

Agriculture Data Space for Sovereign Data Sharing and Semantic Integration

Rosa Gil & Roberto García

Universitat de Lleida

<https://rhizomik.net/~roberto>

Introduction

- Digitalization and **data** critical for the transformation of the agriculture sector
- Sharing **beyond** open data:
 - Complex due to personal, sensible, strategic... data
- **Data Spaces**: an open and decentralized infrastructure for **sovereign** data sharing
 - Participants **control** how data is used
 - No central data repository, **direct interchanges** among participants

What do Data Spaces provide?

- Overcome **barriers** to **data sharing** across organizations by combining necessary tools and infrastructures and addressing issues of **trust**
- Data sovereignty:
 - Ensure stakeholders do not get exploited in the competition for their data
 - Protect how one's data is used and request for fair conditions
 - Sovereignty for states, organizations and also individuals

How to motivate data sharing ?

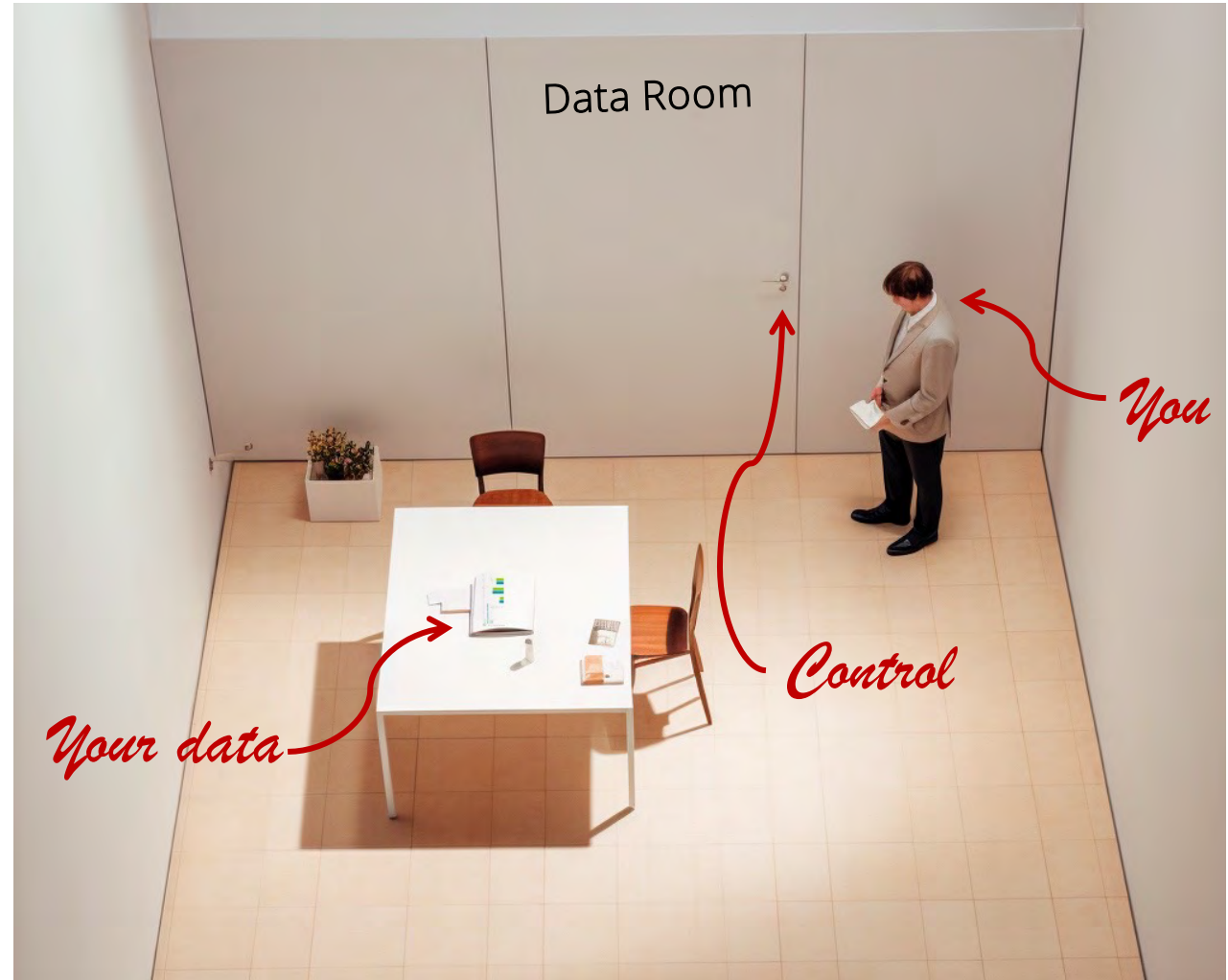
- Data-monetization (tokenization) as a data sharing incentive
- Requires an environment I can trust to share my data in, without losing its control
 - *"tragedy" of the digital*, so easy to copy and disseminate...
 - Example: digital music → Digital Rights Management (DRM),... streaming
- For data... Data Self-Sovereignty by Design?
 - Data Space P2P Connectors or Data API Intermediaries
→ **difficult policy enforcement** beyond trusted environment (like with DRM...)

Data Self-Sovereignty by Design

- Your data:
Never “leaves the room”
- You:
In control on who “enters” and what they “take out” with them

- ✗ Full or partial copies
- ✓ Aggregated data
- ✓ AI trained models
- ✗ Personal data

...



dar01995_a_bird-view_of_the_inside_of_a_whole_room_without_wind_1b2e5c91-5513-471d-9673-6c6493a76c11

Compute-to-Data

- Facilitates **privacy** and **Data Sovereignty** by design
 - Combinable with **Federated Learning**, **Compute-to-Edge**,...
- Algorithms are executed where the data is stored
 - Data originators keep **control** of data, **not copied** anywhere else



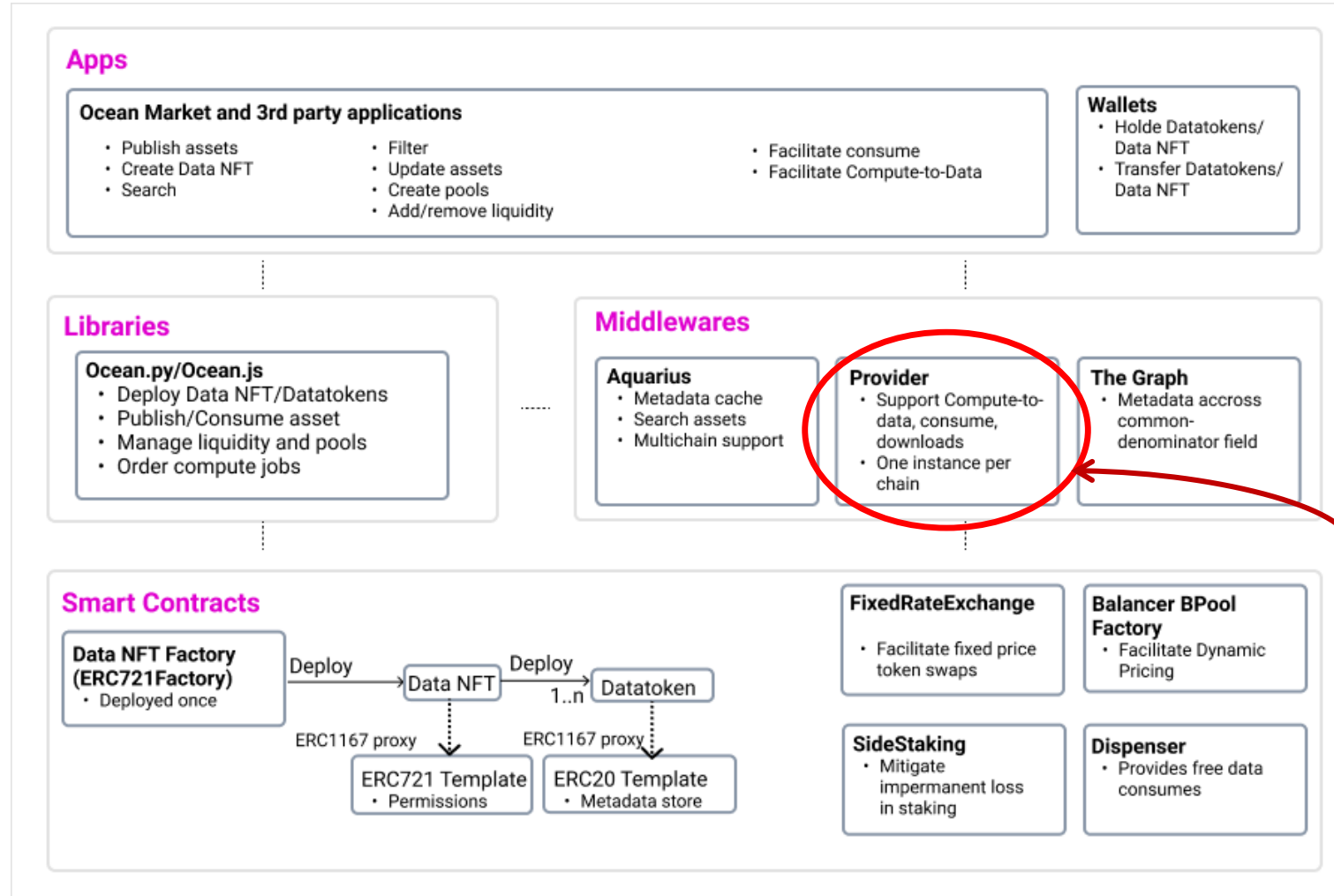
Pontus-X: Gaia-X Web3 Ecosystem

PONTUS-X

powered by
gaia-x



Pontus-X: built on Ocean Protocol's technology stack



Proof of Concept

<https://dataspace.angliru.udl.cat>



UdL Research Data Space

An open research platform following the FAIR guiding principles of findability, accessibility, interoperability and reusability, while guaranteeing data sovereignty.

