Automating Data Unification, Governance, and Analytics for Ruminant Research

A Cloud-Native, AI-Assisted Framework

Puchun Niu, Enhong Liu, Hanlu Zhang, Ermias Kebreab, Jennifer Ellis, Mutian Niu, Stephanie Terry, Fran Cowley, Lyle Rode, Miel Hostens

Methane matters: we must measure to mitigate

Respiration chamber—measurement gold standard

GreenFeed System—becoming the ‘de facto’ system

GreenFeed Emissions Measurement System

Project:
Global collaboration to track ruminant gas emissions and build a FAIR data warehouse for climate-smart livestock research

Mission:
Develop standardized SOPs for GreenFeed data use & interpretation to ensure comparable, defensible emissions measurements across partners.

Primary institution

Additional contributors

GEMS Project Goals



1 Analyze emissions data across methods

2 **Build a Global GreenFeed Data Resource**

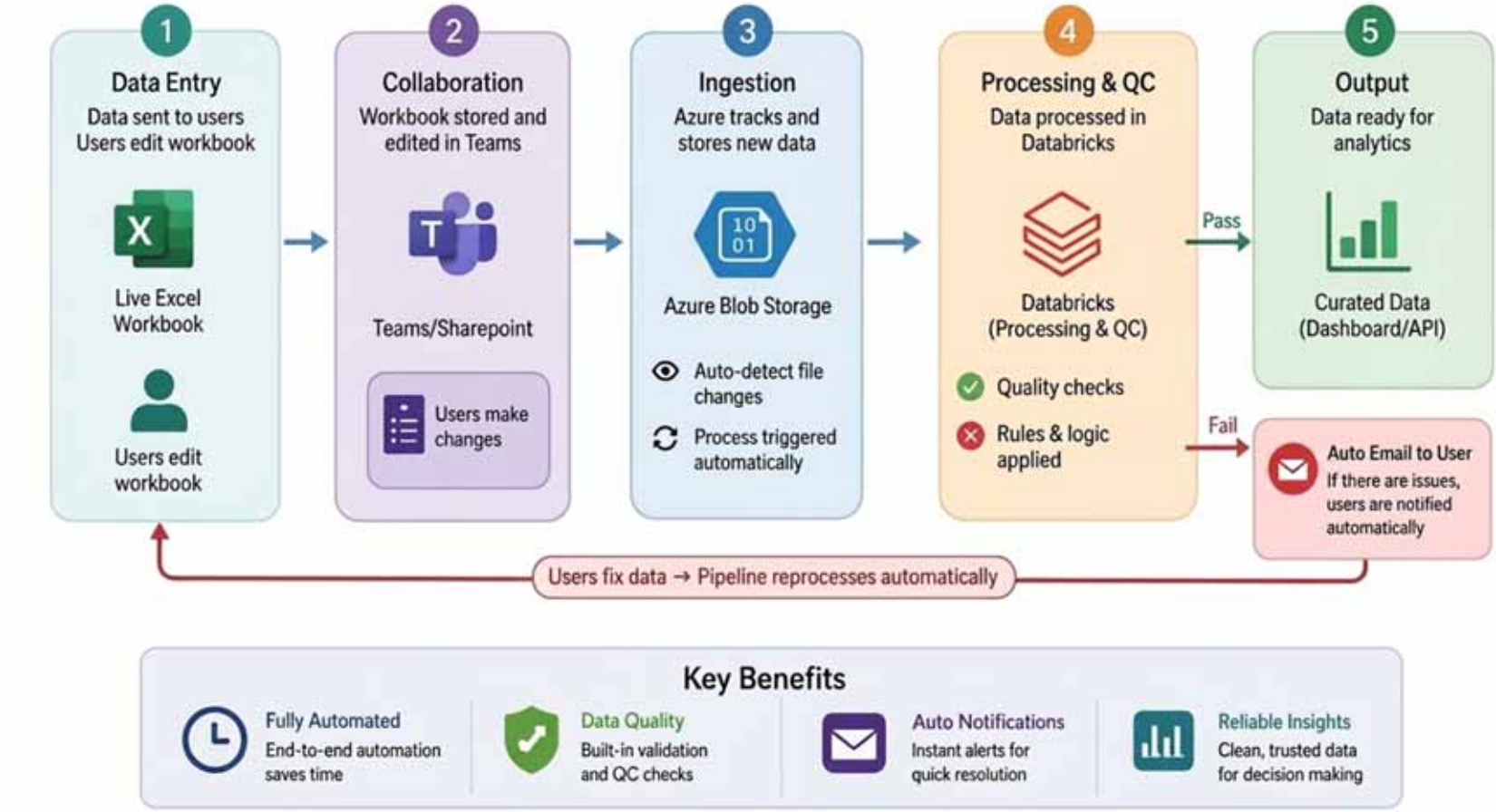
3 **Develop data user agreement and populate GF database**

