air carrier" which is a organization that undertakes directly by lease, or other arrangement, to engage in air transportation. | | Air Operator Designation Code (Design_Code) | Authorized Doing Business As Name for an Air Operator | None | No, does not contain this field. | No, does not contain this field. | None | None | Air Operator Designation Code (Design_Code) can be used to tie this Agency by a governmental registration authority record to an associated Air Operator record in the various datasets containing Air Operator. | The file layout is: av-info.faa.gov/data/AIROPERATORS/OperDBA.doc | | | |
| U.S. Department of Transportation | Federal Motor Carrier Safety Administration | Motor Carrier Census | Owned | Continuously | Norma.Org@dot.gov | | si.fmcsa.dot.gov/SMS/Tools/Downloads.aspx | | The FMCSA regulates carriers engaged in the transportation of people and goods across state lines. | The dataset is publicly available. | is a Census of interstate motor carriers | Establishment-based | Per the Federal Motor Carrier Safety Regulations. See guidance www.fmcsa.dot.gov/registration/unified-registration-system | Only headquarters element. No relationships maintained. | Company name, address, and EIN/TIN (ETN in the case of small owner-operators) | US DOT Number | Company name | Employer Street Address, City Name and State Abbreviation, Zip Code. Contains the full address. | Yes, if a Mexican/Canadian domiciled company. | No, does not contain this field. | None | None | USDOT Number is an agency assigned code. | See download link for file layout. | | |

Employer Data Matching Workgroup
Data Inventory

| Department/ Agency | Component Agency or Office | Dataset Name | Dataset Type: Owned by Data Inventory Respondent, or Desired by Data Inventory Respondent? | Data Update Frequency and Lag | Contact E-mail Address | Link to Data Collection Tool | Data.gov Link | Other Relevant Link(s) | Purpose of the Data Collection | Access Restrictions | Coverage | Unit of Analysis | Definition of "Company" | Corporate Structure and Relationships | Quality of Fields for Matching | Unique Identifier Fields | Name Fields | Address Fields | Country Fields | FIPS Code Fields | Industry Code Fields | Geocoding | Other Fields | Additional Notes |
|------------------------------------|--|-----------------------------------|--|--|------------------------|---|---|---|--|--|--|-----------------------------------|--|--|--|--|---|--|---|------------------|-----------------------|-----------|---|--|
| U.S. Department of Transportation | Federal Railroad Administration | List of railroads and companies | Owned | Continuously | Andrew.Martin@dot.gov | | | safetydata.fra.dot.gov/OfficeofSafety/publicize/downloads/RailroadCompanyFile.aspx | The FRA regulates railroad companies. | The dataset is publicly available. | | Establishment-based | See safetydata.fra.dot.gov/OfficeofSafety/publicize/ReportingLevelDefinitions.aspx | See safetydata.fra.dot.gov/OfficeofSafety/publicize/ReportingLevelDefinitions.aspx | Company Name, address | Railroad code | Company name | Employer Street Address, City Name and State Abbreviation, Zip Code Contains the full address. | No, does not contain this field. | No | None | none | Railroad identifier is industry-assigned | See download link for file layout. |
| U.S. Small Business Administration | Office of Disaster Assistance | Disaster Credit Management System | Owned | Updated Annually | N/A | www.sba.gov/loans-grants/for-whole-sba-offers/sba-loan-programs/disaster-loans/disaster-loan-data | N/A | www.sba.gov/loans-grants/for-whole-sba-offers/sba-loan-programs/disaster-loans/disaster-loan-data | Data are collected for program general public knowledge and evaluation purposes. | None | 2004 - 2015 | Individual Disaster Loan Approval | Individuals / companies are not revealed | N/A | No PII is within files | Declaration ID / zip code / state / county / business or home loan | N/A | N/A | Not required | No | No | zip code | No | N/A |
| U.S. Small Business Administration | Office of Capital Access | EIS | Owned | Updated weekly run | N/A | N/A | N/A | N/A | E-tran - system for electronically capturing loan application information from lenders; Data collected to track 7(A) & 504 loan approvals | None | All loans approved, information updated in real-time with weekly database refreshes. | Individual Loan Approval | Borrower Firm Name, TIN | N/A | High: Full details of loan approval process, including the lender and bank, vetted by bank before the approval | TIN, Demographic, location, details on loan program | Borrower & Lender Names captured | Borrower & Lender Addresses captured | Yes | No | NAICS Code; character | Zip code | No | N/A |
| U.S. Small Business Administration | Office of Capital Access | MPERS | Owned | Updated weekly run | N/A | N/A | N/A | N/A | MPERS - system for Microloan intermediaries to self-report Microloan loan information; Data collected from Microloan intermediaries on loans to Microborrowers | None | Self-reported Microborrower information, reported up to 3 months after Microloan approval | Individual Microloan Approval | Microborrower Name, SSN | N/A | Medium: Data is self-reported from lender-intermediaries; there is no QA on the data | SSN, Demographic, location | Microborrower and Microlender Intermediary Names captured | Microborrower and Microlender Intermediary Addresses captured | Yes | No | NAICS Code; character | Zip code | No | N/A |