AI-based Animal Well-being Assessment without Images Leakage

 COMPUTE
  DATASET
  GXAT

Pig pen images sequence for animal well-being assessment
 0x38f8...840d
 A sequence of images from video surveillance of one of the pens in the Centre of Swine Studies of Catalonia (CEP), an experimental pig farm managed by a consortium made u...
Free
 1 sale  GEN-X Testnet

 COMPUTE
  ALGORITHM
  GXAT

CIDAI pig farm Mask R-CNN segmentation & Tracking (CUDA) (UDL)
 0x3BB6...8E40
 Perform instance segmentation with a Mask R-CNN and tracking of pig images seen from above. The input data must be a zip with images.
Free
 5 sales  GEN-X Testnet

Precision Pig Feeding Semantic Data Integration and Sovereign Data Pooling

 COMPUTE
  ALGORITHM
  CEP-RML

CEP's CSV Data Mapper and Semantic Data Pooler
 0x6207...6822
 This algorithm supports the "Pay-as-you-go" approach when sharing data through a data space. ...
Free
 2 sales  GEN-X Testnet

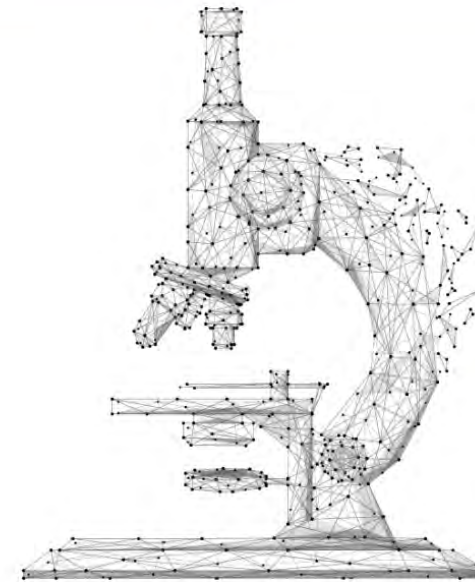
 COMPUTE
  DATASET
  CEP-FEED

CEP - Automatic Pig Feeding - 2021 S1 - 982091062894506
 0x38f8...840d
 Automatic pig feeding data collected at the (Centre of Swine Studies of Catalonia), a consortium made ...
Free
 2 sales  GEN-X Testnet

 DOWNLOAD
  ALGORITHM
  UDL-EDA

Exploratory Data Analysis
 0x6207...6822
 Generate an Exploratory Data Analysis (EDA) report for the input tabular data. The input data is loaded ...
Free
 10 sales  GEN-X Testnet

[ALL DATASETS AND ALGORITHMS →](#)



Goal

The Universitat de Lleida Gaia-X Web3 Ecosystem Portal aims to accelerate the progress of research and science, increase the quality of data, and foster collaboration between research and educational institutions. On the research portal, data scientists and researchers can find, consume and share research data sets and algorithms. Depending on the publisher's decision, the data assets can be consumed either by granting compute or download access to the data assets.

Data Audit Trail for Academia

By tokenizing data assets and using smart contracts to manage access control, each transaction, which includes publishing, sale, exchange, and consumption of data services, is logged automatically and immutable on the distributed ledger. This creates a data audit trail for academia which allows research institutions to monitor the use of their assets and gain attribution for their work.

Data Sovereignty by Design

One core concept of the UdL Research Data Space is the **Compute-to-Data (CtD)** approach. Compute-to-Data is the functionality that solves the current trade-off between the benefits of using private data and the risks of exposing it. It allows data consumers to run compute jobs on private data while the data stays on-premise with the data provider, who retains control.



AI4Pork Use Case



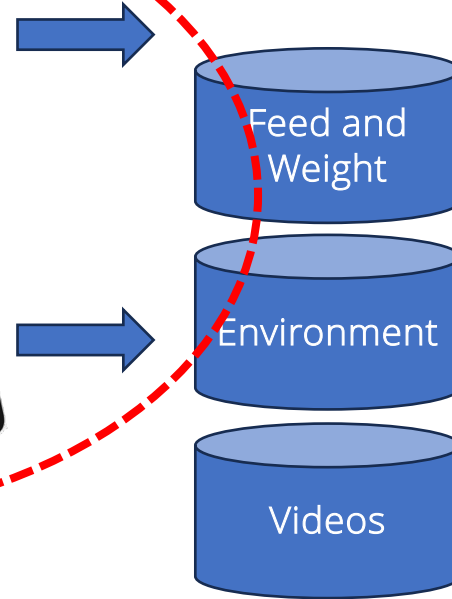
Precision Feeding



Temp., Hum., CO2,
Ammonia,...

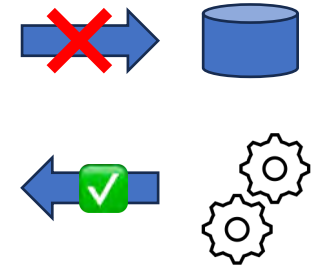


Cameras



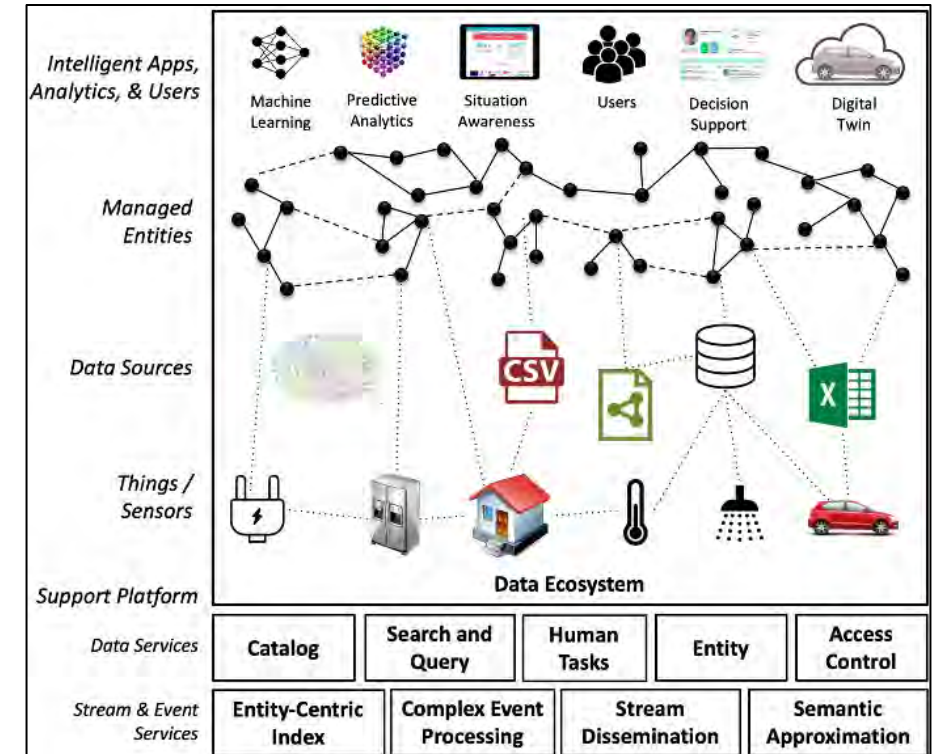
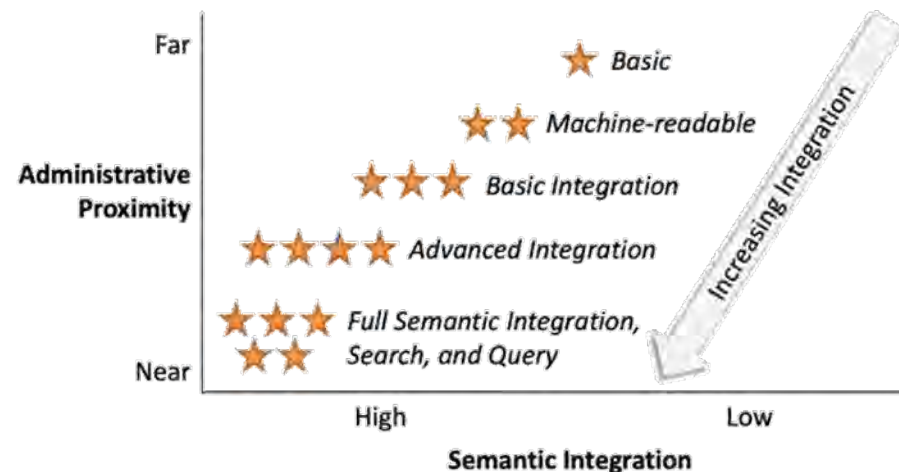
Centre of Swine
Studies of
Catalonia

Compute-to-Data Provider
at Universitat de Lleida



Semantic Integration

- <http://dataspaces.info>
- Pay-as-you-go Model
 - Reduce initial cost to join
 - Integration level with dataspace services improved incrementally, as-needed



PoC: AI4Pork Data Space

Pen Number	Animal ID	Date	Time	Duration (s)	Feed (g)	Animal Weight (g)
4	982091062894196	2021-03-16	01:08	42	0	16500
4	982091062894196	2021-03-16	01:09	45	1	16300
4	982091062894196	2021-03-16	13:05	70	0	15300
4	982091062894196	2021-03-17	10:44	5	0	14400
4	982091062894196	2021-03-17	10:45	9	6	14500
4	982091062894196	2021-03-17	10:45	50	14	14400



Universitat de Lleida

[CATALOGUE](#)
[PUBLISH](#)
[VERIFY](#)
[LOG](#)
[ECOSYSTEM](#)

[GEN-X Testnet](#)
[0x6207...6822](#)

CEP - Automatic Pig Feeding - 2021 S1 - 982091062894196

GEN-X Testnet



Owned by 0x6207...6822
Accessed with CEP-2021-S1-FEED_982091062894196

DATASET

Published 4 days ago

Automatic pig feeding data collected at the Centre for Pig Studies of Catalonia (Centre d'Estudis Porcins), a consortium made up of the Diputació de Lleida, the Regional Council of La Noguera, the Torrelameu Town Hall and the Universitat de Lleida.

The feeding data includes:

- Pen Number:** the pen the pig is located in.
- Animal ID:** the pig identifier.
- Date:** the date the feeding data is about.
- Time:** the time pig feeding happened for the given date.
- Duration (s):** the duration of the feeding event.
- Feed (g):** the amount of food provided, measured in grams.
- Animal Weight (g):** the pig weight, measured by the feeding machine during each feeding event.

udl

pig

agriculture

livestock

feed

automatic-feeding

DATA AUTHOR

Centre for Pig Studies of Catalonia

OWNER

0x6207...6822

DID

did:op:0d9493e7c7be9afdd6119a0099b54347ab053201d04a49ac1090deee8e73e041

Binary

56.98 kB

url

Free

Select an algorithm to start a compute job

Exploratory Data Analysis

Free

RML Mapper for Daily Pig Weight CSV Data

Free

You will pay

BUY COMPUTE JOB

You bought this dataset already allowing you to use it without paying again. The C2D resources required to start the job are available, no payment is required for them. Please note that network gas fees still apply, even when using free assets.