4 Understand What Affects Emission Estimates

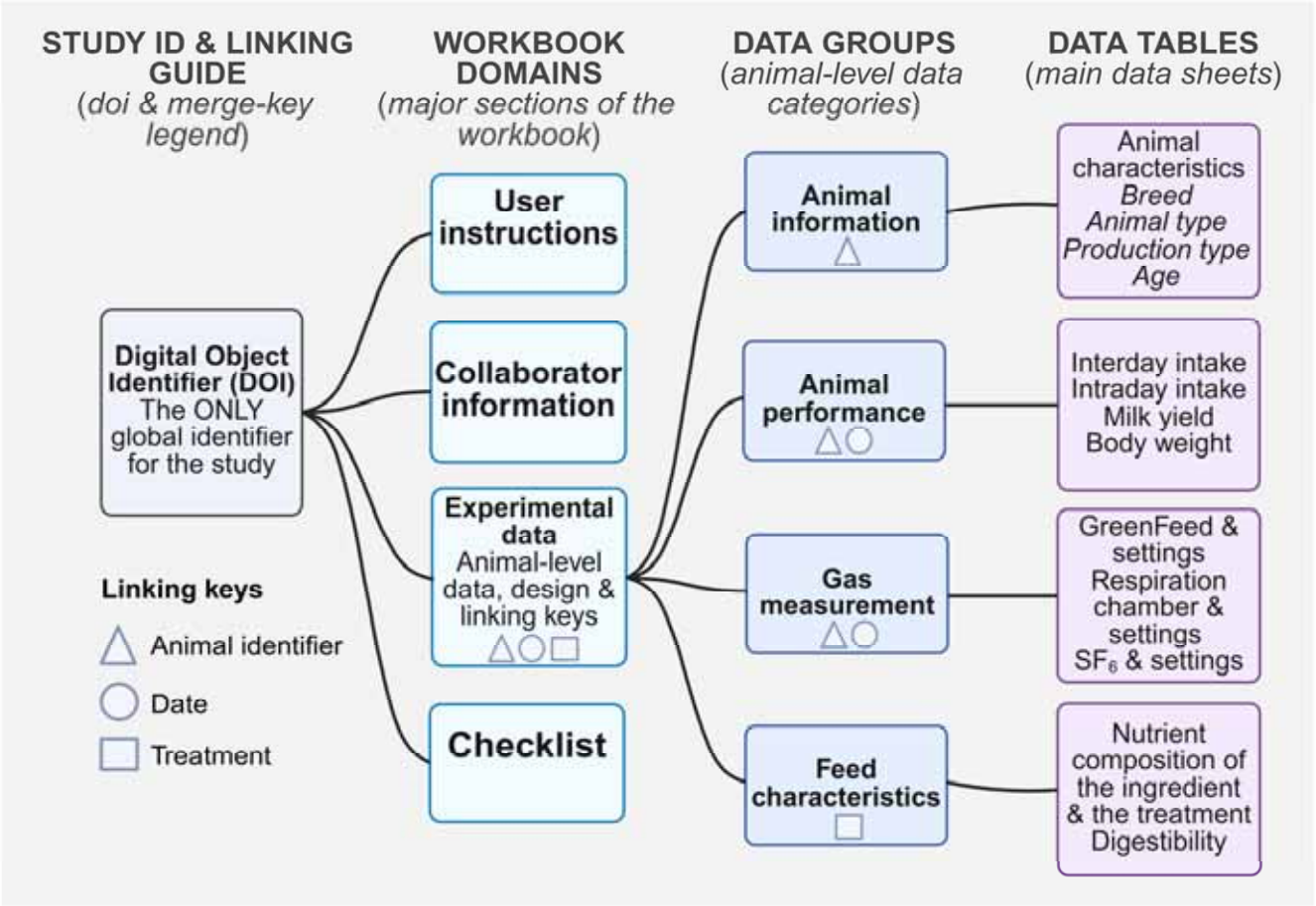
5 Validate GreenFeed with Respiration Chambers