| U.S. Department of Labor | Employee Benefits Security Administration (EBSA) | Form 5500 Data Sets | Owned | The Form 5500 Data is updated on the website first business day of each month. | foarequest@dsi.gov | www.doi.gov/agencies/ebsa/about-ebsa/about-activities/public-disclosure/foia-form-5500-datasets | catalog.data.gov/datasets/foia-form-5500-datasets | www.ebsa.doi.gov/portal/apps/disseminate?executive=211 | Data is used for compliance, enforcement, and research. The Internal Revenue Service (IRS), Department of Labor (DOL), and Pension Benefits Guaranty Corporation (PBGC) have consolidated certain returns and report forms to reduce the filing burden for plan administrators and employers. Employers and administrators who comply with the instructions for the Form 5500 generally will satisfy the annual reporting requirements for the IRS, and DOL, and PBGC. | The data sets are shared by IRS, PBGC, and DOL. Some data containing Personal Identifying Information (such as SSN) will not be available. | Over 700,000 Pension and Welfare filings for a given plan year. Specific plan information are attached (participant, asset, etc.). A return/report must be filed every year for every pension benefit plan, welfare benefit plan, and for every entity that files as a DFI as specified below (pursuant to Code section 6058 and ERISA sections 104 and 4065). | Plan based unit of analysis | Usually based on entire firm, not on a single establishment. Firms will be reported in data base as plan sponsors or plan service providers. | Sponsor to benefit plan(s) structure. Same sponsor may provide multiple plans. Each company may be represented in various plans. If any connection can be identified, it's usually through the EIN | Name fields are not very dependable. Numerous different names can be spelled, shortened, etc for 1 given plan or sponsor. Ack ID and EIN fields are primarily used. Other variables such as participant, asset, and plan characteristic fields are used to ensure quality. Various edit checks are performed to ensure no duplicate filings are tabulated. | EIN, Ack_ID | Yes; Plan name, Sponsor name, Plan administrator name. | Yes, all are simple mailing addresses and not where the plans are established; Sponsor mailing address and Plan administrator mailing address. | Theoretically, the data can contain non-US address. If outside US, enter the foreign routing code if applicable. Leave U.S. state and zip code blank if entering a foreign routing code and country name. | No | 6-digit NAICS codes; | No | Possibly match name fields and EINs to other sources. Not sure about the resulting quality. | Page 35 from the following pdf has data variables: www.doi.gov/agencies/ebsa/researchers/data/retirement/pension-user-guide-2013.pdf |
| U.S. Department of Treasury | Internal Revenue Service - Statistics of Income | Business Master File | Owned | Data are updated continuously; Data becomes available weekly. | sis@irs.gov | Not applicable | Data.gov doesn't host any of our data, it just points at our pages on IRS.gov. As far as us pointing out to data.gov, there's only a couple places where we direct users out to the Data.gov site through Tax Stats on IRS.gov. | www.irs.gov/uz/tax-stats | IRS: The purpose of data collection is mainly for determining Federal tax liability for businesses required to file. SOI: The data are used to produce statistics on income, deductions, credits and other taxes, as reported by businesses. The current design is a probability sample stratified by business type (as indicated by the IRS form filed) and either by size of total assets alone or size of total assets and a measure of income. | Federal Tax Information (FTI) is confidential (Title 26, US Code) and shared with other government agencies under IRC 6109(j) provisions. | IRS: All Federal business tax returns that meet IRS filing requirements. SOI: Selected active Federal business tax returns based on SOI's sample design. | Tax Return. | Definition of "company" is based on Title 26 requirement. | The dataset consists of individually filed returns which can represent: corporations (a single establishment; subsidiary establishment or consolidated filed returns representing a group of establishments); partnerships and other pass-through entities; or sole proprietorships. | Most matching is accomplished using the EIN provided by entities as reported on Federal tax returns. SOI uses exact matching with EINs and data processing begins with information already extracted for IRS administrative purposes. SOI performs limited internal "data cleaning" for statistical purposes. This includes organizing data to make it structurally consistent, coding data items to make them analytically useful, and validating values to ensure mathematical consistency. Contact information is validated as part of routine administrative processing of returns at the time they are received by the IRS. | Employer Identification Number (EIN) | Yes; Parent Company Name | Yes; Tax Filing Address, City, State Abbreviation, Zip Code | - | - | 6-digit NAICS codes | - | - | The focal point of the answers provided is based on data collection for Statistics of Income (SOI) unless otherwise noted. |