Your Compute Jobs

SHOW

2 sales

CEP - Automatic Pig Feeding - 2021 S1 - 982091062894196

GEN-X Testnet

Owned by Ox6207...6822
Accessed with CEP-2021-S1-FEED_982091062894196

DATASET Published 4 days ago

Automatic pig feeding data collected at the Centre for Pig Studies of Catalonia (Centre d'Estudis Porcins), a consortium made up of the Diputació de Lleida, the Regional Council of La Noguera, the Torrelameu Town Hall and the Universitat de Lleida.

The feeding data includes:

- Pen Number:** the pen the pig is located in.
- Animal ID:** the pig identifier.
- Date:** the date the feeding data is about.
- Time:** the time pig feeding happened for the given date.
- Duration (s):** the duration of the feeding event.
- Feed (g):** the amount of food provided, measured in grams.
- Animal Weight (g):** the pig weight, measured by the feeding machine during each feeding event.

udl pig agriculture livestock feed automatic-feeding

DATA AUTHOR
Centre for Pig Studies of Catalonia

OWNER
Ox6207...6822

DID
did:op:0d9493e7c7be9afdd6119a0099b54347ab053201d04a49ac1090deee8e73e041

Binary 96.98 kB url

Free

Select an algorithm to start a compute job

Search by title, datatoken, or DID...

- Exploratory Data Analysis** d1d:op:e73ac9c332adfea204acefb4b308e... **Free**
- RML Mapper for Daily Pig Weight CSV Data** d1d:op:f421d1ff824262a7a61e42370f65cc... **Free**

You will pay

BUY COMPUTE JOB

You bought this dataset already allowing you to use it without paying again. The C2D resources required to start the job are available, no payment is required for them. Please note that network gas fees still apply, even when using free assets.

Your Compute Jobs **SHOW**

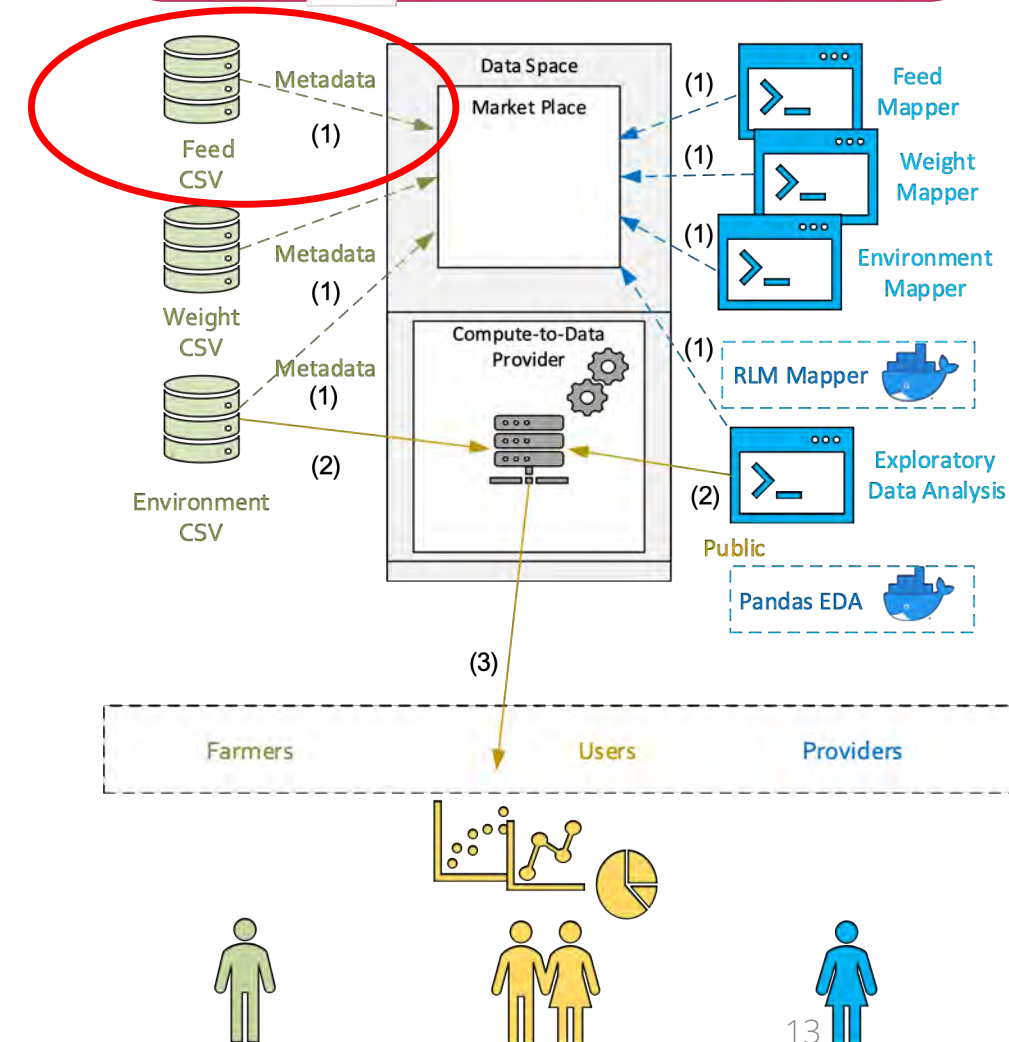
2 sales

A14
Pork



STEP 1

Publishing & Reading



Exploratory Data Analysis

GEN-X Testnet

Owned by 0x4cF0...2E17
Accessed with GXAT

ALGORITHM Published 19 days ago

Generate an **Exploratory Data Analysis (EDA)** report for the input tabular data.

The input data is loaded using **pandas** and then the EDA report is generated using **pandas-profiling**. The HTML version of the report can be downloaded when the algorithm finishes.

eda exploratory-data-analysis profiling pandas tabular-data

udl

DATA AUTHOR: OWNER: 0x4cF0...2E17

DOCKER IMAGE: rogarcon/pandas-profiling: DID: did:op:e73ac9c332adfea204acefb4b300e0dde977ca9dee99017a8ed94242dc4bfa53

METADATA HISTORY: NFT metadata updated 19 days ago

DDO

plain 695 B url Free

This algorithm has been set to private by the publisher and can't be downloaded. You can run it against any allowed datasets though!

Datasets algorithm is allowed to run on

CEP - Automatic Pig Feeding - 2021 S1 - 982091062894196	Free
CEP - Automatic Pig Feeding - 2021 - Pens 4 & 5	1
CEP - Daily Pig Weight - 2021 - Pens 4 & 5	Free

Consent

Your Compute Jobs SHOW

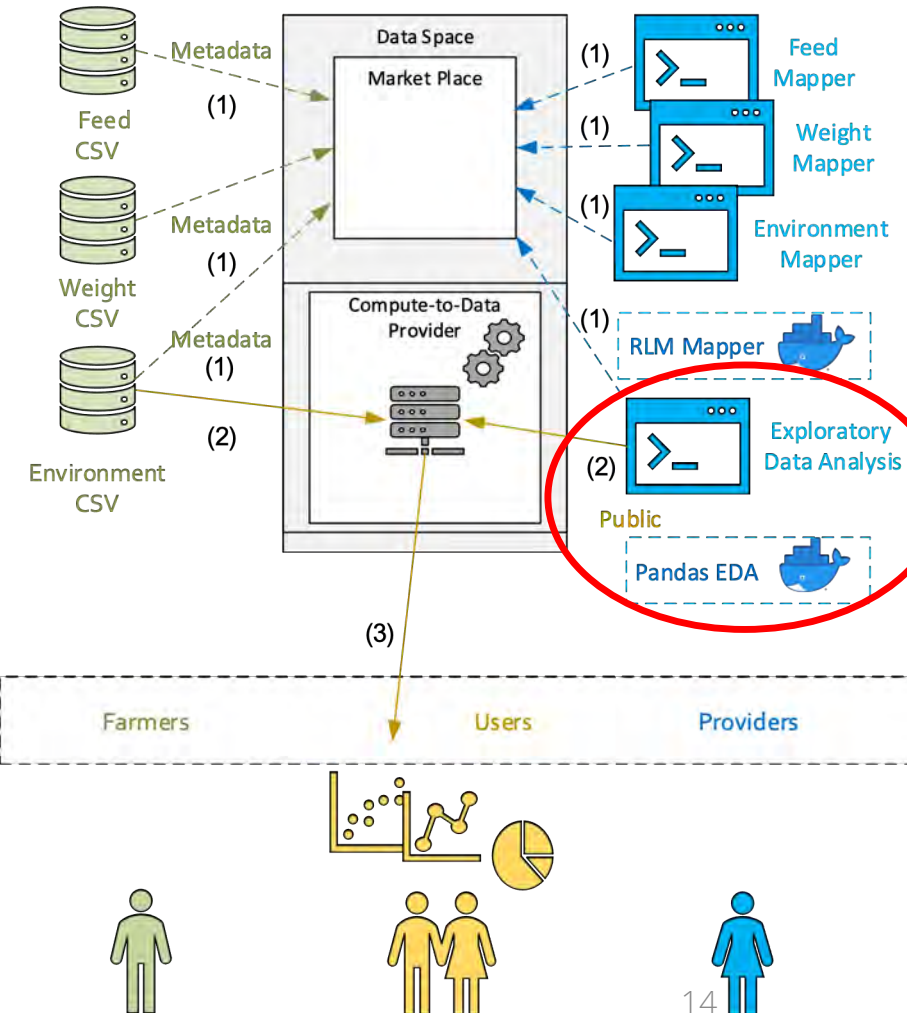
6 sales

A14
Pork



STEP 1

Publishing & Reading



Compute-to-Data in the "Data Room"

STEP 1

Publishing & Reading

1



CATALOGUE PUBLISH VERIFY LOG ECOSYSTEM

Exploratory Data Analysis

GEN-X Testnet

Owned by 0x4cF0...2E17
Accessed with GXAT

ALGORITHM Published 19 days ago

Generate an **Exploratory Data Analysis (EDA)** report for the input tabular data.

