COMMERCIAL BUILDING STOCK ASSESSMENT 4

Database User Manual

TOGETHER We Are Transforming the Northwest

Table of Contents

Introduction

From 2017 to 2019, Cadmus and its subcontractors (the "Cadmus team") completed the fourth Commercial Building Stock Assessment (CBSA) on behalf of the Northwest Energy Efficiency Alliance (NEEA). A broad, regional study, the CBSA assessed 932 commercial buildings across 12 building types. Since its inception, NEEA has conducted research on Northwest commercial building stock characteristics, conducting its first comprehensive, regionally representative study in 2003.

To serve as a resource for efficiency planners and program designers in the region, a publicly available database contains data gathered through the study. The Cadmus team collected these data through on-site assessments of building equipment and characteristics, building staff interviews, and historical energy consumption data provided by utilities in the region.

This user manual addresses three areas:

- **Database Overview and Relationships**: This section provides guidance on how to access data collected during the CBSA.
- **Estimation:** This section provides database users with guidance regarding formulas they should use to estimate CBSA metrics for strata or subpopulations of interest.
- **Importing into Microsoft Access:** This section provides instructions to users who wish to import data into a Microsoft Access database.

Please direct questions and inquiries regarding these data to NEEA's Market Research team.

Database Overview and Relationships

This section includes several resources that may be helpful for users of the database:

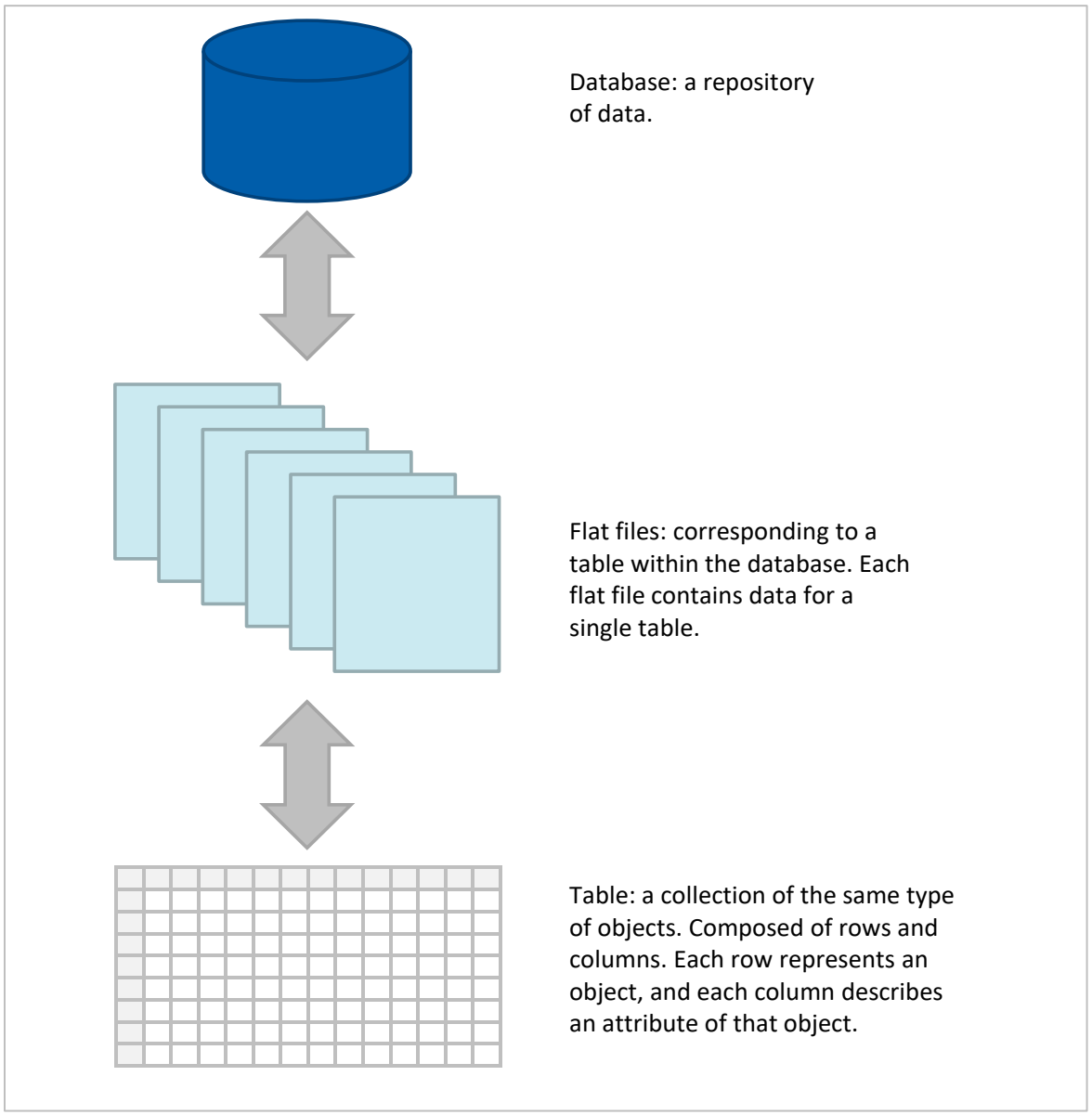
- A synopsis of the database structure and relationships
- Definition of the objects within each table in the database
- An example entity relationship diagram that visually depicts the relationships between the datasets housed within the database

In addition to this document, a data dictionary is provided in Microsoft Excel format as an appendix. The data dictionary provides definitions for each field in the database.

Database Overview

The CBSA database is a relational database provided as a collection of comma separated values (CSV), or flat, files. A relational database presents information in a collection of tables using rows and columns. Each table serves as a collection of objects of the same type. Each of the CSV files consists of a single table. Columns within each table represent the attributes for that set of objects. Figure 1 illustrates the database concept.

Figure 1. Relational Database Components



Tables in the CBSA database relate to one another using keys or unique identifiers. Each row in a table has a unique primary key (PK) making it possible to reference that specific object, and may include other relational keys, or foreign keys (FKs), that make it possible to relate to other tables in the database.

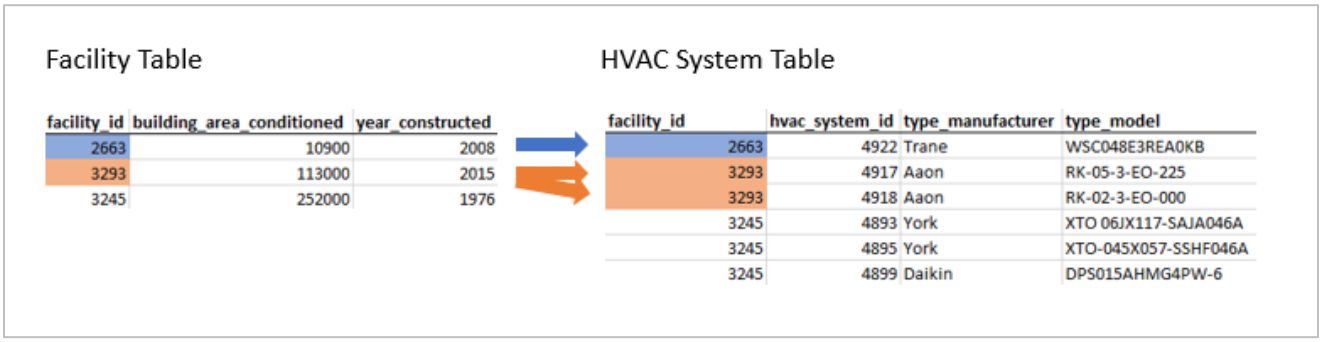
For each site surveyed during the CBSA, field assessors collected a variety of data at different levels. The assessors recorded information about the site itself, as well as information about individual facilities at the site. Within a facility, the assessors may have also sampled individual businesses, rooms, and miscellaneous spaces. The assessors also collected information about objects within the building, such as HVAC systems and

lighting fixtures. Many objects have their own set of attributes independent of the space in which they are located. For example, lighting characteristics (e.g., type of fixture, lamp, lamp wattage) are independent of the space itself. By relating tables using PKs and FKs, these different levels of data can be combined in a multitude of meaningful ways. For example, relating lighting fixtures to the rooms in which they are located allows the user to calculate properties such as lighting power density and the average number of lamps in each type of room.