6 Deliver Practical Tools and Guidelines

GEMS centralized end-to-end data platform



GEMS data entry workbook



GEMS data entry workbook and DUA

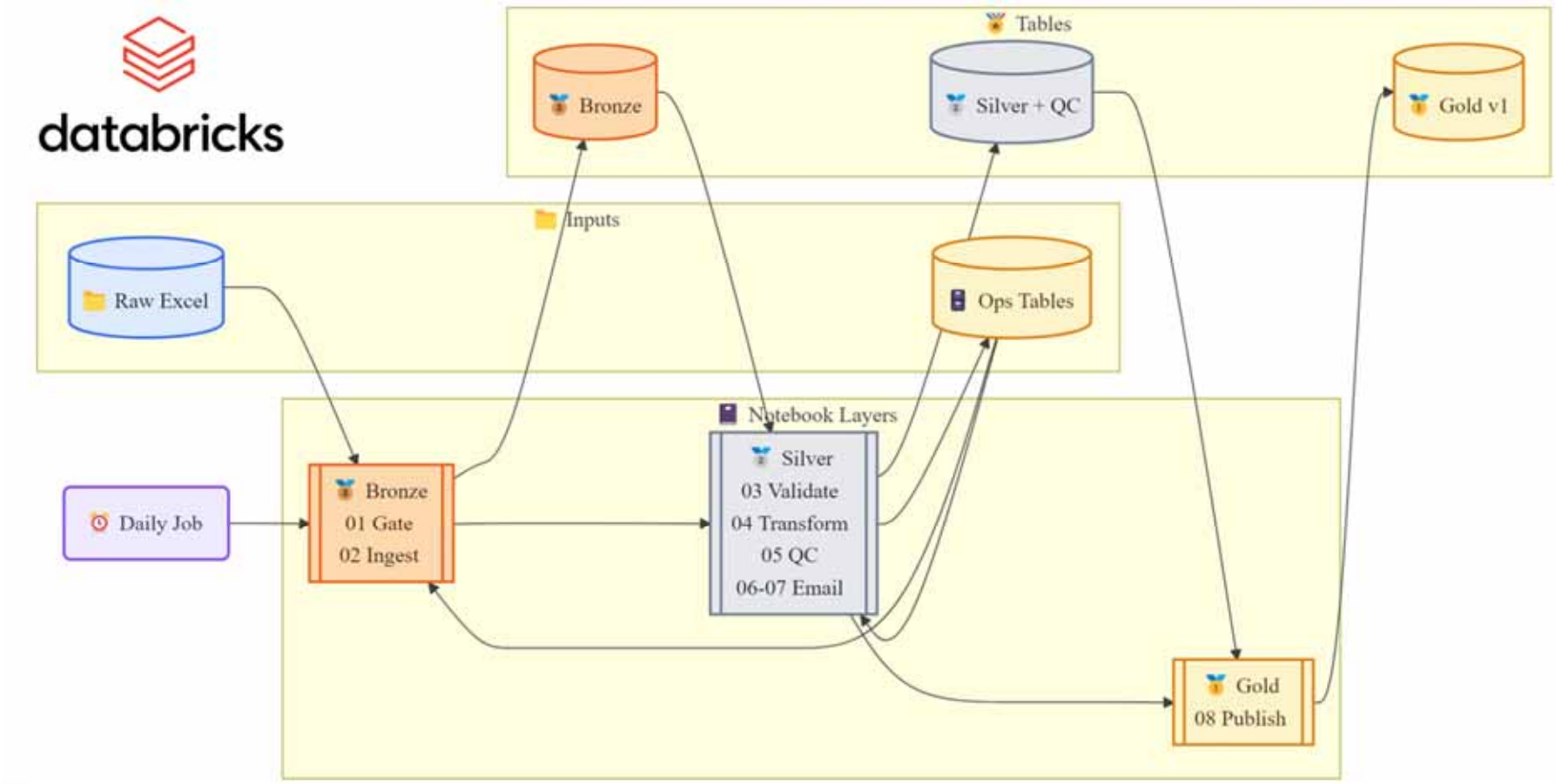
	A	B	C	D	E	F	G	H	I	J	K	L
1	Expression	AnimalIdentifier	RadioFrequencyIdentification	Breed	ProductionType	AnimalType	BirthDate	Age	AnimalStatus	Parity	DaysInMilk	Cannulation
2	This sheet is MANDATORY. List all animals used in the experiment, with one animal per row. Column headers shown in dark blue link to an official reference (e.g., ICAR breed list)	Identification number assigned to the animal by the researcher or farm. This field is mandatory and must be consistently used across all relevant sheets to enable data linkage. If Radio Frequency Identification (RFID) is used as the animal's identifier, enter the RFID here to ensure each animal has a unique and consistent identifier.	Enter the Radio Frequency Identification (RFID) tag used by the GreenFeed system. If a tag is replaced, record the new RFID on a new line while retaining the same AnimalIdentifier to maintain data continuity.	Holstein-Friesian, Jersey, Brown Swiss, Angus, Hereford, Simmental, etc. If the animal is crossbred, list the contributing breeds, separated by "×" (e.g., Holstein×Jersey)	Dairy, Beef, Dual-Purpose	Cow, Heifer, Bull, Steer, Calf, bullock	Enter the animal's date of birth in the format YYYY-MM-DD	Age of the animal (in months) at the start of the experiment	The status of the animal at the start of the experiment, 'Lactating' or 'Dry' (if applicable)	Enter the parity number of the cow at the start of the experiment (1 = first lactation, 2 = second lactation, etc.)	Days in milk at the beginning of the experiment	Was the cow cannulated: 'Yes' or 'No'
3	Prioritization	Must have	Must have	Must have	Must have	Must have	Nice to have	Nice to have	Nice to have	Nice to have	Nice to have	Nice to have
4	Unit	Text	Text	Text	Text	Text	Date	Month	Text	Number	Number	Text
5		2919	840003145678912	Holstein							100	No
6		2923	840009876543210	Holstein							101	No
7		154	840000123456789	Angus							103	No

Table 1 – Required Data Variables and Contributor Commitments *[Please fill in the correspon information and check the boxes for datasets that will be provided from this study. Please create a table for each additional study and change the 'study number' respectively]*