The input data is loaded using **pandas** and then the EDA report is generated using **pandas-profiling**. The HTML version of the report can be downloaded when the algorithm finishes.

eda exploratory-data-analysis profiling pandas tabular-data

udl

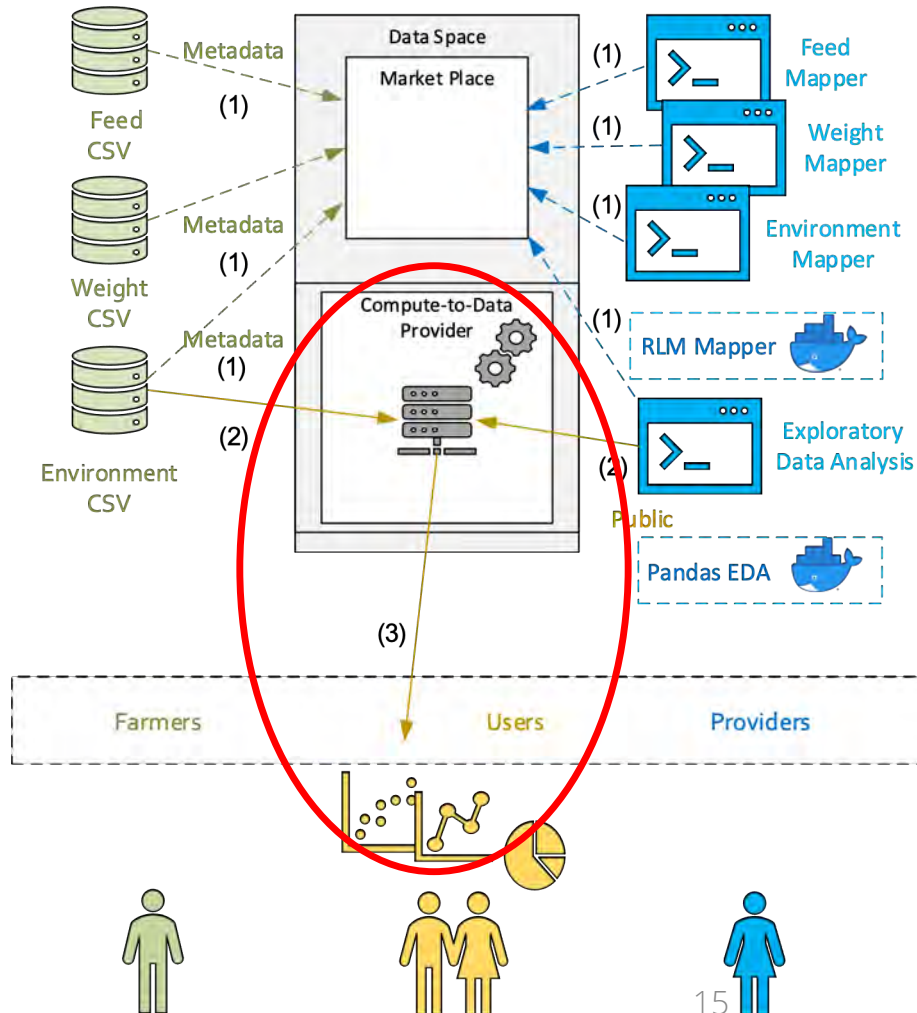
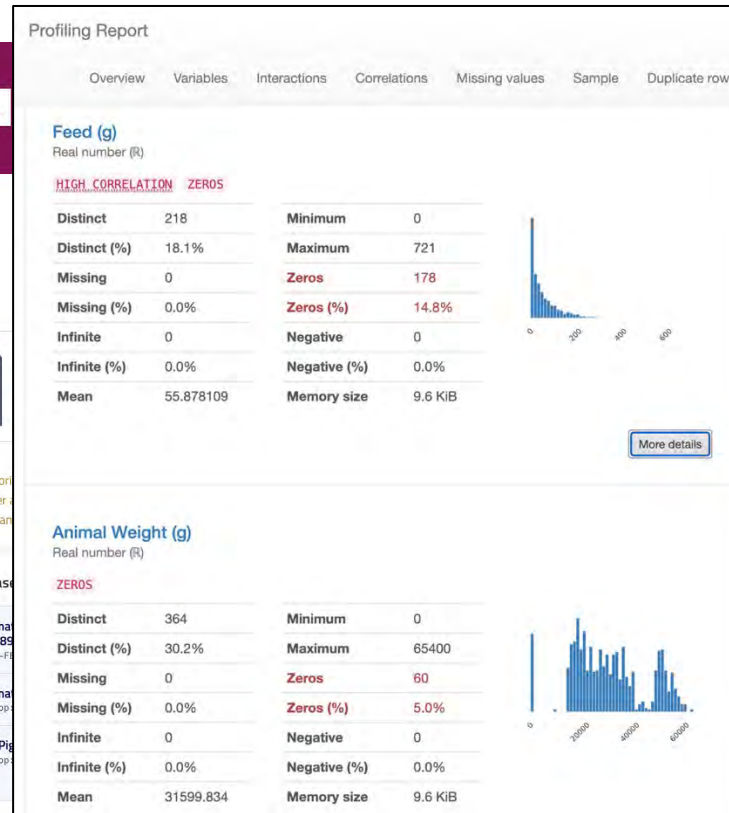
DATA AUTHOR

OWNER
0x4cF0...2E17

DOCKER IMAGE
rogargon/pandas-profiling:

METADATA HISTORY
• NFT metadata updated 19 days ago

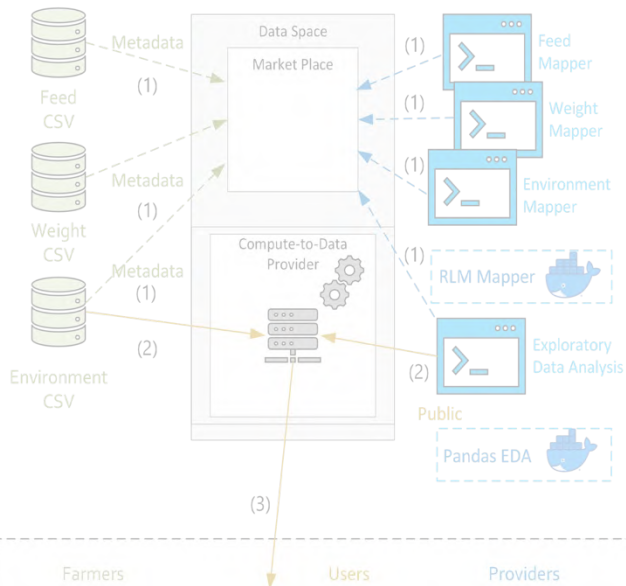
DDO



AI4Pork Use Case

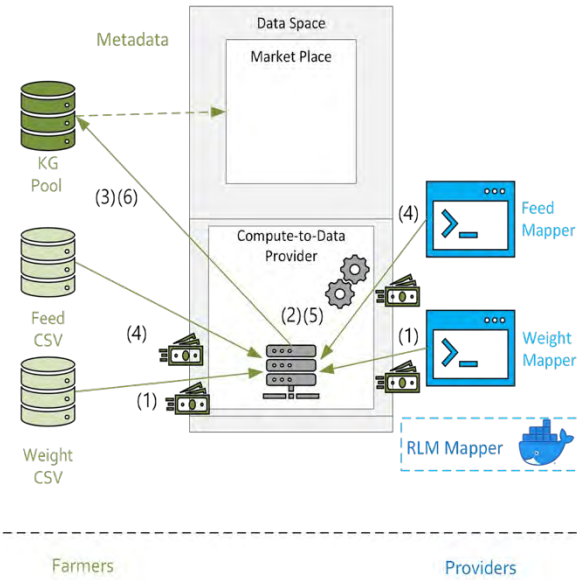
STEP 1

Publishing & Reading



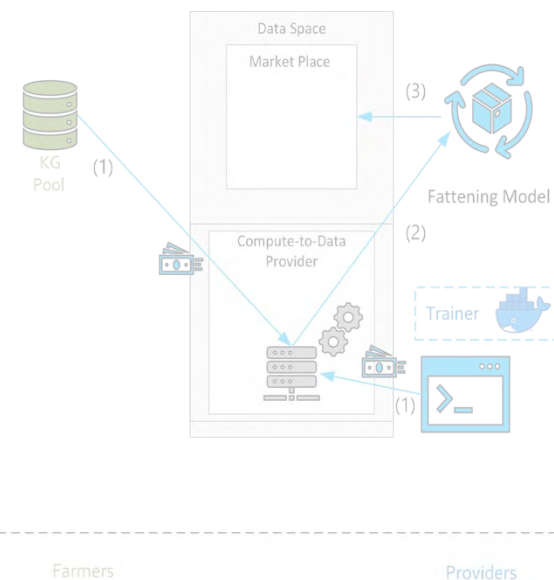
STEP 2

Semantic integration & pooling



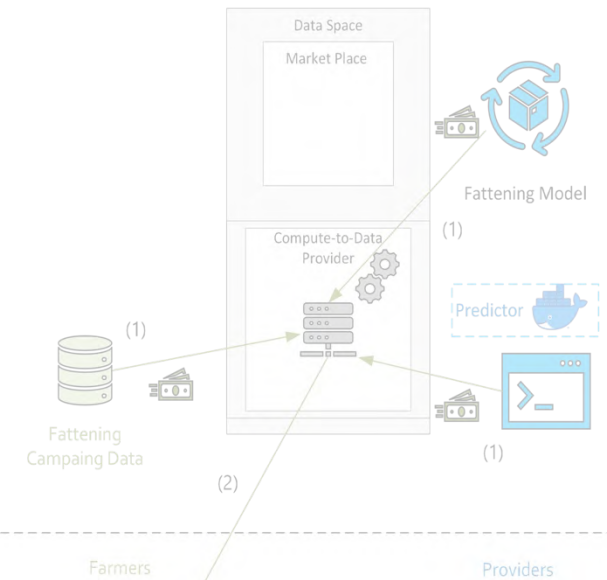
STEP 3

Training (Federated Learning)