Types of Relationships

Two primary types of relationships exist within the CBSA database: one-to-many relationships and many-to-many relationships. In a one-to-many relationship, the PK table contains only one record, which may relate to none, one, or multiple records in the related table. Figure 2 illustrates this type of relationship.

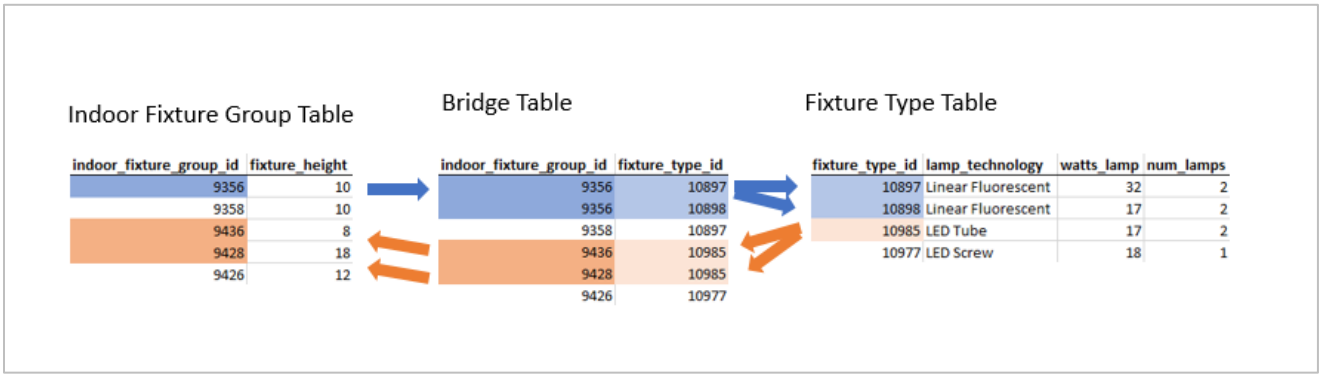
Figure 2. One-to-Many Relationship



In this relationship, an individual facility may be associated with multiple HVAC systems. The PK on the facility table is present as an FK on the HVAC system table. This relationship shows which HVAC systems are located within a given facility.

In a many-to-many relationship, each of the related tables contains only one record. Each of the two related tables is also related to a bridge table (also known as an associative table, join table, or junction table). The bridge table allows a single record in the first table to relate to many records in the second table as with a one-to-many relationship, while also allowing a single record in the second table to relate to many records in the first table. Figure 3 illustrates this type of relationship.

Figure 3. Many-to-Many Relationship



For example, the indoor fixture group table stores information about specific groups of lighting fixtures within a space. The fixture type table stores information about the types of lighting within these groups. Because more than one type of lighting may be located in a given group of fixtures and a particular type of lighting may be found in more than one fixture group, there is a many-to-many relationship between fixture groups

and fixture types. In Figure 3, the indoor fixture group highlighted in blue contains two different types of linear fluorescent lamps: 17 watt and 32 watt. Looking at this type of relationship in the opposite direction, the fixture type highlighted in orange is associated with two indoor fixture groups: one with a fixture height of 8 feet and another with a fixture height of 18 feet.

Sites, Facilities, Businesses, and Spaces

A site is the highest level of organization in the database. All collected data are grouped under a site. Some basic information that applies to the entire site is reported in the site table. Most data are collected within a facility. The following facility types are available:

- **Building**: Most data are collected in building facility types. Every site will have one or more buildings, and many sites will only have building facility types.
- **Central Plant**: If a site has a central plant that provides services to other buildings on the site, and the central plant itself was not sampled as a surveyed building, then hydronic systems in the central plant may be grouped under a central plant facility type.
- **Site Lighting**: If there is exterior lighting at the site that is not attached or adjacent to an existing building entry, it may be grouped under a site lighting facility type. Site lighting will only contain exterior lighting.

There are three types of spaces which can be defined within a building. Other data tables contain relationships to these space tables to group related equipment together or to document within-building sampling.

- **Business** entries are created when there are multiple businesses within one building. If business sampling occurred, there may be fewer business entries in the data than the total number of businesses present in the building. Refer to the *We calculated* within-site sampling weights based on selection probabilities derived from within-site sampling plans. The database reports the within-site sampling weights for each building, business, and room, and users may combine them with site case weights to calculate regional metrics. Additionally, the database includes the combined weights, which users may apply to calculate regional metrics directly.

All levels of sampling were subject to the practical limitations of the site visit, including safety and security restrictions and participant agreement. In some cases, the team could not assess businesses or rooms selected by the sampling plan. Technicians attempted to sample replacement spaces whenever possible while on site. We adjusted the sample weights reported in the CBSA database to reflect the actual spaces visited during each site assessment.

- Building Sampling section of this manual for additional details.
- **Room** entries are only created if room sampling was conducted within a business or building. Each room entry represents a discrete room within the building. Room entries are always subject to sampling weights. Not every physical room within the building or business is recorded in the data. Rooms are used primarily in conjunction with lighting groups to define where lighting is installed, although rooms may also be associated with general office equipment, food service equipment, and HVAC systems.
- **Miscellaneous Spaces** are created if room sampling was not conducted. A miscellaneous space is a grouping of floor areas with a common space-use type. A single miscellaneous space may encompass multiple similar rooms. Miscellaneous spaces are only used in relation to lighting groups and do not necessarily represent all of the floor area within a building or business.

Some information is collected at the building or facility level and is not related to any spaces. Other data may be related to a business, room, or miscellaneous space depending on how within-site sampling may have been applied. See the *Example Entity Relationship Diagram* section for more information on how to identify relationships between equipment, spaces, and sites.

Non-Response Flags

The CBSA data collection effort included over 500 separate variables with detailed information about every type of equipment. Field staff were not able to fill out every field in every situation. Each column in the database has a corresponding flag column. If there is a NULL value in a column, the adjacent flag column will have one of the flags listed below to indicate why that value was left blank:

- **Skipped** indicates that the question was automatically skipped in the data collection software due to a previous answer. In many cases this indicates that the question did not apply. For example, the heating fuel will be skipped on an HVAC system that only provides cooling. It may also appear if the technician marked a previous question as Not Applicable or Unable to Determine, and the current question relies on a previous question. For example, if the technician marked the heating efficiency units as Unable to Determine, the heating efficiency value will be skipped. Skipped is the only flag that is created automatically by the data collection software. All other flags are selected manually by the field staff completing the assessment.
- **Not Applicable** indicates that the field technicians determined that the question does not apply to the given situation. For example, in the Building Controls table, if the building has multizone HVAC systems but does not have any multizone HVAC reset strategies enabled, the technician may have selected Not Applicable when prompted to select all reset strategies that are in use.
- **Unable to Access** indicates that the field technician was not able to physically access the equipment to answer the question. This generally means that the equipment was either in a restricted area, could not be safely reached, or that labels or nameplates could not be seen without disassembling the equipment.
- **Unable to Determine** indicates any other situation where the field technician could not identify the answer to the question. For example, a nameplate might be too faded to read or the site contact might not know how a piece of equipment operates.

Database Contents

Object Definition

As previously discussed, each database table serves as a collection of details for objects of the same type. Columns within each table represent the attributes for that set of objects. These tables are organized into groups—for example, the hydronic systems group includes tables for boilers, chillers, and water heaters. Table 1 describes the type of objects housed within each table group. For more detailed information about each individual table, refer to the data dictionary.

