NOTE: The underlined text in this document indicates additions made in this revision (posted on August 16, 2024).

1 Introduction

To better understand emissions trends in the State's dairy sector, CARB staff utilized data sources from Federal, State, and Local agencies to develop the California Dairy and Livestock Database (CADD). The database contains facility-level cattle population and digester information for California dairies from 2012 to 2022. CADD is based on publicly accessible documents and improves CARB's ability to track the California dairy and livestock sector trends. This brief document describes CADD data content. For additional information, please refer to the CADD Technical Document that will be posted on the *Short-Lived Climate Pollutants Meetings & Workshops* page or contact cadd@arb.ca.gov.

2 Database Structure

The database is composed of three interconnected tables: the **Facility General Information table,** the **Facility Herd Size table,** and the **Anaerobic Digesters table**.

2.1 Overview

1. **Facility General Information Table**: Stores information about each facility in CADD, including unique identifiers, names, location information, and the associated Regional Waterboard offices.
2. **Facility Herd Size Table**: Contains herd size data of seven (7) cattle categories (described in Section 3.2) for all facilities in the Facility General Information table for every year from 2012 to 2022. Herd size data for some non-dairy cattle facilities[1] in CADD spans from 2019 to 2022.
3. **Anaerobic Digesters Table**: Contains facilities with anaerobic digesters, the year the digesters became operational, and the year they shut down (if applicable).

2.2 Relationships

Figure 1 shows the relationships between different tables. `CADDID`, a unique identifier associated with each facility in CADD, is used across all tables to connect information. While each `CADDID` appears only once in the Facility General Information table, it can appear multiple times in the other two (2) tables.

---

[1] Some non-dairy cattle facilities started submitting annual reports in 2019, as required by the Bovine General Order (R5-2017-0058). If a facility did not submit an annual report before 2019 and a) its "Order #" in the California Integrated Water Quality System (CIWQS) database is "R5-2017-0058", or b) appeared in the Vista-CA Feed Lot layer, its CADD's coverage years is 2019 to 2022. Otherwise, it is 2012 to 2022. Please refer to data sources in Section 3.1 for descriptions of CIWQS and Vista-CA.
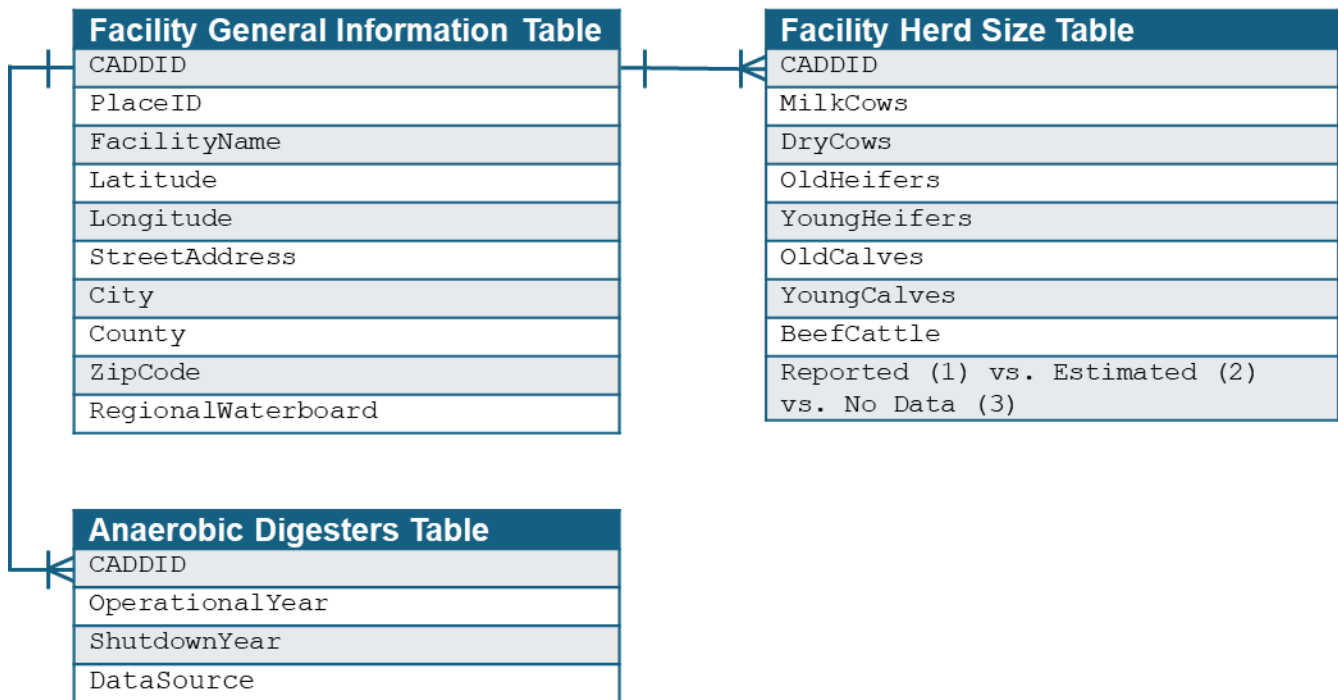
Facility General Information Table

CADDID
PlaceID
FacilityName
Latitude
Longitude
StreetAddress
City
County
ZipCode
RegionalWaterboard

Facility Herd Size Table

CADDID
MilkCows
DryCows
OldHeifers
YoungHeifers
OldCalves
YoungCalves
BeefCattle
Reported (1) vs. Estimated (2) vs. No Data (3)

Anaerobic Digesters Table

CADDID
OperationalYear
ShutdownYear
DataSource

Figure 1. Relationships between different tables in CADD.