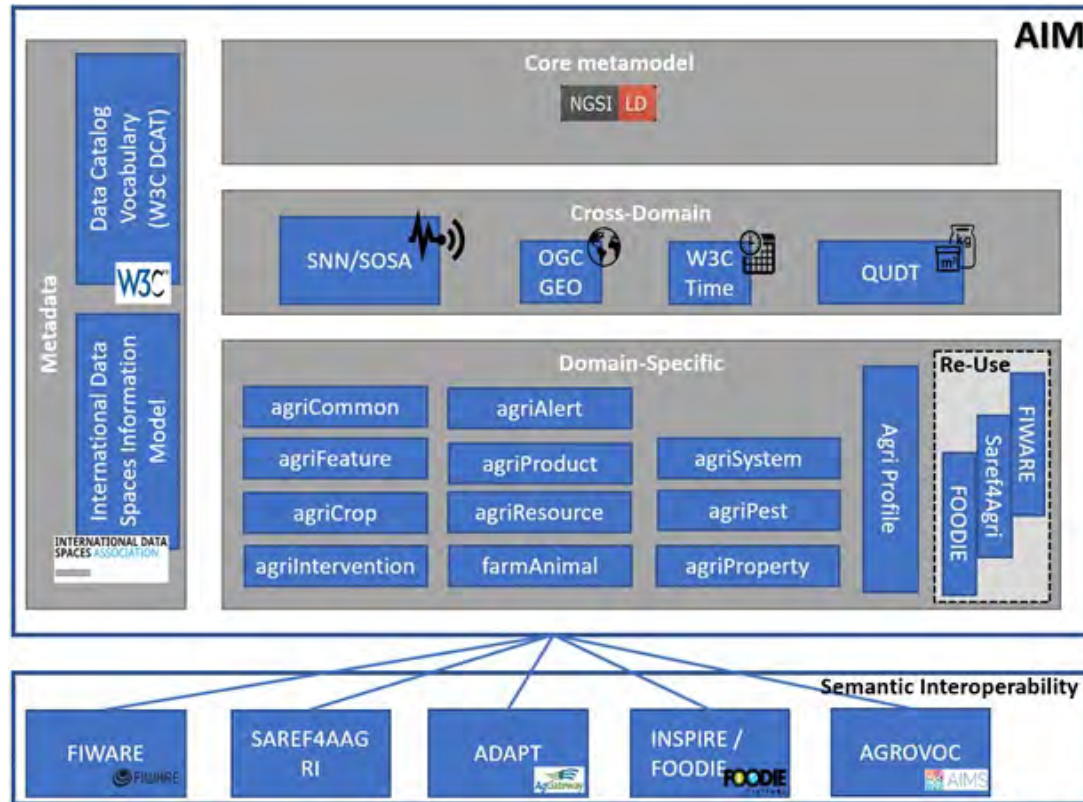


STEP 4

Predicting

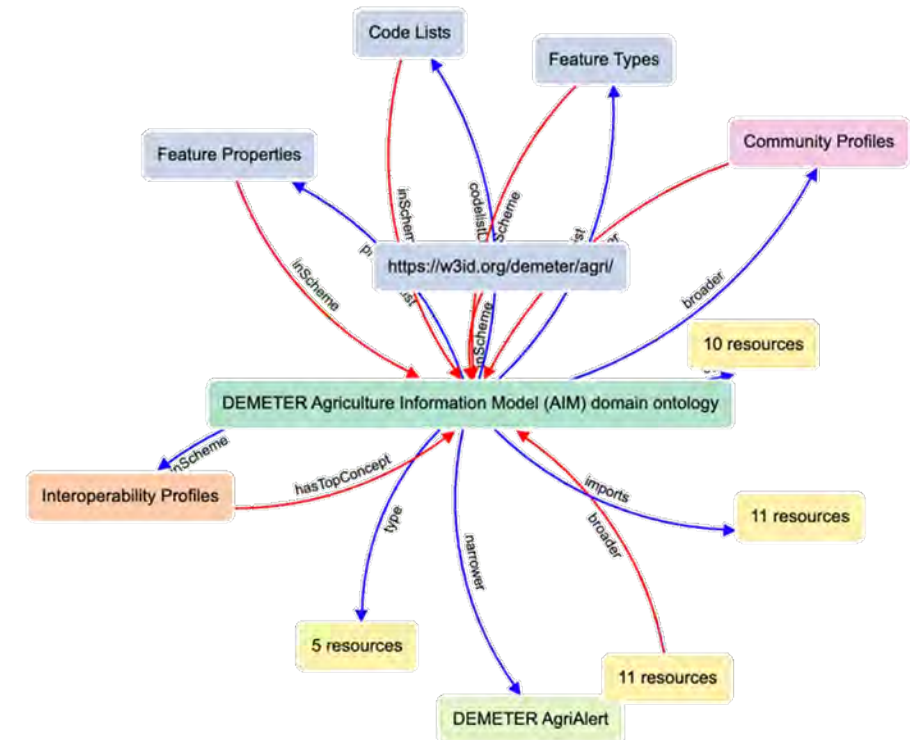


Vocabularies and Ontologies Agriculture Information Model (AIM)



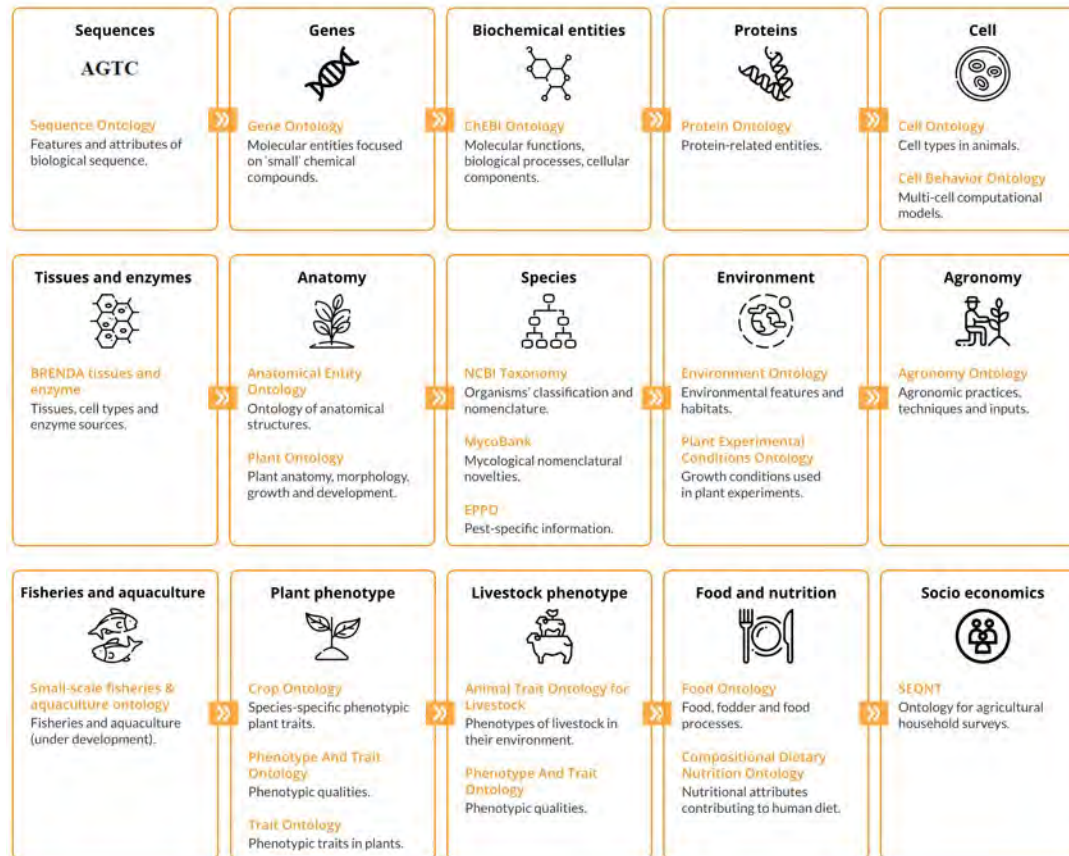
<https://w3id.org/demeter/agri>

OGC Vocabulary:



Vocabularies and Ontologies

CGIAR: Ontologies most used in Agriculture



- **GENERIC ONTOLOGIES AND THESAURI**
 - **AGROVOC**: Multilingual and controlled vocabulary designed to cover concepts and terminology under FAO's areas of interest.
 - **CAB Thesaurus**: Controlled vocabulary for agriculture, forestry, horticulture, soil science, entomology, mycology, parasitology, veterinary medicine, nutrition, rural studies.
 - **CG Core Metadata Schema**: Minimum set of elements applicable for data and publication annotation and curation across CGIAR.
 - **Comparative Data Analysis Ontology (CDAO)**: Formalization of concepts and relations relevant to evolutionary comparative analysis.
 - **EMBRACE Data and Methods Ontology (EDAM)**: Comprehensive ontology for bioscientific data analysis and data management.
 - **GAZ**: Represents places through their names using an ontological approach to promote semantic coherence.
 - **Geonames**: Covers all countries and contains over eleven million placenames.
 - **National Agricultural Library's Agricultural Thesaurus and Glossary (NAL)**: Online vocabulary tools English and Spanish to select controlled vocabulary terms for subject indexing of AGRICOLA, PubAg and other databases.
 - **Research Organization Registry (ROR)**: Registry of open, sustainable, usable, and unique identifiers for every research organization in the world.
 - **Units of Measurement Ontology (UO)**: Ontology of units of measurements.
 - **Variation Ontology (VariO)**: Ontology for standardized, systematic description of effects, consequences and mechanisms of variations.
 - **VIVO Ontology**: Open-source software and an ontology for representing scholarship.