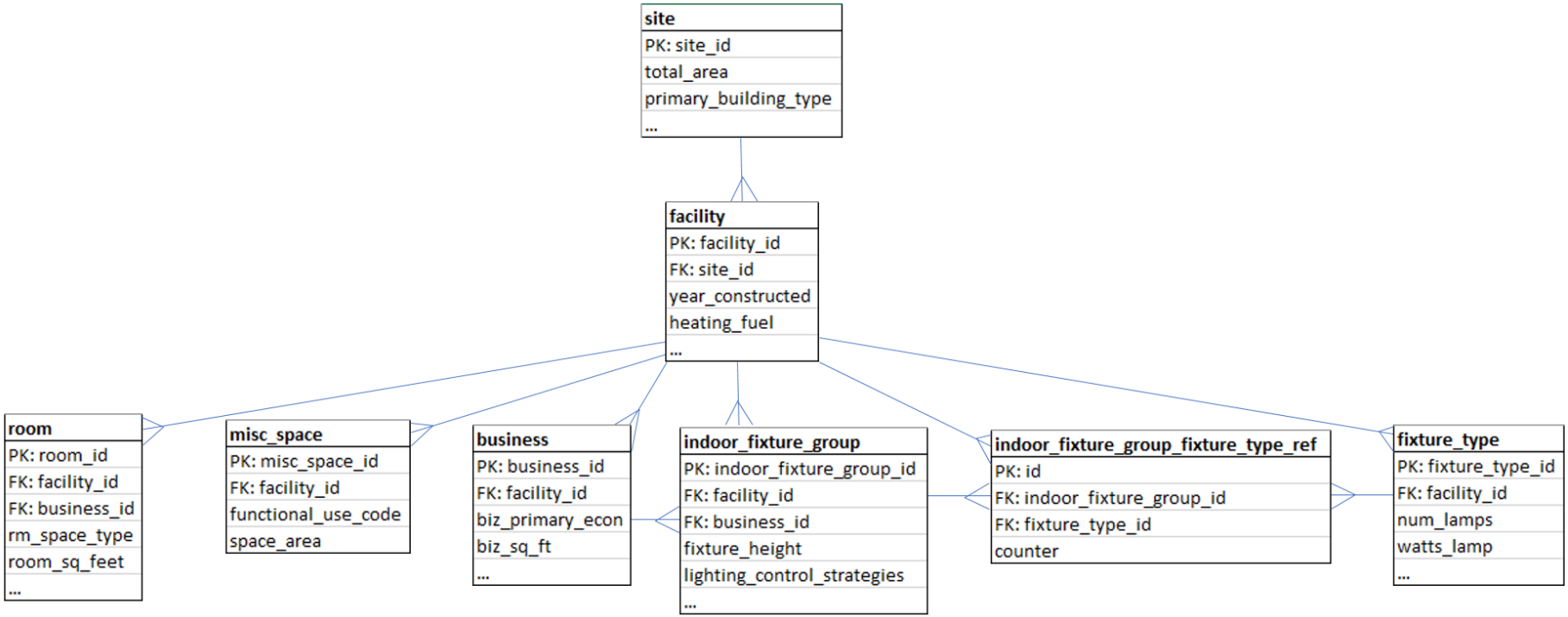
Table 1. Object Description by Table Group

Database Table Group	Definition of Objects (Rows) within the Table Group
site_details	Information at the site or facility (building) level. These tables contain general information about the site such as state and sample weight. They also contain summary tables, detailed in the *Building Summary Table* section.
spaces	Information about spaces below the facility level. These include businesses, rooms, and miscellaneous spaces.
general_building_information	Additional information at the facility level, including building controls and on-site generation.
building_envelope	Information about building envelope components including walls, roofs, windows, and floors. Only includes exterior walls and floors.
data_centers	Information about any data centers/server rooms at the site.
hvac	HVAC system components. In some cases, a quantity field is used to indicate the presence of multiple, identical units.
hydronic_systems	Hydronic systems, including boilers, chillers, and water heaters. In some cases, a quantity field is used to indicate the presence of multiple, identical units.
lighting	An entry that represents a unique light bulb type/base type/shape/wattage. Each entry may represent one or more individual light bulbs with the same characteristics. A quantity field is provided.
miscellaneous_equipment	Information about miscellaneous pieces of equipment, including laptop and desktop computers, TVs, and dishwashers. In many cases, a quantity field is used to record the number of appliances of the same type.
refrigeration_equipment	Information about refrigeration equipment, including walk-in and display cases. Length, depth, and height are provided where relevant.

Example Entity Relationship Diagram

An example of an entity relationship diagram that visually depicts the relationships between the tables in the database is provided in Figure 4. Relationships between tables are indicated by blue connections. The diagram shows relational keys and depicts one-to-many relationships by a single line from the table with one record branching into three lines on the table that may have many associated records. These relationships are standardized across the database (i.e., they are the same for all equipment type tables). Bridge tables are always indicated by the common suffix, "_ref".

Figure 4. Example Entity Relationship Diagram



Aggregating Data

Users can aggregate or roll up data using relational keys in each table. This section provides an overview of the data aggregation process and illustrates the process through an example.

Note: These instructions are for unweighted data only. Though the aggregation process is similar for weighted data, applying appropriate weighting to each line item requires additional steps. See the *Estimation Method* section for more information.

Aggregating Data to the Building Level

The CBSA database captures some data on a room-by-room basis, such as information about lighting. Other information is captured at the building level, such as floor type, window type, and wall type. To understand total quantity and averages for a given level of granularity (e.g., site or building), cells must be summed or averaged. For example, wall information is captured with one wall record for each type of wall with a different set of characteristics (e.g., framing type), so a single building may have multiple types of walls. To determine the total net wall area in a given building, sum all wall areas where the facility_id matches a specified value, as exemplified in Table 2.

Table 2. Sample Wall Data

walls_id	site_reference_number	facility_id	walls_framing_type	wall_area_net (sq. ft.)
3180	14352	2810	Wood Frame	540
3181	14352	2939	Concrete	1,200
3182	14352	2939	Wood Frame	1,800
3183	14352	3337	Wood Frame	3,760
3184	14352	3337	Concrete	660

Table 3 shows the totals of summing wall_area_net across each building.

Table 3. Sum of Net Wall Area by Building from Sample Wall Data

facility_id	wall_area_net (sq. ft.)
2810	540 = 540
2939	1,200 + 1,800 = 3,000
3337	3,760 + 660 = 4,420

Aggregating Data to the Site Level

The steps for aggregating data to the site level are roughly the same as those used for aggregating data to the building level. The primary difference between building-level and site-level aggregation is that aggregation occurs based on site_reference_number rather than facility_id.

Revisiting the previous example, Table 4 shows the total net wall area at site 14352.

Table 4. Sum of Net Wall Area by Site from Sample Wall Data

site_reference_number	wall_area_net (sq. ft.)
14352	540 + 3,000 + 4,420 = 7,960

Aggregating Data to Different Levels

To aggregate data using grouping variables different from site_reference_number or facility_id, apply the same concepts discussed in the preceding sections, making sure to identify appropriate grouping variables.

Estimation Method

This section provides an overview of how users should apply the weighting estimation developed for the CBSA. The overview includes a brief review of the sampling approaches used as the basis for weighting, and contains instructions on how database users can apply the weights to estimate building, site, and regional values from sampled data.

Regional Case Weight Development

The Cadmus team developed regional case weights as a function of the inclusion probability of each CBSA site, the sample design, recruitment metrics, and within-site sampling weights. More detail on the regional case weight development process can be found in the 2019 CBSA final report. We provide a variable called sample_weight_site to represent the regional sampling weight for the site. A user can multiply that by the metric of interest to weight up to the regional level.

Within-Site Sampling Weights

We calculated within-site sampling weights based on selection probabilities derived from within-site sampling plans. The database reports the within-site sampling weights for each building, business, and room, and users may combine them with site case weights to calculate regional metrics. Additionally, the database includes the combined weights, which users may apply to calculate regional metrics directly.

All levels of sampling were subject to the practical limitations of the site visit, including safety and security restrictions and participant agreement. In some cases, the team could not assess businesses or rooms selected by the sampling plan. Technicians attempted to sample replacement spaces whenever possible while on site. We adjusted the sample weights reported in the CBSA database to reflect the actual spaces visited during each site assessment.

Building Sampling

At any site with more than three buildings, we considered within-site building sampling when there was insufficient time to collect data from all buildings. We selected any building that accounted for more than one-third of the site total floor area with certainty, and selected other buildings with probability proportional to size. The Cadmus team calculated building sampling weights as the inverse of each building's selection probability. Users can apply the building sampling weights to building metrics within a site to estimate site totals and can apply site weights to site totals to estimate regional totals.