3 Tables Description

3.1 Table 1: Facility General Information Table

- **Description**: Stores information about each facility in CADD, including unique identifiers, names, location information, and the associated Regional Waterboard offices. Facilities in CADD include those subject to reporting to the Regional Waterboards. Additional facilities are included if they a) received funding from AMMP or DDRDP (as of March 2023), b) have been issued air permits in 2019 by the SJVAPCD, or c) based on Google Earth imagery, housed cattle any time between 2012 and 2022 (for additional information, please refer to CADD Technical Document that will be posted on the *Short-Lived Climate Pollutants Meetings & Workshops* page).
- **Columns:**
    - `CADDID`: Whole number (integer) – a unique identifier associated with each facility in CADD,
    - `PlaceID`: Whole number (integer) – Place ID of the corresponding facility in the State Waterboard system, if applicable,
    - `FacilityName`: Text – Name of each facility; since facilities may change their names frequently, `FacilityName` may not reflect the current name,
    - `Latitude`: Decimal (double) – Geographic coordinate specifying the north-south position of each facility on the Earth's surface, expressed in decimal degrees,
    - `Longitude`: Decimal (double) – Geographic coordinate specifying the east-west position of each facility on the Earth's surface, expressed in decimal degrees,

- o StreetAddress: Text – The detailed location of each facility, including building number and street name,
  - o City: Text – The city associated with each facility,
  - o County: Text – The county associated with each facility,
  - o ZipCode: Whole number (integer) – A five-digit zip code associated with each facility,
  - o RegionalWaterboard: Text – The Regional Waterboard office associated with each facility.
- **Data Sources**: The sources of data used to develop this table are listed below:
  - o **California Integrated Water Quality System (CIWQS) from State Waterboard[2]**: CIWQS, managed by the California State Waterboard, is a computer system that allows the State and Regional Waterboards to track information about places of environmental interest, manage orders, track inspections, etc. CIWQS contains general information about facilities, such as unique identifiers (Place ID), facility names, and location information. CIWQS was the primary dataset used to construct the Facility General Information table.
  - o **California Climate Investment (CCI) Implemented Projects Database from CARB[3]**: provides detailed information on facilities that have received funding from several State programs, such as the Alternative Manure Management Program (AMMP) and Dairy Digester Research and Development Program (DDRDP). The database includes facility names and location information. The location information of the DDRDP and AMMP awardees in CADD was manually verified by comparing it with the location information reported in the CCI implemented projects database for DDRDP and AMMP.
  - o **Vista-CA[4]**: consists of detailed spatial maps for facilities and infrastructure in California that are known (or expected) sources of methane emissions. This dataset was created using publicly available data sources from local, state, and federal agencies for the years 2005 – 2019. Three Google Earth layers (Dairies, Feed Lots, and Digesters) from Vista-CA were used in CADD. The closest CADD facility to each Vista-CA facility was identified, and the location information was manually verified if the distance between the Vista-CA facility and its matching CADD facility exceeded 500 meters or if one CADD facility was linked to more than one Vista-CA facility.

3.2 Table 2: Facility Herd Size Table

- **Description:** Contains average herd size data of seven (7) cattle categories for all facilities in the Facility General Information table for every year from 2012 to 2022. Herd size data for some non-dairy cattle facilities in CADD spans from 2019 to 2022. The

---

[2] *Surface Water - Water Quality Regulated Facility Information - Confined Animal Facilities Regulated by the Water Boards - California Open Data* (last accessed November 21, 2023).

[3] *CCIRTS GGRF Application (arcgis.com)* (last accessed October 23, 2023).

[4] *Sources of Methane Emissions (Vista-CA), State of California, USA, https://doi.org/10.3334/ORNLDAAC/1726* (last accessed December 28, 2023).

seven (7) cattle categories include: Milk Cows, Dry Cows, Old Heifers (Heifers 15 – 24 months old), Young Heifers (Heifers 7 – 14 months old), Old Calves (Calves 4 – 6 months old), Young Calves (Calves 0 – 3 months old), and Beef Cattle.