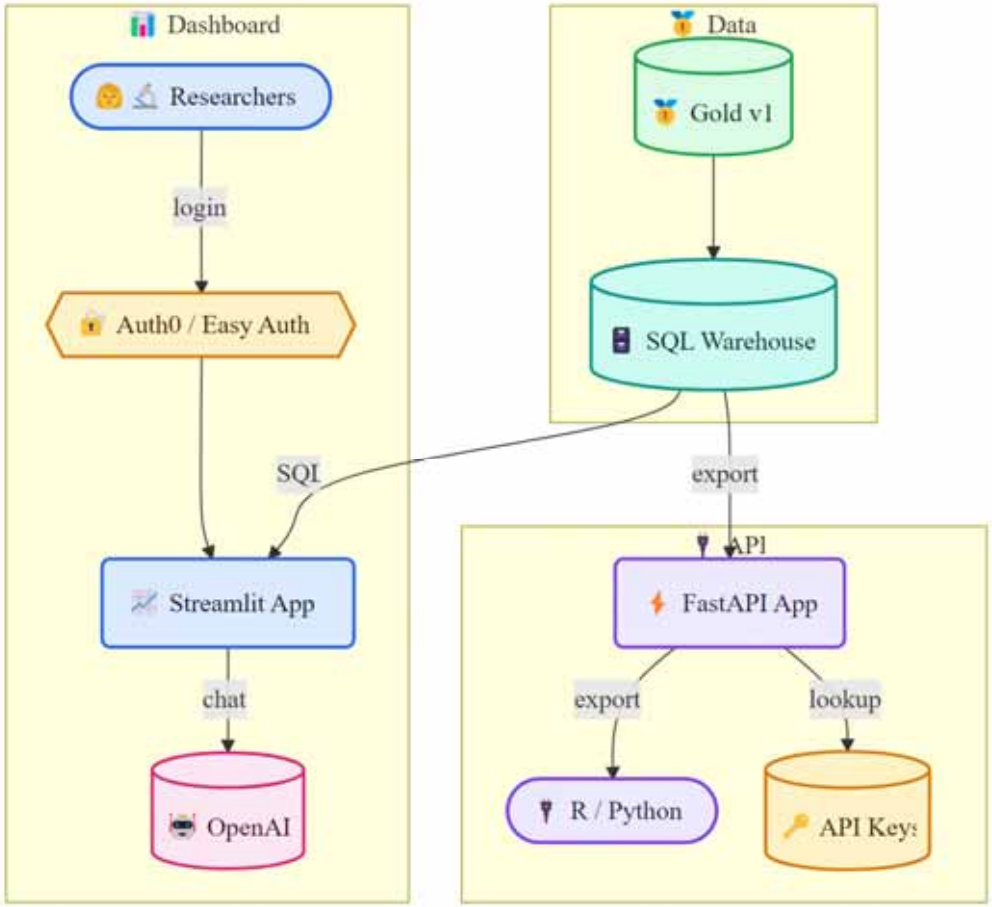
Study number[1]	1
Contact person (Data Manager)	
Contact email (Data Manager)	
Number of animals	
Study length (days)	
GreenFeed pellet components (for non-SF_6 studies)	☑
GreenFeed data file reference (for non-SF_6 studies)	☑
Animal characteristics	☑
Experimental design	☑
Average daily dry matter intake	☑
Diet nutrient composition	☑
Milk yield (if applicable)	☑
Milk composition (if applicable)	☑
Dry matter intake intraday	☐
Feed components	☐
Body weight	☐
Diet digestibility	☐
Nutrient digestibility	☐
GreenFeed calibration	☐
Respiration chamber settings (required for chamber vs GreenFeed studies)	☐
Respiration chamber measurements (required for chamber vs GreenFeed studies)	☐
SF_6 settings (required for SF_6 studies)	☐
SF_6 measurements (required for SF_6 studies)	☐