The building-level sample weight is provided in the sample_weight_facility column of the facilities table, and the regional expansion weight for each building is provided in the facility_expansion_weight column of the facilities table. These columns are blank for rows representing facilities which are not buildings (site lighting or central plant facilities).

Business Sampling

Within any building with multiple businesses, we considered business sampling when there was insufficient time to collect data from all businesses. If one business accounted for more than 50% of the floor area, we selected it with certainty. We selected a simple random sample of the remaining businesses from within each business type category present. We calculated business sampling weights as the inverse of each business' selection probability. Users can apply the business weight to metrics measured at the business level within a building to estimate total building metrics, and can then apply the building weight to estimate site metrics and the site weight to estimate regional metrics. The business table in the database lists the business sampling weights.

Business weights may be related to equipment tables in three ways:

- Some tables contain a foreign key to a single business, indicating that the information described in that table only applies to a single business. To get the business weight that applies to these tables, use the business_id column in the table to look up the corresponding row in the business table, then apply the sample_weight column from the business table to the metric of interest. The following tables reference businesses in this manner: pool_and_hot_tub, laundry_equipment, lab_equipment, lodging_information, medical_facility_information, reach_ins_display_cases, refrigerated_walk_ins_storage_boxes, refrigeration_compressors, and refrigeration_condensers.

- Some tables represent equipment that could serve multiple businesses. For these tables, the business ID relationship is defined in a bridge table. To get the business weight for these tables, find the corresponding bridge table using the data dictionary, where the "primary table" is the table of interest and the "referenced table" is the business table. Find all records in the bridge table matching the ID for the table of interest, look up the corresponding business ID in the business table, then use the sample_weight column for those businesses. The following tables reference businesses in this manner: onsite_generation, building_controls, water_heaters, and air_compressor.

- The remaining tables that reference business equipment also involve room sampling, and the business weight has been incorporated with the room weight as described below.

Room Sampling

We conducted room sampling within any building with more rooms than we could individually visit during the site visit. If one room accounted for over 50% of the floor area in the business or building, we selected it with certainty. We selected a simple random sample of the remaining rooms from within each of 11 space type categories. We calculated room sampling weights as the inverse of each room's selection probability.

Aggregated weights, accounting for room sampling, business sampling, and all other data relationships between tables, are provided directly on the hvac_summary, hvac_system, general_equipment, and food_service_equipment tables. Users can apply the sample_weight column (sample_weight_hvac on the hvac_summary table) directly to the metric of interest to estimate building-level values. Users can then apply the building weight to estimate site-level values and apply the site weight to estimate regional values.

Sampling weights for lighting are listed in the lighting summary table, which joins together data from all of the lighting tables and is described in further detail in the Summary Tables section below. We did not report weights directly on the underlying lighting tables because the fixture type and lighting group tables must be considered together due to the complex database relationships between them. Users may join the outdoor_lighting_group, indoor_lighting_group, and/or fixture_type tables to the lighting_summary table, and then use the sample_weight_lighting column from the lighting_summary table to estimate building-level values. Users can then apply the building weight to estimate site-level values, and the site weight to estimate regional values.

The sample_weight_lighting column in the lighting_summary table is the product of the related_spaces_weight column and the sqft_correction_ratio column. The related_spaces_weight column is the aggregated weight of the businesses and rooms audited in the lighting group represented by a row in the lighting_summary table. The sqft_correction_ratio column adjusts all lighting weights in the building to account for discrepancies between the weighted sum of sampled floor area and the actual total floor area in some buildings. We provided a detailed description of how we calculated the lighting weights in Appendix D of the 2019 CBSA report.

Outdoor Lighting Sampling

Field technicians collected all outdoor lighting on site when possible. When it was not practical to collect all outdoor lighting, technicians recorded the fraction of the total linear feet (for building perimeter, narrow walkways, and driveways) or square feet (for other space types) of outdoor lighting audited in each outdoor lighting group. The sample weight is the inverse of the percent of the lighting audited and is reported in the lighting_summary table in the sample_weight_lighting column. The correction factor does not apply to outdoor lighting because room sampling was only used indoors.

Data Center Sampling

Data center spaces represent the one variance from the within-site sampling approach described above. For these spaces, field technicians compiled the total floor area for all data center spaces (e.g., server rooms, server closets, localized data centers). The data collection protocols required the field assessor to select the largest data center space at the facility for auditing purposes. The assessor did not gather any additional data on other data center spaces except their floor area. There are no within-site sampling weights to apply with data center spaces.

Summary Tables

In addition to the detailed database tables described in the previous chapter, the CBSA database also includes three summary tables. These tables contain basic analysis and aggregation for each site, summarizing key variables of regional interest. The data dictionary defines individual columns in each summary table.

Building Summary Table

The building summary table contains one row per building in the study. Numerical totals and percentages for building-level values (e.g., areas, perimeters, and capacities) are summed across all underlying elements. Some HVAC system, boiler, and chiller columns are reported for the predominant and secondary system. For these values, all underlying systems (e.g., all HVAC systems) are grouped by the value being summarized. The individual system capacities are summed within each group, and the groups are then ranked by capacities. The grouping with the largest total capacity is reported as the predominant system type, and the group with the second largest total capacity is reported as the secondary system type, if applicable. Note that the predominant and secondary values are calculated independently for each column in the summary table, and the predominant category may vary depending on which value is being summarized. The building summary table can be joined to other tables using the facility_id column.

HVAC Summary Table

The HVAC summary table contains one row per HVAC system and may contain many rows per building or no rows if the building is unconditioned. Many of the columns in the summary table are identical to the data columns found in the hvac_system data table. Other columns add additional data summarized from relationships with other tables or calculated from the data in the HVAC table. Each column is described in detail in the data dictionary. The HVAC summary table can be joined to the hvac_system table using the hvac_system_id column present in both tables.