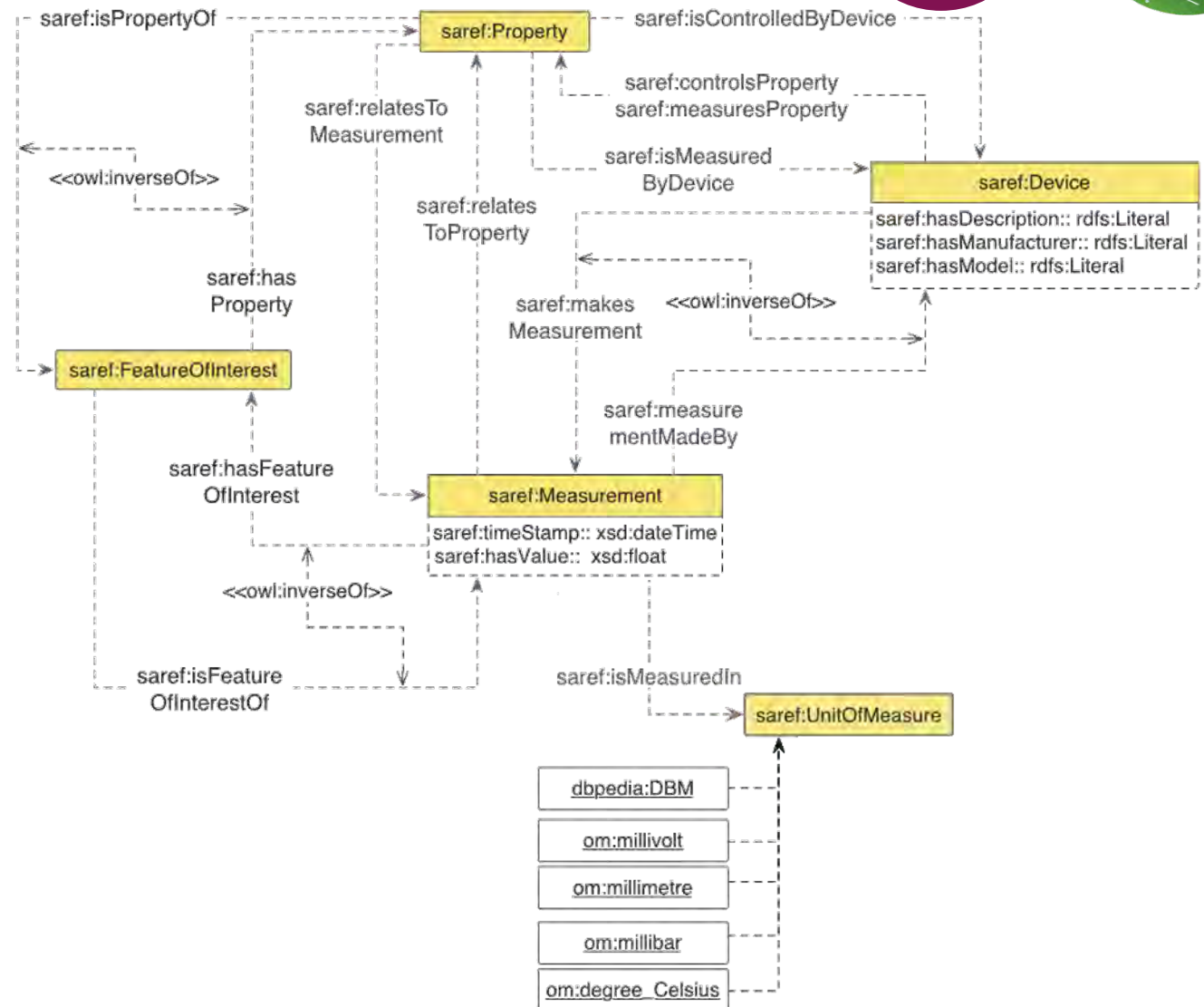
SAREF

Smart Applications REference ontology

<https://saref.etsi.org>

Extensions:

- [SAREF4ENER](#): Energy domain
- [SAREF4ENVI](#): Environment Domain
- [SAREF4BLDG](#): Building domain
- [SAREF4CITY](#): Smart Cities domain
- [SAREF4INMA](#): Industry and Manufacturing domains
- [SAREF4AGRI](#): Smart Agriculture and Food Chain domains
- [SAREF4AUTO](#): Automotive domain
- [SAREF4EHAW](#): eHealth/Ageing-well domain
- [SAREF4WEAR](#): Wearables domain
- [SAREF4WATR](#): Water domain
- [SAREF4LIFT](#): Smart Lifts domain
- [SAREF4GRID](#): Smart Grid domain



CEP's CSV Data Mapper and Semantic Data Pooler

GEN-X Testnet

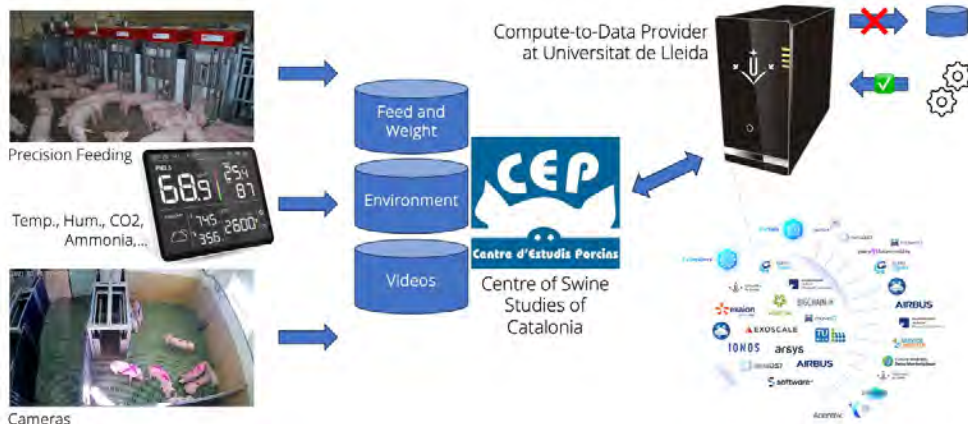
Owned by 0x6207...6822
Accessed with CEP-RML 71

COMPUTE ALGORITHM Published 2 months ago

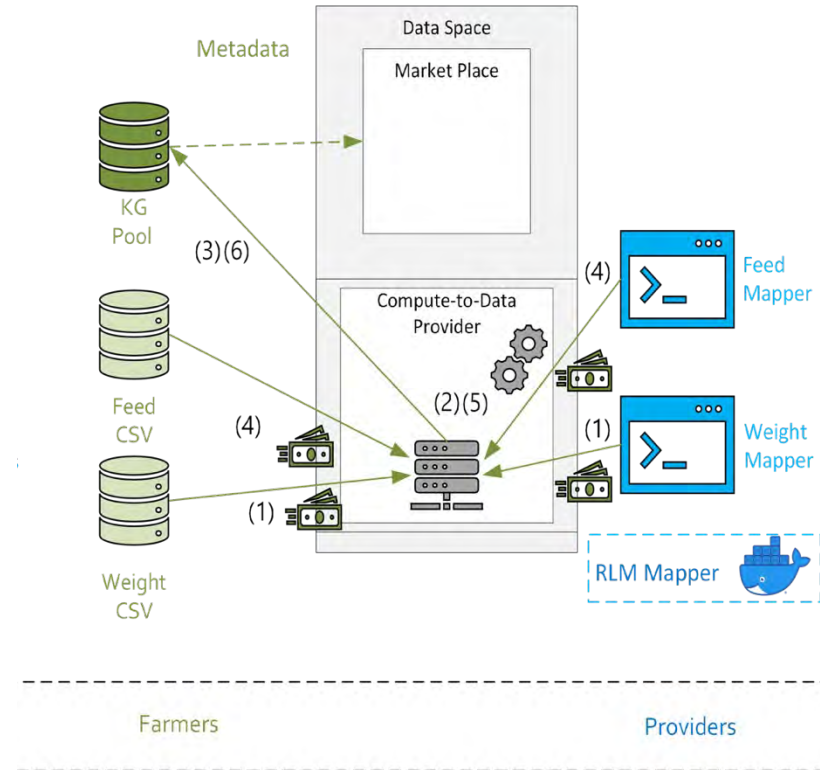
This algorithm supports the "Pay-as-you-go" approach when sharing data through a data space. Instead of requiring that publishers integrate the data based on existing schemas, which causes a significant upfront overhead, the pay-as-you-go paradigm favours an incremental approach. This way, entry barriers are lowered and it becomes easier to share data.

Moreover, data sovereignty is guaranteed by design through a **Data Room** implemented using "Compute-to-Data". The mapped data does not leave the room, it is processed and then stored into a Knowledge Graph that stays inside the room. This way, it remains under the control of the data originator, the **Centre of Swine Studies of Catalonia (CEP)**.

Later, CEP can decide to grant access to trusted algorithm to visit the Data Room and slice the Knowledge Graph to extract the semantically integrated data relevant to their computations. Also in this case, data sovereignty is guaranteed as just computation results, like aggregations or AI trained models, can leave the room, not the original data or subsets of it.