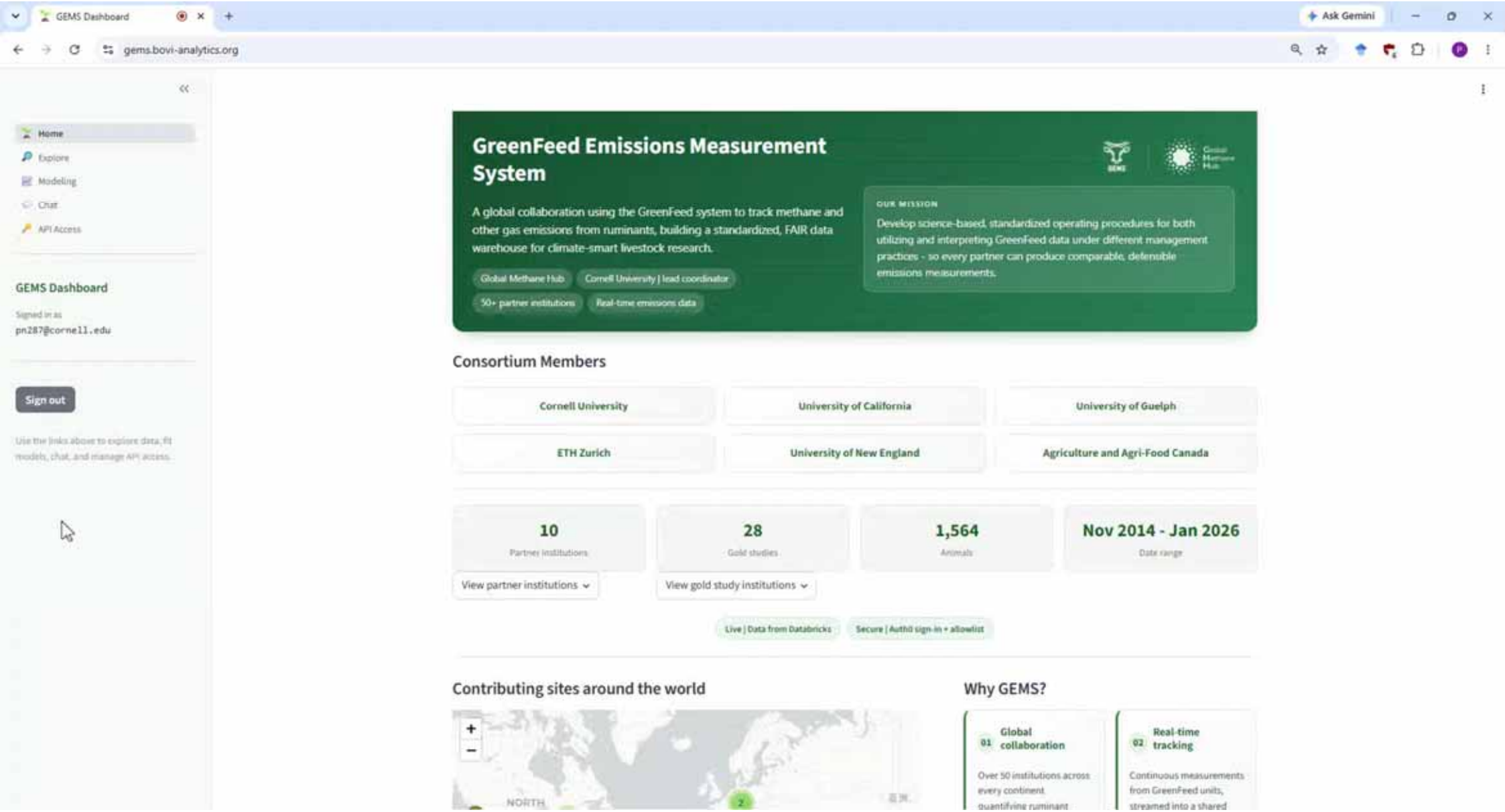
GeneralInstruction | Checklist | Contributor | AnimalCharacteristics | ExperimentalDesign | FeedComponents | DietNutrientComposition | IntakePerDay | IntakeIntraDay | Milk | BodyWeight | Digestibility | GreenFeedSettings

GEMS Medallion Pipeline



GEMS Data Sharing and Access




GEMS Dashboard
Ask Gemini
gems.bovi-analytics.org
Home
Explore
Modeling
Chat
API Access
GEMS Dashboard
Signed in as
pn287@cornell.edu
Sign out
Use the links above to explore data, fit models, chat, and manage API access.
GreenFeed Emissions Measurement System
A global collaboration using the GreenFeed system to track methane and other gas emissions from ruminants, building a standardized, FAIR data warehouse for climate-smart livestock research.
Global Methane Hub
Cornell University | lead coordinator
50+ partner institutions
Real-time emissions data
OUR MISSION
Develop science-based, standardized operating procedures for both utilizing and interpreting GreenFeed data under different management practices - so every partner can produce comparable, defensible emissions measurements.
Consortium Members
Cornell University
University of California
University of Guelph
ETH Zurich
University of New England
Agriculture and Agri-Food Canada
10
Partner institutions
28
Gold studies
1,564
Animals
Nov 2014 - Jan 2026
Date range
View partner institutions
View gold study institutions
Live | Data from Databricks
Secure | Auth0 sign-in + allowlist
Contributing sites around the world
NORTH
Why GEMS?
01 Global collaboration
Over 50 institutions across every continent quantifying ruminant
02 Real-time tracking
Continuous measurements from GreenFeed units, streamed into a shared

Thank You

We welcome collaborations, data partnerships, and questions.