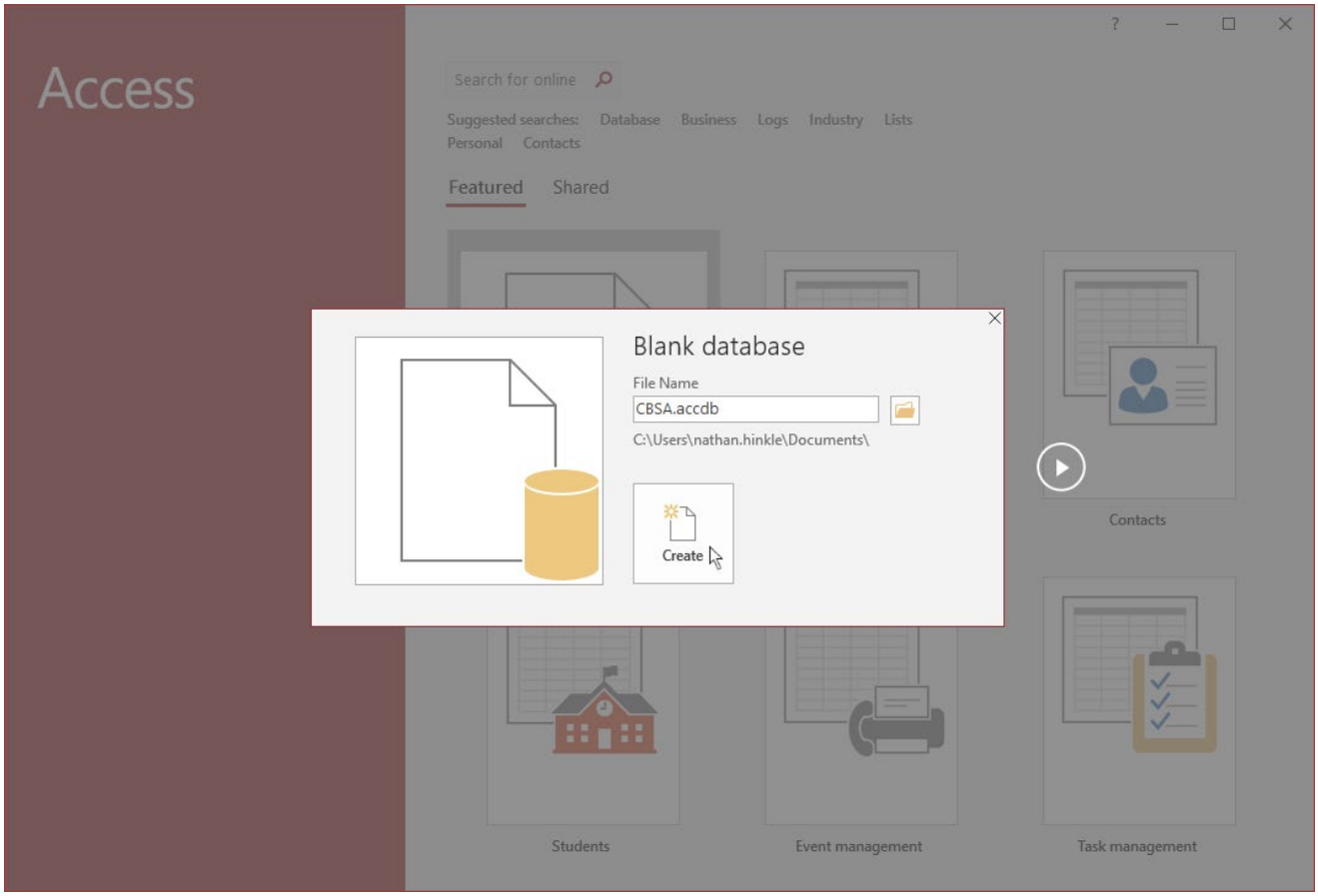
Lighting Summary Table

The lighting summary table contains one row for each fixture type within each indoor or outdoor lighting group. If a fixture type appears in two indoor lighting groups and one outdoor lighting group, it will appear on three rows in the lighting summary table, along with the lighting group details. If an indoor lighting group has two fixture types in it, there will be one row for each. The lighting summary table can be joined to the fixture_type table using the fixture_type_id column present in both tables. It can also be joined to the indoor_lighting_group and outdoor_lighting_group tables using the fixture_group_id column in conjunction with the location_type column. Be careful to join to the correct table, because the same ID could appear in both the indoor_fixture_group table and the outdoor_fixture_group table representing separate, unrelated groups.

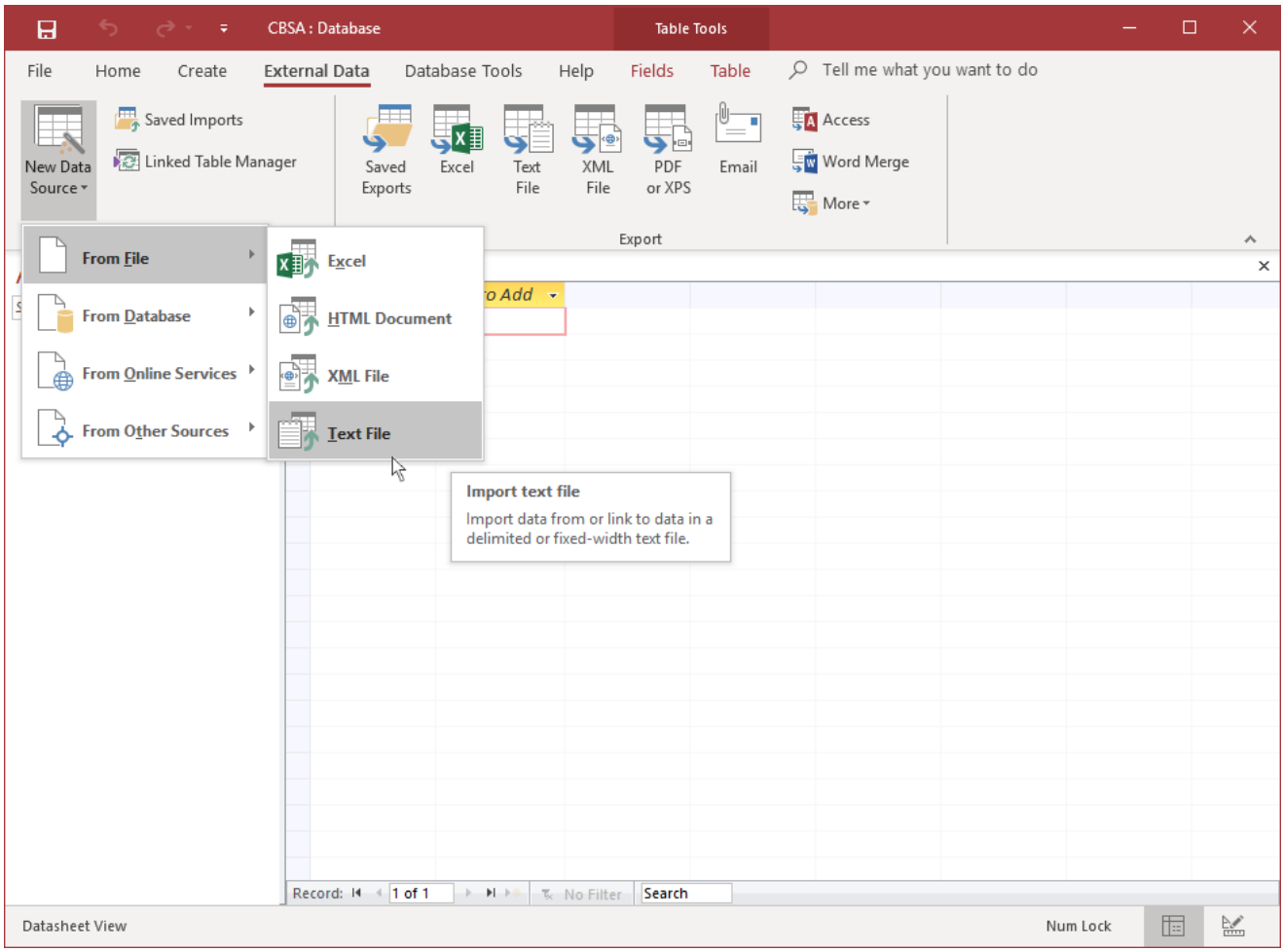
Importing into MS Access

As noted, the CBSA database is made available as a collection of CSV files. Some users may wish to import these data into a Microsoft Access database—a format used for previous CBSA databases. Use the process below to import individual tables of interest into Microsoft Access. These instructions are written for Microsoft Office 365, but the same import options can also be used in older versions.