STEP 2 Semantic integration & pooling



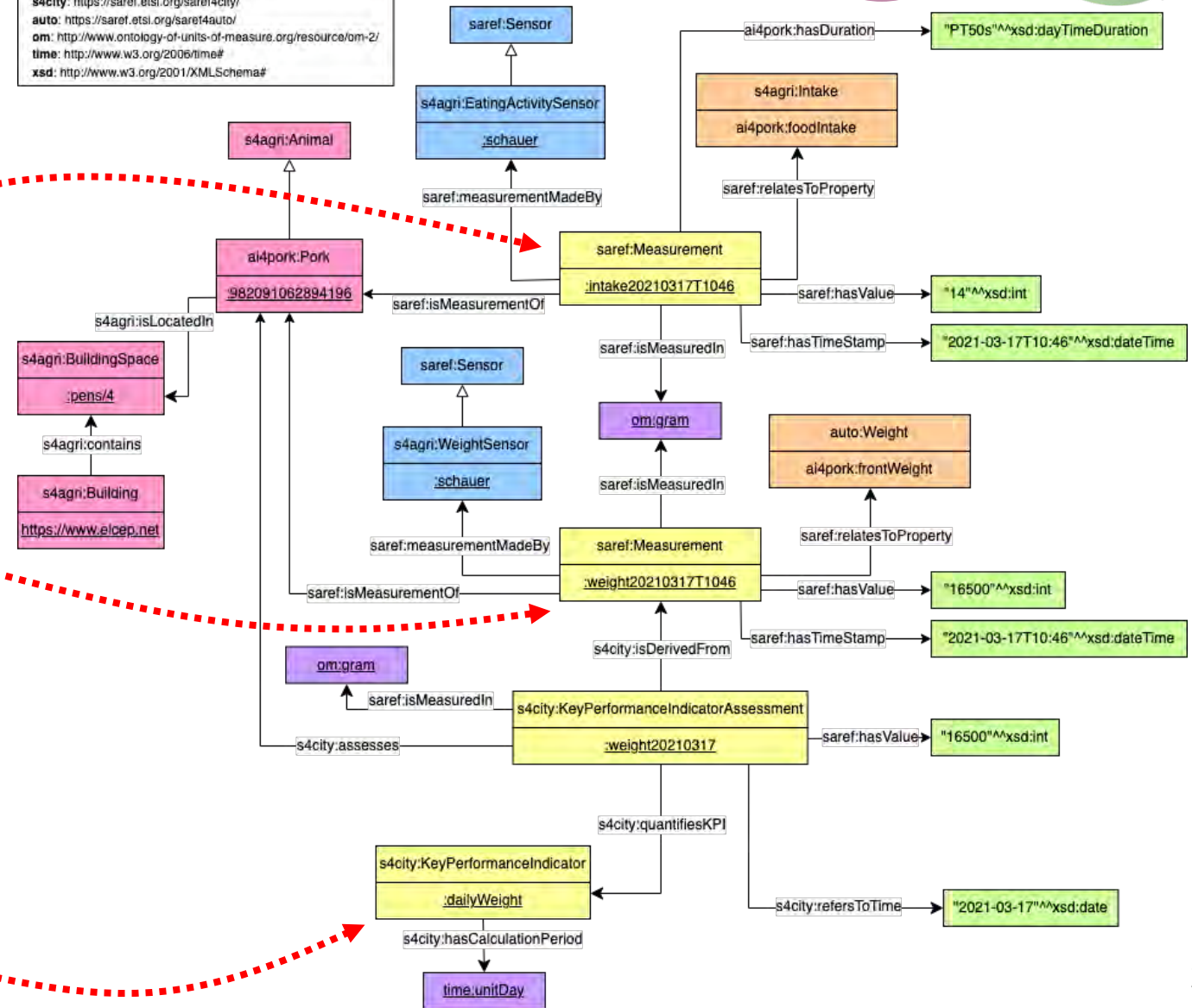
Feed CSV

Pen Number	Animal ID	Date	Time	Duration (s)	Feed (g)	Animal Weight (g)
4	982091062894196	2021-03-16	13:05	70	0	15300
4	982091062894196	2021-03-17	10:44	5	0	14400
4	982091062894196	2021-03-17	10:45	9	6	14500
4	982091062894196	2021-03-17	10:46	50	14	16500

Daily Weight CSV

Animal ID	Date	Weight (g)
982091062894196	2021-03-16	16300
982091062894196	2021-03-17	16500

base: https://ai4pork.angliru.udi.cat/schauer/
 ai4pork: https://ai4pork.angliru.udi.cat/ontology/
 saref: https://saref.etsi.org/core/
 s4agri: https://saref.etsi.org/saref4agri/
 s4city: https://saref.etsi.org/saref4city/
 auto: https://saref.etsi.org/saref4auto/
 om: http://www.ontology-of-units-of-measure.org/resource/om-2/
 time: http://www.w3.org/2006/time#
 xsd: http://www.w3.org/2001/XMLSchema#

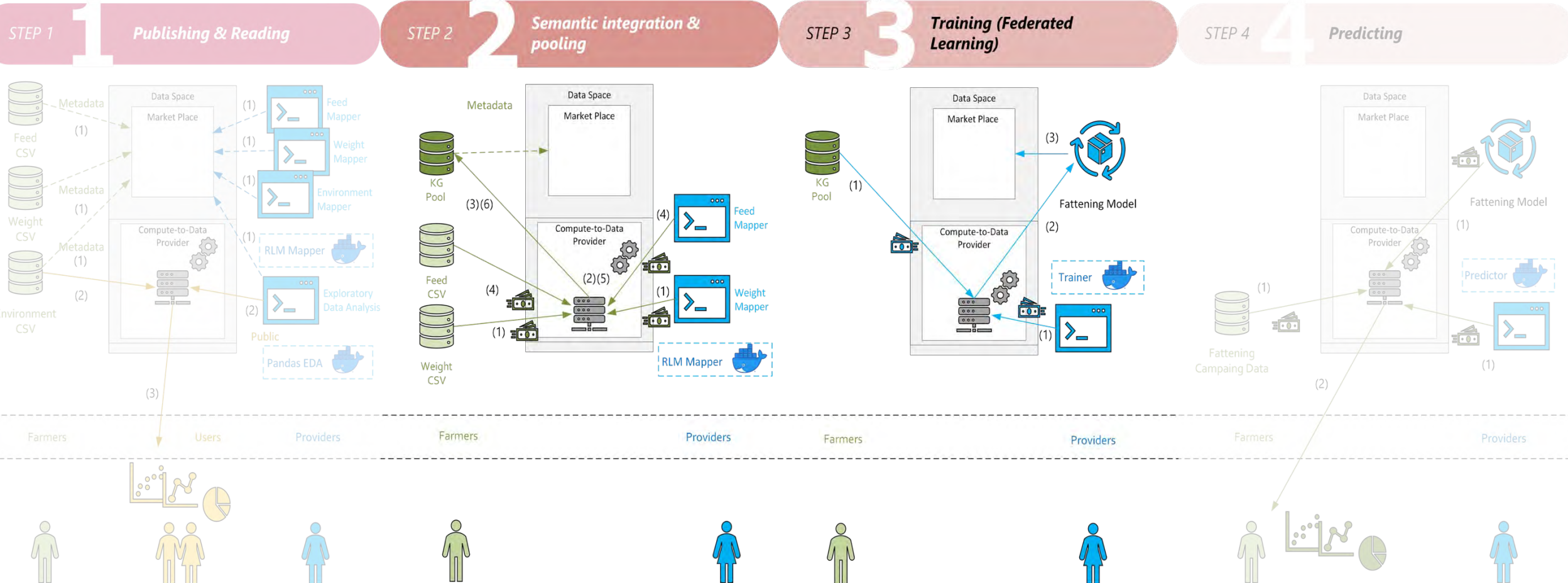




YARRRML Mapping

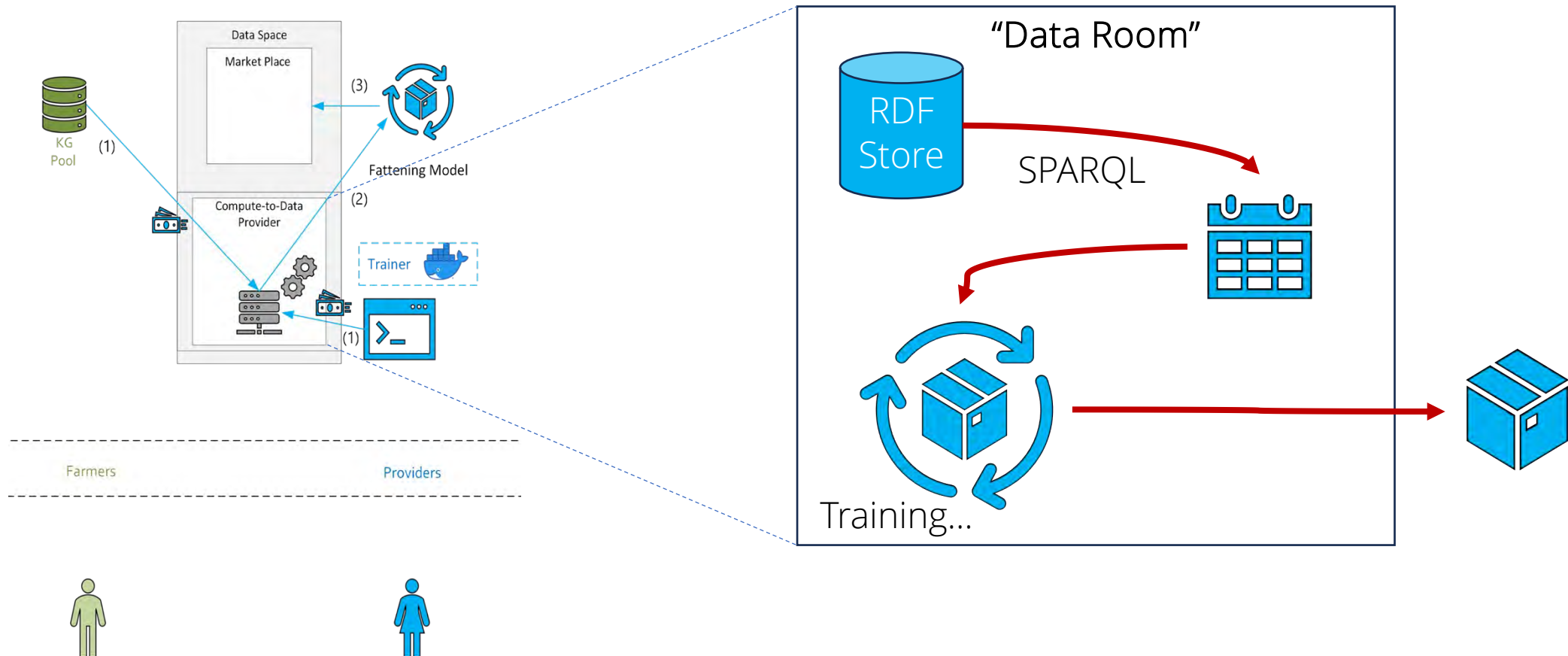
- A human readable text-based representation for declarative Linked Data generation rules.
- Input
 - Tabular: CSV, TSV, MySQL, PostgreSQL...
 - Tree: JSON or XML
- Output: Knowledge Graph
- Interactive example:
 - YARRRML Online: <https://rml.io/yarrml/matey/>
 1. Input: <https://raw.githubusercontent.com/rogargon/ocean-algo/rmlmapper/data/inputs/8f67E08be5dD941a701c2491E814535522c33bC2/0>
 2. YARRRML Mapping: <https://raw.githubusercontent.com/rogargon/ocean-algo/rmlmapper/feed-map.yml>
 3. Visualize output: <https://rdf2svg.redefer.rhizomik.net/>