GEMS Project

Project Inquiries

gems@cornell.edu

Miel Hostens

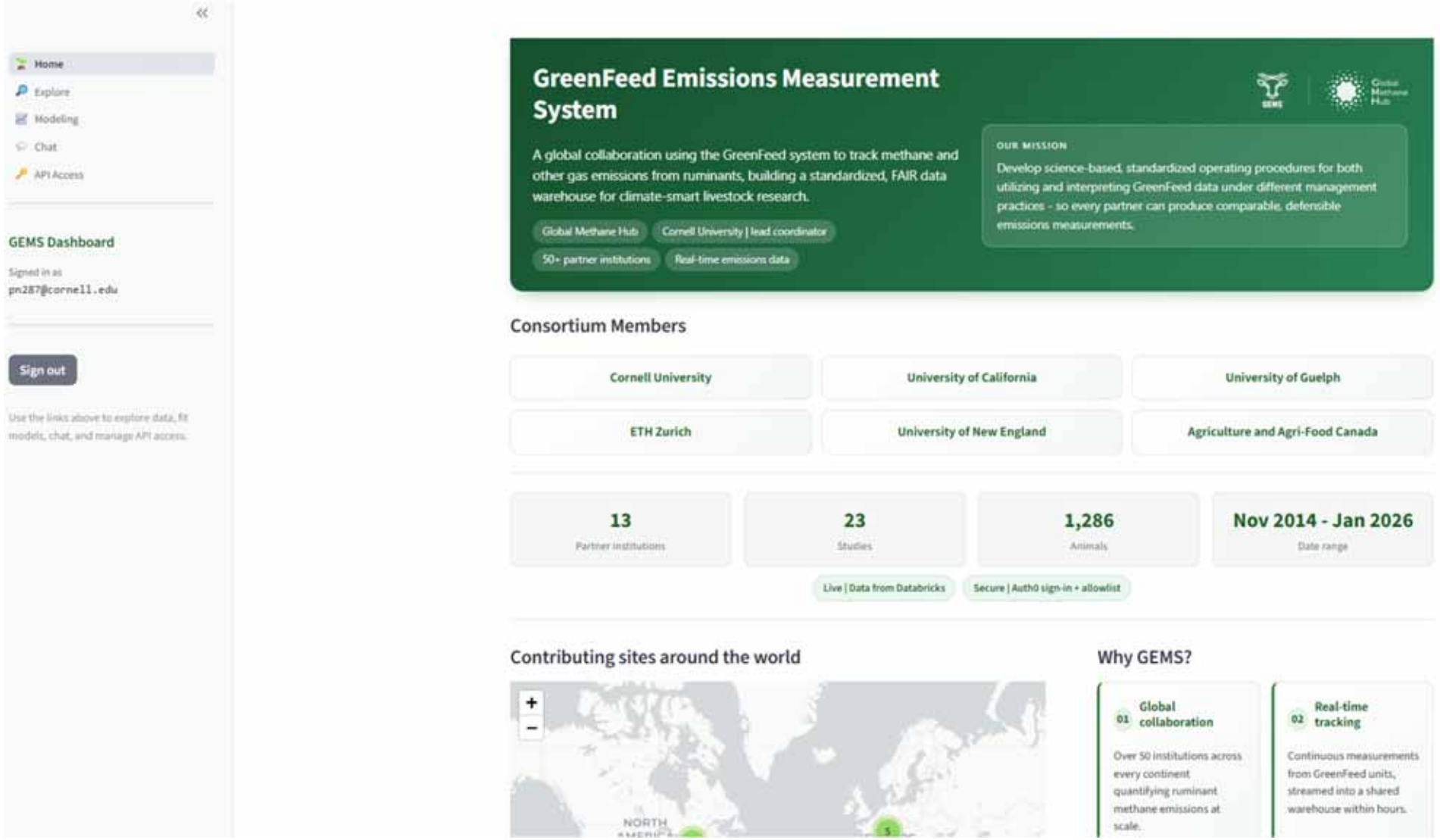
Principal Investigator

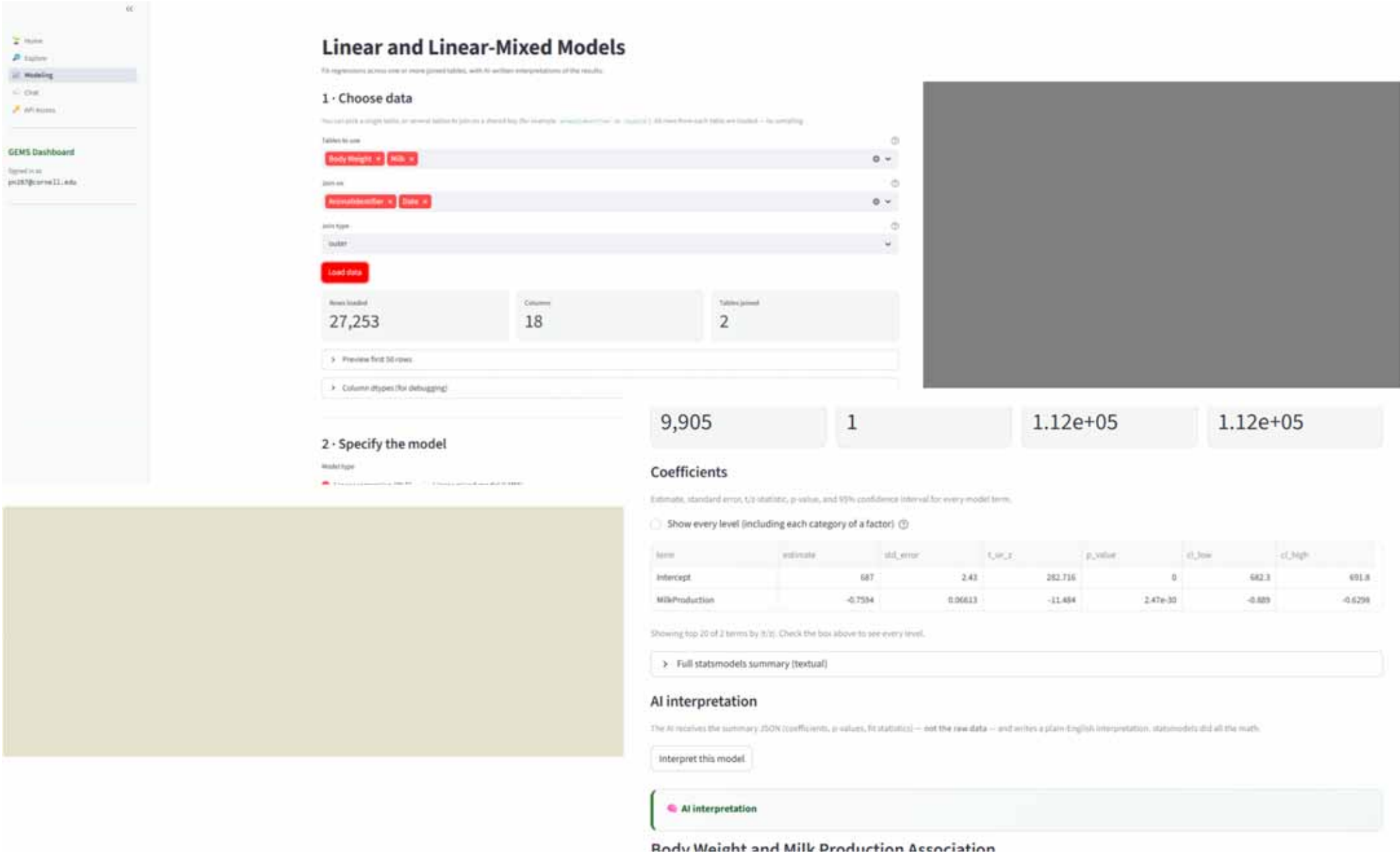
miel.hostens@cornell.edu

Puchun Niu

Data & Analytics

pn287@cornell.edu


Home
Explore
Modeling
Chat
API Access
GEMS Dashboard
Signed in as
pn287@cornell.edu
Sign out
Use the links above to explore data, fit models, chat, and manage API access.
GreenFeed Emissions Measurement System
A global collaboration using the GreenFeed system to track methane and other gas emissions from ruminants, building a standardized, FAIR data warehouse for climate-smart livestock research.
Global Methane Hub
Cornell University | lead coordinator
50+ partner institutions
Real-time emissions data
OUR MISSION
Develop science-based, standardized operating procedures for both utilizing and interpreting GreenFeed data under different management practices - so every partner can produce comparable, defensible emissions measurements.
Consortium Members
Cornell University
University of California
University of Guelph
ETH Zurich
University of New England
Agriculture and Agri-Food Canada
13
Partner institutions
23
Studies
1,286
Animals
Nov 2014 - Jan 2026
Date range
Live | Data from Databricks
Secure | Auth0 sign-in + allowlist
Contributing sites around the world