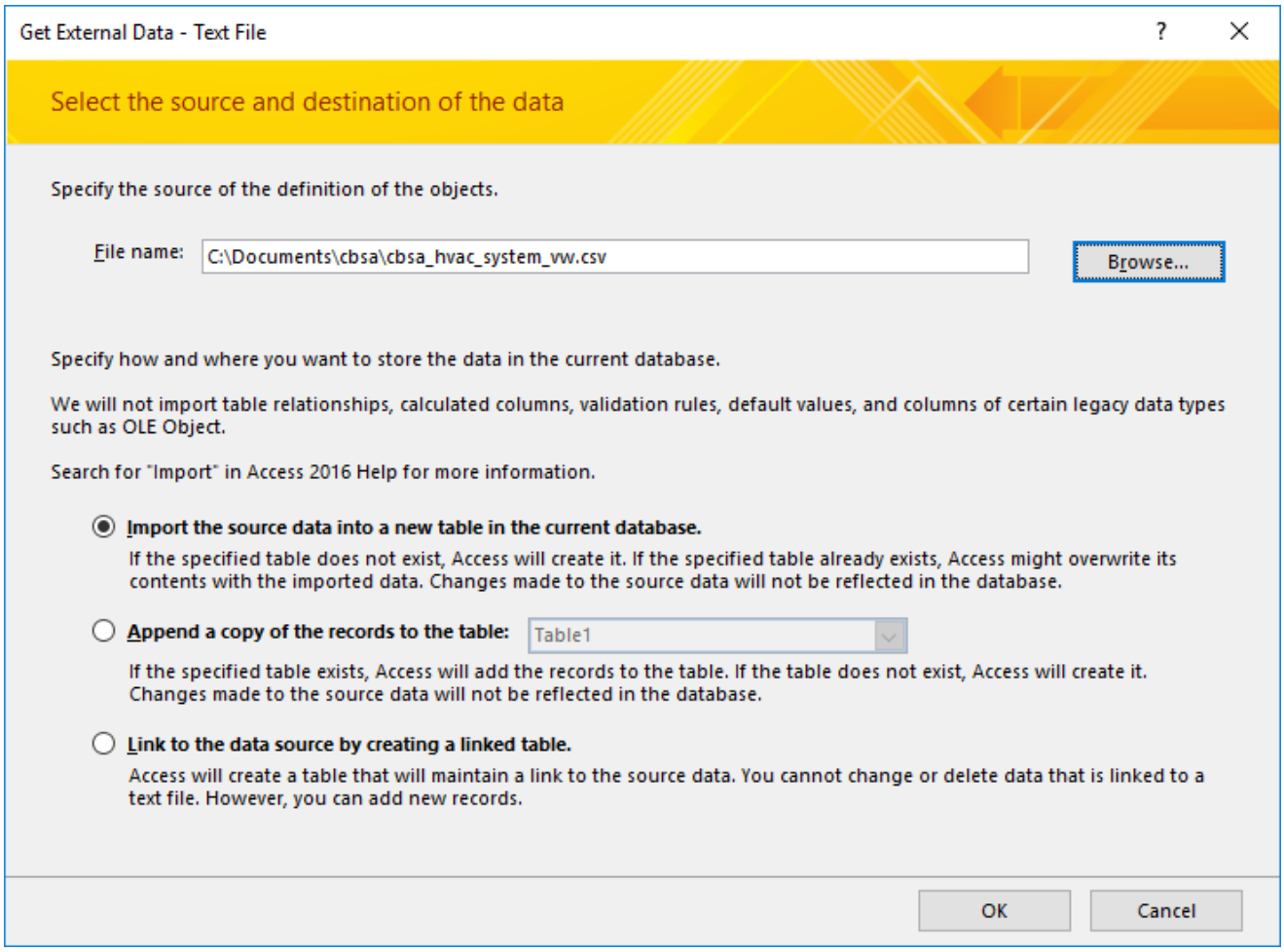
Open Access and create a new blank database:



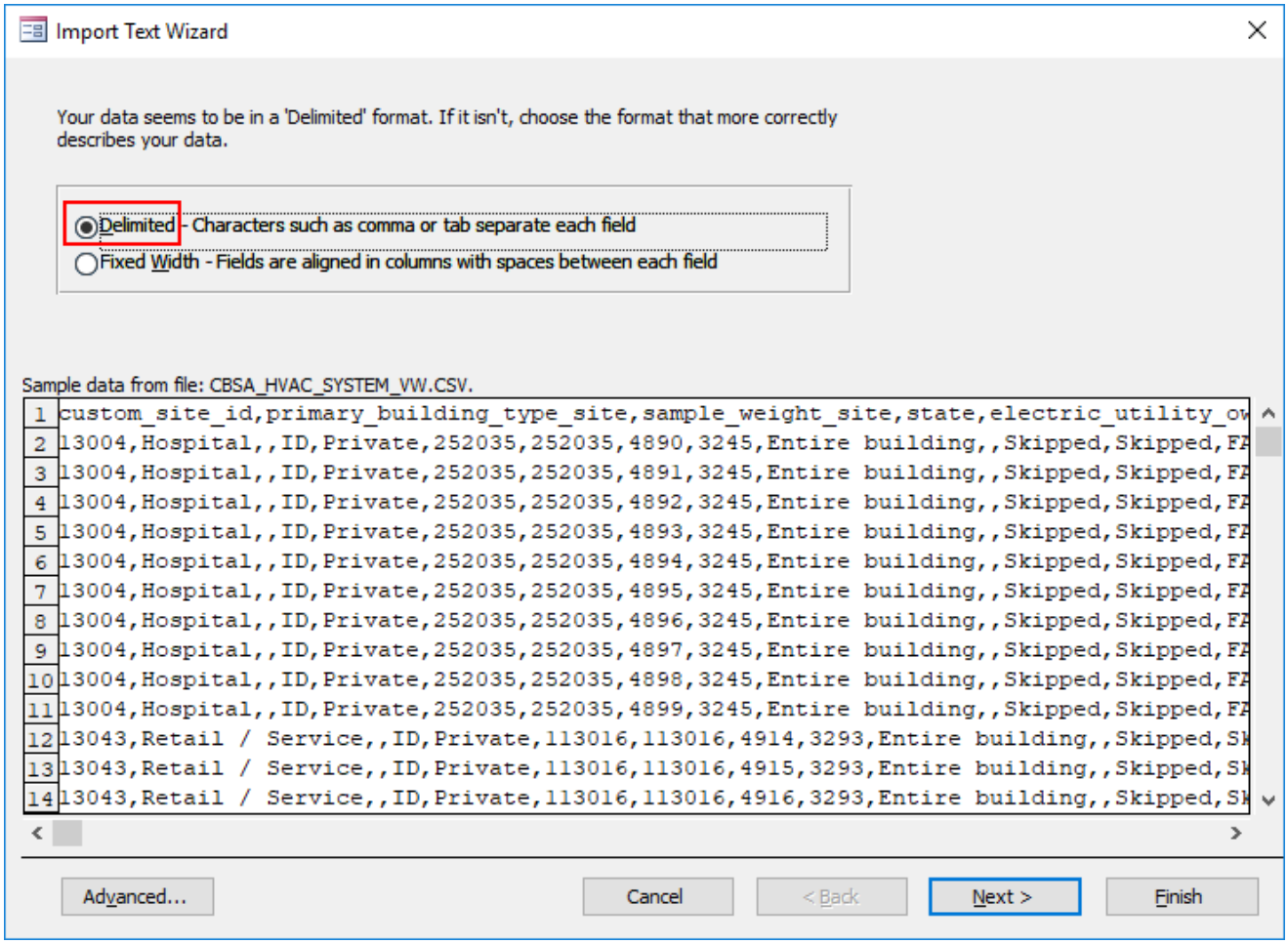
Click New Data Source > From File > Text File



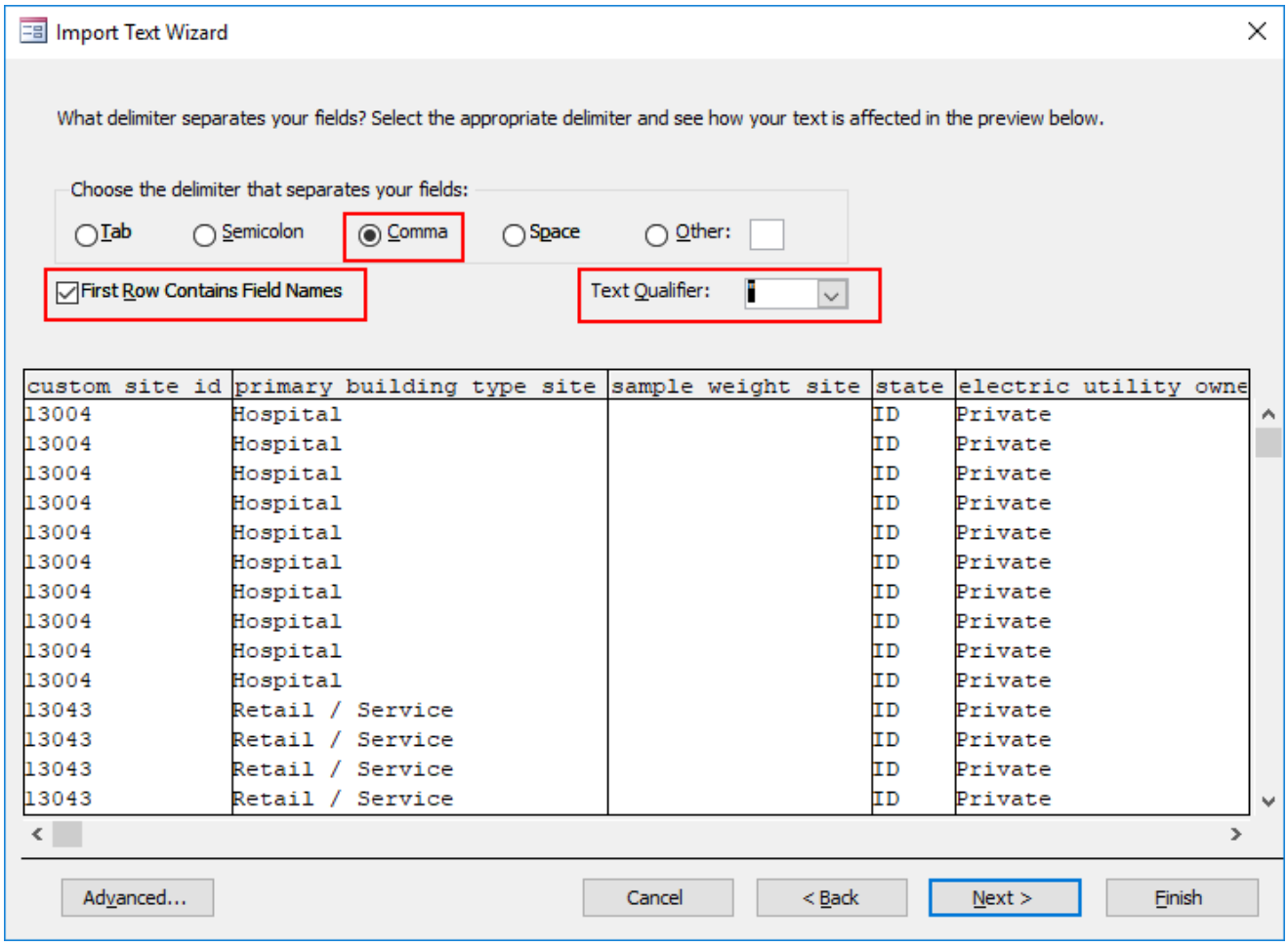
In the Get External Data dialog, browse to the location of the desired CBSA CSV table. Select “Import the source data into a new table in the current database.”