- **Columns:**
  - `CADDID`: Whole number (integer) – a reference to the facility; linked to the `CADDID` in the Facility General Information table,
  - `Year`: Whole number (integer) – the calendar year in which the herd size is associated with,
  - `MilkCows`: Whole number (integer) – average milk cow herd size of `CADDID` during the `Year`,
  - `DryCows`: Whole number (integer) – average dry cow herd size of `CADDID` during the `Year`,
  - `OldHeifers`: Whole number (integer) – average old heifer herd size of `CADDID` during the `Year`,
  - `YoungHeifers`: Whole number (integer) – average young heifer herd size of `CADDID` during the `Year`,
  - `OldCalves`: Whole number (integer) – average old calf herd size of `CADDID` during the `Year`,
  - `YoungCalves`: Whole number (integer) – average young calf herd size of `CADDID` during the `Year`,
  - `BeefCattle`: Whole number (integer) – average beef cattle herd size of `CADDID` during the `Year`,
  - `Reported (1) vs. Estimated (2) vs. No Data (3)`: Whole number (integer) – indication of how herd size information is determined:
    - One (1): herd size information is based on an annual report submitted to Regional Waterboards,
    - Two (2): no annual report found for `CADDID` during the `Year`. Herd size information is estimated following a very comprehensive process (for additional information, please refer to the CADD Technical Document that will be posted on the *Short-Lived Climate Pollutants Meetings & Workshops* page).
    - Three (3): no herd size information (historical and current) is available for `CADDID` in any of the data sources below. In such cases, the herd size of all cattle categories was assumed to be zero (0).
  - `LabeledAsDairy`: Whole number (integer) – indication of whether or not `CADDID` housed milk cows during the `Year`:
    - Zero (0): Labeled as not a dairy since `MilkCows` is equal to zero (0) for `CADDID` during the `Year`,
    - One (1): Labeled as a dairy since `MilkCows` is greater than zero (0) for `CADDID` during the `Year`.
- **Data Sources**: The sources of data used to develop this table are listed below:
  - **Annual reports from Regional Waterboards**: In California, all operators of cattle facilities, with some exceptions, are required to submit annual reports to their regulating Regional Waterboards. These reports include average herd size

information for multiple animal types, such as milk cows, dry cows, heifers, calves, and beef cattle.

- o **Inspection reports[5]**: are conducted by the Regional Waterboards' staff and uploaded to the CIWQS website. These reports typically provide facility's herd size data at the time of inspection, serving as an alternative source of information in case of a missing annual report.
- o **Air permits**[6]: issued by the San Joaquin Valley Air Pollution Control District (SJVAPCD); these permits specify the maximum permitted herd size for each SJVAPCD-permitted facility, serving as an alternative source of information in case of a missing annual report.
- o **Regulation information from CIWQS**[7]: provides information[8] about the operational status of a facility in a given year in case of a missing annual report.
- o **Google Earth images**: used to evaluate the operational status of a facility in a given year in case of a missing annual report.
- o **Waste management plans, nutrient management plans, notices of intent**: These documents are generally submitted to the Regional Waterboards once by the facility owners/operators to enroll in the Dairy and Bovine General Orders. Typically containing the facility's herd size data at the time of document submission, they serve as alternative sources of information in case of a missing annual report.

3.3 Table 3: Anaerobic Digesters Table

- **Description**: Contains facilities with anaerobic digesters, the year when they became operational, and the year they shut down (if applicable).
- **Columns**:
  - o `CADDID`: Whole number (integer) – a reference to the facility; linked to the `CADDID` in the Facility General Information table,
  - o `OperationalYear`: Whole number (integer) – the calendar year in which the anaerobic digester became operational. For DDRDP projects, `NaN` means "verification in process"; for non-DDRDP projects, `NaN` means the digester is under construction,
  - o `ShutdownYear`: Whole number (integer) – the calendar year in which the anaerobic digester shut down, if applicable. `NaN` means the digester has not shut down (either is not constructed yet or is operational),
  - o `DataSource`: Text – The data source for each record: DDRDP or AgSTAR (described below).
- **Data Sources**: The sources of data used to develop this table are listed below:

---

[5] *California Integrated Water Quality System* (last accessed February 16, 2024).

[6] *Permits Public Information Portal (valleyair.org)* (last accessed July 3, 2019).

[7] *Surface Water - Water Quality Regulated Facility Information - Confined Animal Facilities Regulated by the Water Boards - California Open Data* (last accessed November 21, 2023).

[8] Includes information on the "Order #" a facility is subject to, the date of "recission of coverage letter" due to going out of business (if applicable), termination date of the order for that facility, etc.

- o **Project-level data from DDRDP[9]**: published quarterly by the California Department of Food and Agriculture, lists information about each DDRDP-funded project, including the award year, facility name, location information, start date of digester construction, and completion date. `OperationalYear` column of CADD for DDRDP-funded projects is based on the completion date in these documents.
- o **AgSTAR Livestock Anaerobic Digester database from US EPA[10]**: contains information about the status of dairy digester projects within the United States, including project names, cities, project statuses (operational, under construction, or shutdown), the year of starting operation (for operational and shutdown projects), and the shutdown year, if applicable.

NOTE: If the operational years reported in these two data sources were inconsistent, information from DDRDP was used.

---

[9] *CDFA - OEFI - Dairy Digester (ca.gov)* (published on March 24, 2023 and accessed on April 13, 2023).

[10] *Livestock Anaerobic Digester Database | US EPA* (last accessed on January 10, 2024; based on data available through May 2022).