IM3 Open Source Data Center Atlas

Description

This dataset contains locations of existing data center facilities in the United States (U.S.) derived from OpenStreetMap (OSM), a crowd-sourced database. Data points from OSM are processed in various ways to determine additional variables provided in the data including: facility area (square feet), associated U.S. county, and U.S. state. This dataset can be used to identify areas of concentrated data center development and inform government and private sector planning strategies for future buildout of data centers and the necessary supporting infrastructure.

Usage Notes

- Validation of OSM-derived data center locations is an ongoing development under the IM3 project, and the database will be updated as new information becomes available.
- In some instances, both the data center area (e.g., campus) and individual data center buildings are included as overlapping areas in the database. Both values are retained.
- Data center points, buildings, and campus areas are provided as separate layers in the downloadable data package. Note that data items are not necessarily complete across layers. That is, a specific data center may only be present as a single point geometry in the "point" layer while other data centers are represented in both the campus and building layers.
- In some cases, data center campuses and/or buildings straddle a county boundary line. Mappings to both counties are retained in the database as separate rows. These data rows will have the same data center id information, but each will have different county information.
- Crowd-sourced data, by nature, relies on individuals and communities to provide information. As a result, some data may be missing where it has not yet been reported. As we collect information on additional data center locations and as OSM receives additional contributions, the database will be updated to capture additional data points not yet shown.
- Data items will occasionally be removed from OSM if they are misidentified, if they no longer exist, if they are duplicates of another item, or similar. For that reason, updated versions of this database may not contain all data center locations included in previous versions.

Technical Information

Data is available for download under the following formats:

- GeoPackage (GPKG)
- CSV

Geospatial data is provided in the WGS84 (EPSG:4326) coordinate reference system. The GeoPackage download contains the following layers. See usage notes for more information.

- “point”
- “building”
- “campus”

The “point” layer includes all data from OSM that had POINT geometry type (i.e., individual coordinates). The “building” layer includes all OSM data that did not have POINT geometry and where the building tag in the OSM export was neither equal to “no” or null. Data that did not meet the “point” or “building” qualification was assumed to be a facility campus and included in the “campus” layer.

The dataset contains the following parameters. Variables provided by OSM are labeled with (OSM-provided).

- **id** - unique identification number (OSM-provided with prefix of “node/”, “relation/” and similar attributes removed)
- **state** - name of U.S. state
- **state_abb** - two letter U.S. state abbreviation
- **state_id** - state ID number
- **county** - name of U.S. county
- **county_id** - county ID number
- **ref** - reference numbers or codes (OSM-provided)
- **operator** - the name of the company, corporation, or person in charge of facility (OSM-provided)
- **name** - name of facility (OSM-provided)
- **sqft** - surface area of facility polygon, measured in square feet. Only available for “building” and “campus” types
- **lat** - latitude of data centroid point
- **lon** - longitude of data centroid point
- **type** – represented spatial information. One of “point”, “building”, or “campus”.
- **geometry** – POLYGON geometry of area footprint (in “campus” and “building” layers) or POINT geometry of locations (in “point” layer). This parameter is not included in the csv download

Attribution

Data center locations were derived from OpenStreetMap, which is made available at openstreetmap.org under the Open Database License (ODbL).

U.S. state and county boundary information were collected from the U.S. Census Bureau for the year 2024.

Acknowledgment

IM3 is a multi-institutional effort led by Pacific Northwest National Laboratory and supported by the U.S. Department of Energy's Office of Science as part of research in MultiSector Dynamics, Earth and Environmental Systems Modeling Program.

License

Disclaimer