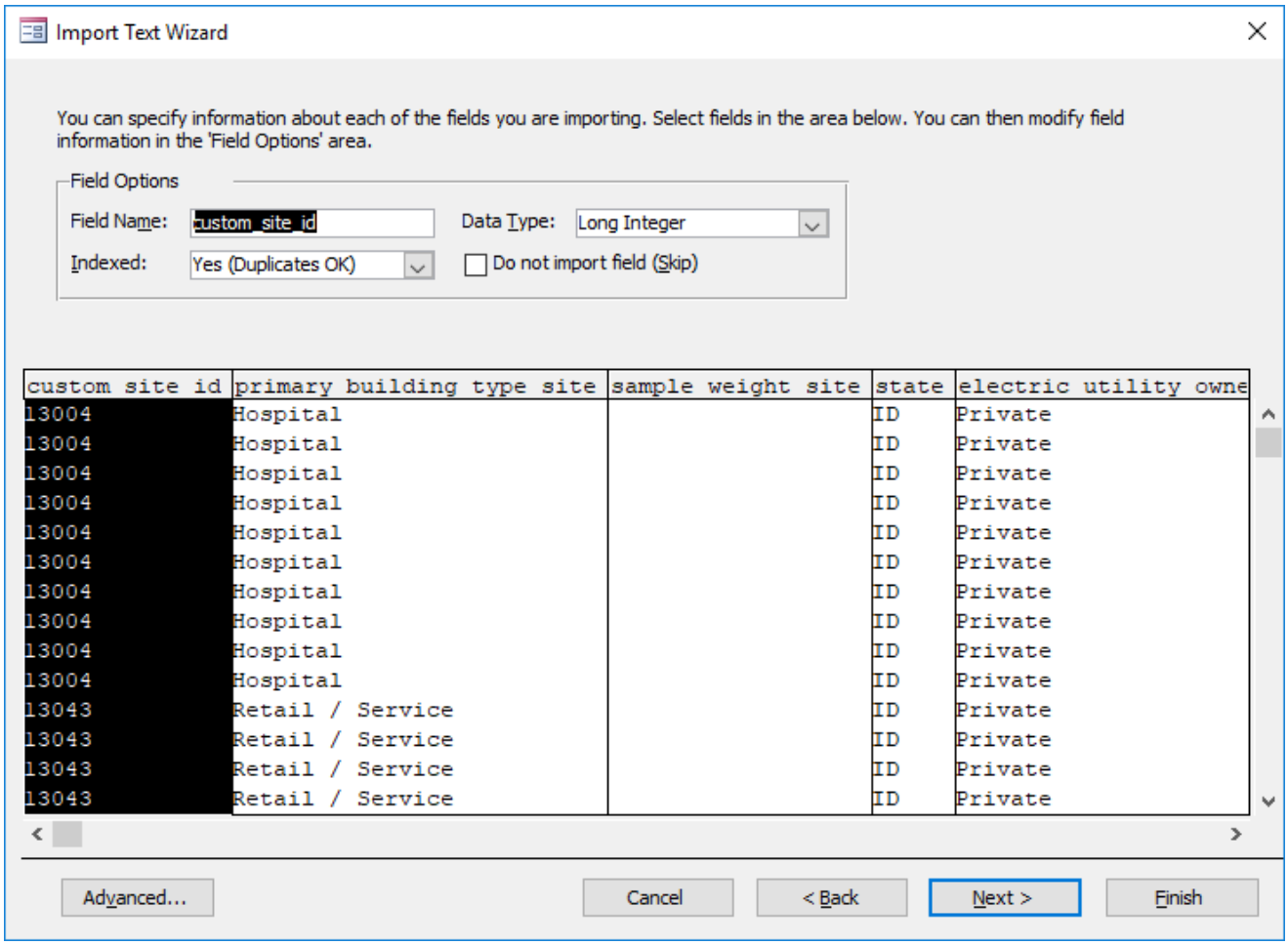
In the next screen of the Import Text Wizard, select “Delimited” for the format, then click Next.



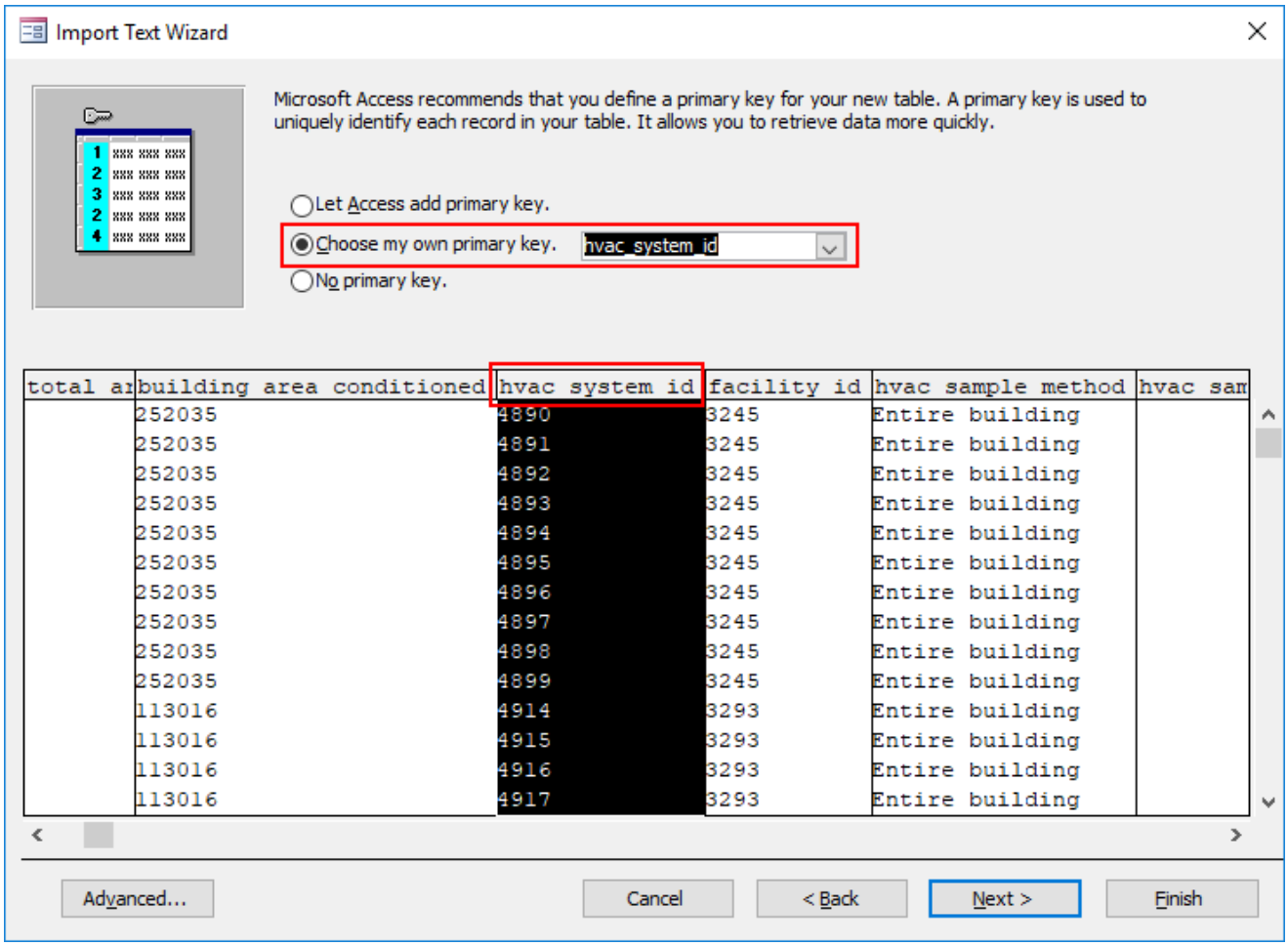
On the next screen, select “Comma” for the delimiter, select the quotation mark (") for the text qualifier, and check “First Row Contains Field Names”, then click Next.