NORTH
Why GEMS?
01
Global collaboration
Over 50 institutions across every continent quantifying ruminant methane emissions at scale.
02
Real-time tracking
Continuous measurements from GreenFeed units, streamed into a shared warehouse within hours.


Home
Explore
Modeling
Chat
API Access
GEMS Dashboard
pn287@cornell.edu
Linear and Linear-Mixed Models
1 · Choose data
Body Weight
Milk
AnimalIdentifier
Date
Load data
27,253
18
2
Preview first 50 rows
Column dtypes (for debugging)
2 · Specify the model
9,905
1
1.12e+05
1.12e+05
Coefficients
Show every level (including each category of a factor)
Intercept 687 2.43 282.716 0 682.3 691.8
MilkProduction -0.7594 0.06613 -11.484 2.47e-30 -0.889 -0.6298
Full statsmodels summary (textual)
AI interpretation
Interpret this model
AI interpretation
Body Weight and Milk Production Association

Chat with your data

what is the distribution of methane emission in grams per day

Summary: Distribution of Methane Emissions in Grams per Day

1. GreenFeed Raw Visitation Data (460,083 records)

- **Range:** -999 to 2,330 g/day
- **Mean:** 100.3 g/day
- **Median:** 171.1 g/day
- **Quartiles:**
  - Q1: 105.2 g/day
  - Q3: 228.4 g/day
- **Distribution:** Highly right-skewed with most measurements concentrated between 100-250 g/day, with a long tail extending to higher values