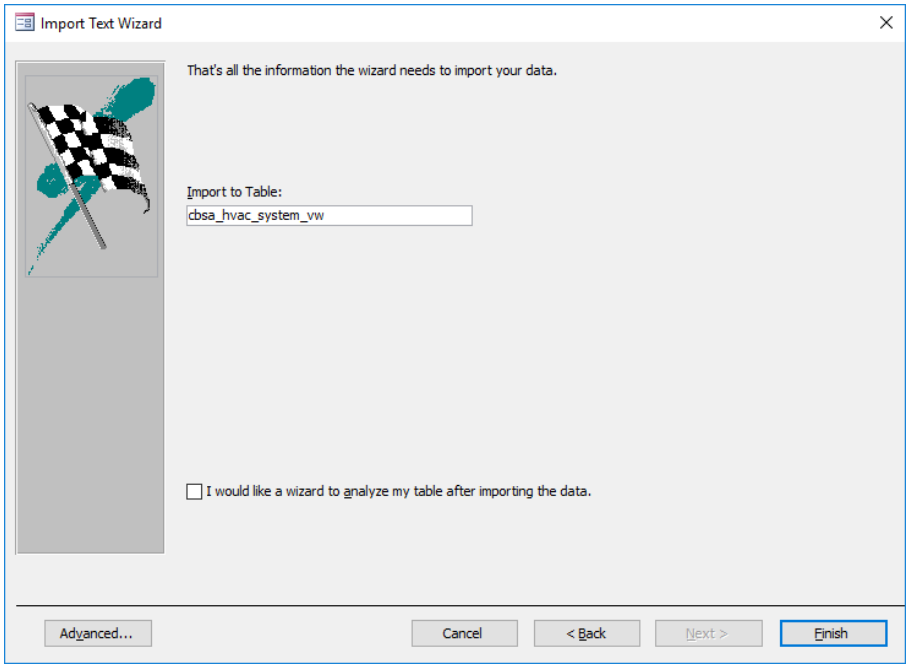
The next screen will show a preview of the imported data. It is not necessary to make changes to this page, but it possible to configure database indexes and change column data types if desired for analysis. Click Next to proceed with the import.

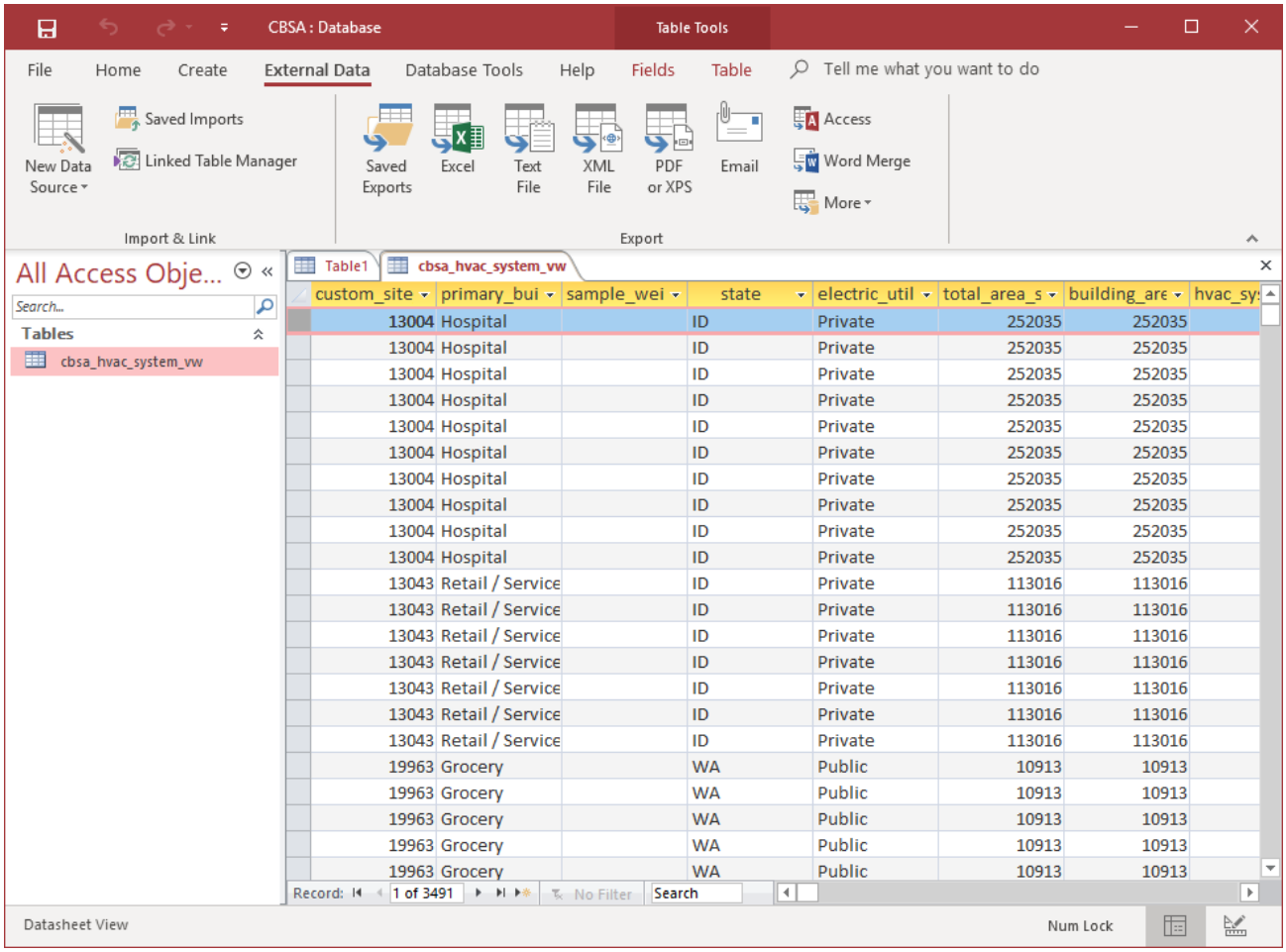


The next screen allows a primary key to be defined for the table. Select the option for “Choose my own primary key,” and select the ID column for this table. The ID column is the first column in each table. This will usually be a column matching the name of the table plus “id” (e.g., “hvac_system_id” for the HVAC Systems table).



Finally, set the name for the imported table, and click Finish to view the table.





Repeat this process as necessary to import additional tables into Microsoft Access.