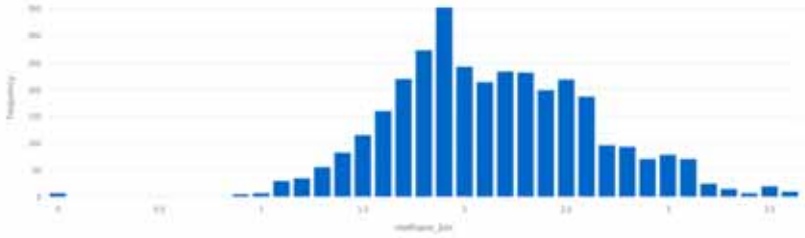
2. Respiration Chamber Measurements (3,342 records)

- **Range:** 0 to 11 g/day
- **Mean:** 2.24 g/day
- **Median:** 2.05 g/day
- **Quartiles:**
  - Q1: 1.65 g/day
  - Q3: 2.49 g/day

Key Observations:

- The **GreenFeed data** shows much higher methane values (100-250 g/day typical range) compared to respiration chamber data
- The **respiration chamber data** is more tightly distributed (1.8-2.5 g/day typical range)
- Both datasets show right-skewed distributions with outliers at the high end
- The GreenFeed data includes some negative values (-999), which may represent missing or error codes

Distribution of Methane Emissions (Respiration Chamber) - g/day



Show query

API Access

Generate API keys and use version-aware scripts to query or refresh GEMS data.

Create an API key for Python, R, curl, or other tools. Keys are shown only once. Only API-tier users can access this page.

Key name

e.g., My laptop, R/Quarto script, shared analysis workflow

Generate API key

Your API keys

Name	Prefix	Created	Last used
API key	gems_live_BoolyTsE...	2026-05-04T15:41:51.324245+00:00	2026-05-12T19:58:54.698830+00:00
first test	gems_live_5yLC7Q...	2026-04-28T13:04:22.462696+00:00	2026-04-28T13:38:53.383414+00:00

Revoke a key

Active key

API key (gems_live_BoolyTsE...)

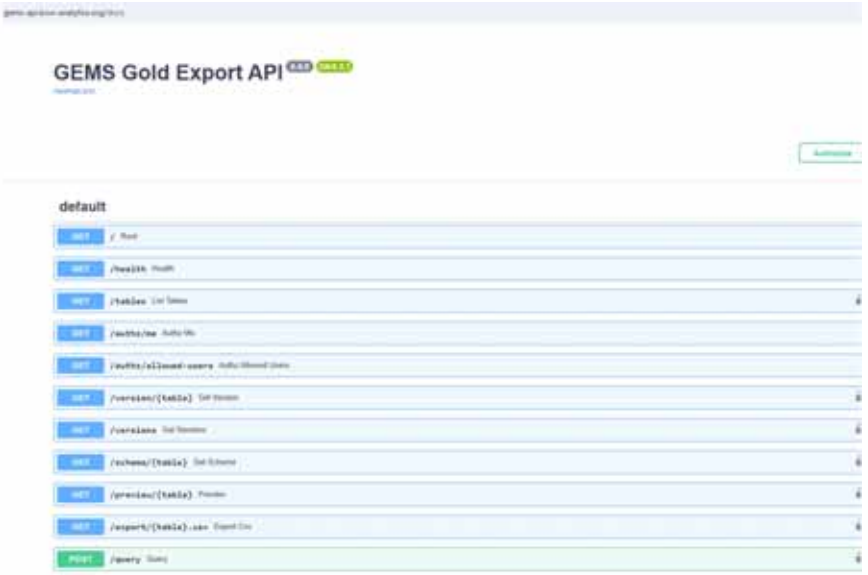
Revoke selected key

API connection details

https://gems-api.bovi-analytics.org

Interactive API documentation:

Open Swagger UI



GEMS Gold Export API

default

GET / Root

GET /health Health

GET /tables List Tables

GET /auth/me Auth Me

GET /auth/allowed-users Auth Allowed Users

GET /version/{table} Get Version

GET /versions Get Versions

GET /schema/{table} Get Schema

GET /preview/{table} Preview

GET /export/{table}.csv Export Csv

POST /query Query

Open Swagger UI

Every data request must include an `X-API-Key` header. The examples below read the key from `GEMS_API_KEY` instead of pasting it directly into commands.

> Available table names

v API documentation and examples

Quick start | Python | R | Endpoints | Security

Recommended workflow

1. Generate an API key above.
2. Create a `.env` file in the same folder as your Python or R script:

```
GEMS_API_KEY="gems_live_your_real_key_here"
```

3. Use the Python or R **All-table version-aware refresh** script in the next tabs. It checks table versions, prints which tables are already current or changed, downloads only changed tables, and overwrites local CSV files.

Run the refresh script whenever you want to check for updates. If a table has a newer version, the script tells you and downloads the latest snapshot. If nothing changed, it skips the download.

Opening the base API URL directly may show a short JSON service message. For interactive browser testing, use Swagger UI:

```
https://gems-api.bovi-analytics.org/docs
```