AI4Pork Use Case

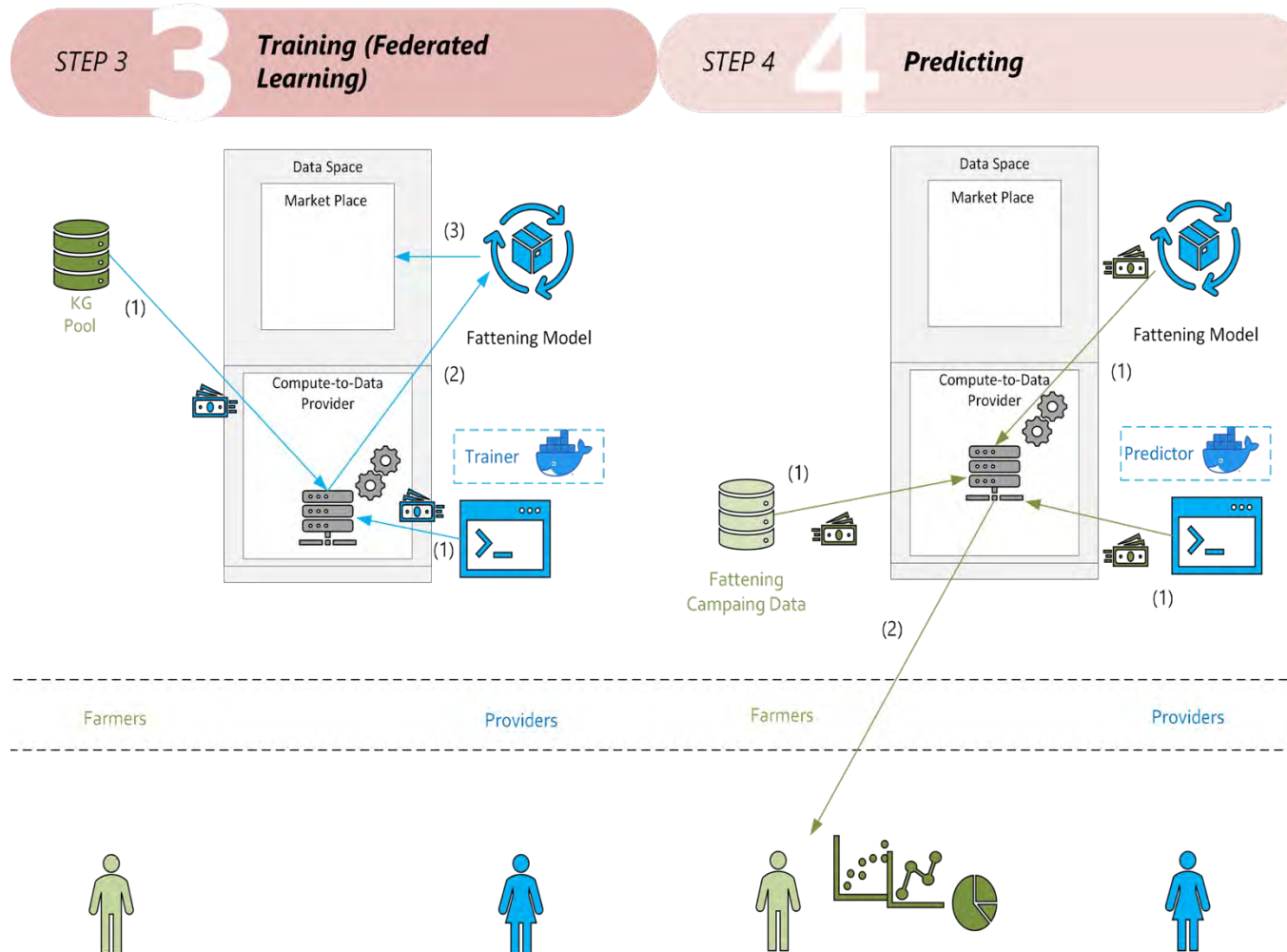


Sovereignty-preserving Data Exploitation

STEP 3 **3** Training (Federated Learning)



Data & AI Models Monetization



You will pay **2.125 EUROe** ⓘ

START	DATASET	1 EUROe
	+ ALGORITHM	1 EUROe
	+ C2D RESOURCES	0.125 EUROe for 898 seconds
	=	2.125 EUROe

You bought this dataset already al
already bought the selected algori
again. The C2D resources required
required for them. Please note tha
free assets.

Access allowed ✓

Your Compute Jobs **SHOW** ▼

1 sale

Gaia-X Integration

Exploratory Data Analysis

GEN-X Testnet

Owned by 0x6207...6822
Accessed with UDL-EDA

DOWNLOAD ALGORITHM Published 3 months ago by Universitat de Lleida (UdL)

Service Credential
Credential ID match
version: 22.10
last check: less than a minute ago

Generate an **Exploratory Data Analysis (EDA)** report for the input tabular data.

The input data is loaded using **pandas** and then the EDA report is generated using **pandas-profiling**. The HTML version of the report can be downloaded when the algorithm

The report includes the following information:

Data Quality Alerts

Overview

Overview Alerts 25 Reproduction

Alerts

- source**: has constant value "NASA"
- Dataset** has 10 (< 0.1%) duplicate rows
- name**: has a high cardinality: 45716 distinct values
- recclass**: has a high cardinality: 466 distinct values
- GeoLocation**: has a high cardinality: 17100 distinct values
- reclat**: is highly correlated with **reclong** and 1 other fields
- reclong**: is highly correlated with **reclat** and 1 other fields

Datasets algorithm is allowed to run on

Dataset	Access
CEP - Automatic Pig Feeding - 2021 S1 - 982091062894506 CEP-FEED did:op:bb5f52d48cf3294e9e979ac591ac75830ea98d8730e183d389d26620...	Free
CEP - Automatic Pig Feeding - 2021 S1 - 982091062894406 (UdL) GXAT did:op:a11ea853ef78f2210e154a9785ac23801880b616fa71dab95b6eeca67bc...	Free

Access allowed

10 sales

Verifiable Presentation
https://angliru.udl.cat/well-known/2210_serviceOffering-eda.json

gx-compliance 2.0.0 GAS 1.0

Prototype for a compliance service as defined in <https://gaia-x.github.io/policy-rules-committee/trust-framework/>

credential-offer

POST /development/api/credential-offers Check Gaia-X compliance rules and outputs a VerifiableCredentials from your VerifiablePresentation

Parameters

Name	Description
vcid string (query)	Output VC ID. Optional. Should be url_encoded if an URL Example : https://storage.gala-x.eu/credential-offers/b3e0a068-4bf8-4796-932e-2fa83043e203 https://storage.gala-x.eu/credential-offers/b3e0a068-4bf8-4796-932e-2fa83043e203

Request body *required*

A VerifiablePresentation in JSON-LD or JWT format

Examples:
Participant

PONTUS-X
Testnet Explorer

Blocks Transactions Tokens APIs Search by address, token symbol, name, transact

Daily transactions

Average block time: 4.2 seconds
Total transactions: 26,603
Total blocks: 13,052,368
Wallet addresses: 4,193

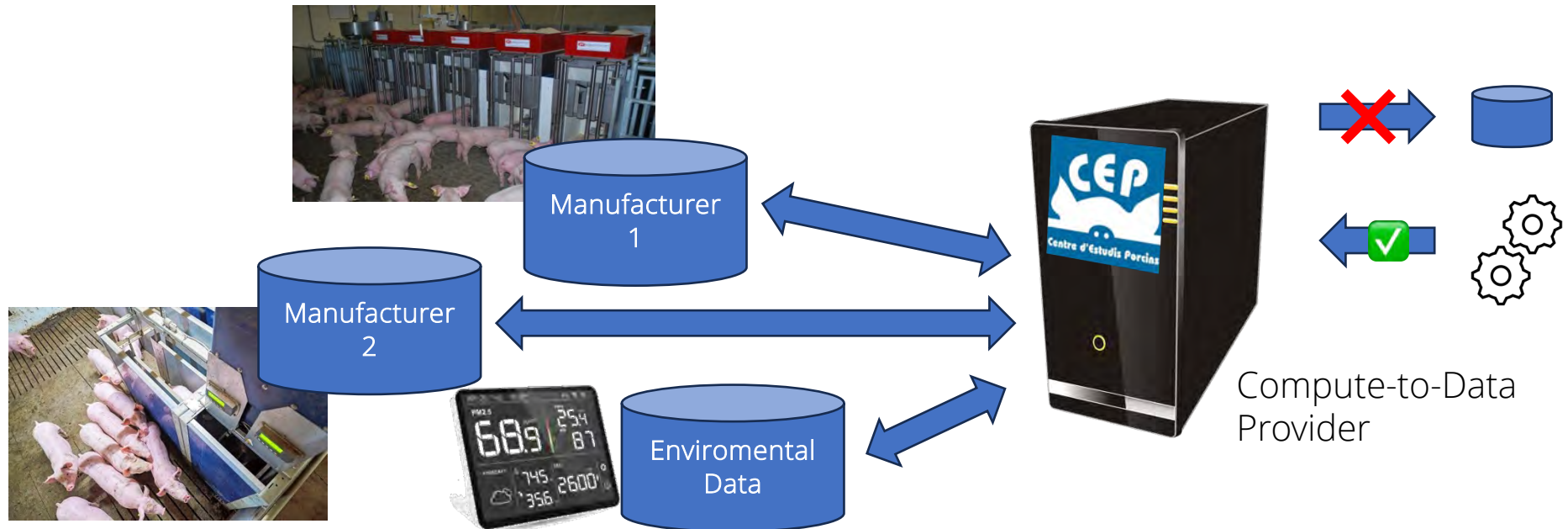
Blocks

View All Blocks

Block ID	Transactions	Time	Miner	Reward
13052367	0 Transactions	8 seconds ago	Miner 0x00000000000000000000000000000000...	Reward 0 GX
13052366	0 Transactions	10 seconds ago	Miner 0x00000000000000000000000000000000...	Reward 0 GX
13052365	0 Transactions	12 seconds ago	Miner 0x00000000000000000000000000000000...	Reward 0 GX
13052364	0 Transactions	14 seconds ago	Miner 0x00000000000000000000000000000000...	Reward 0 GX

Future Work: AgrifoodTEF Data Space

- Data Space's "Data Room" at the Centre for Swine Studies (CEP)
- Validation of precision feeding solutions from different manufacturers
- AI models training without revealing their data
 - Advantageous to share with other manufacturers to improve one's own model



AgrifoodTEF

EU Testing and Experimenting Facilities

- European reference for **testing** and **experimentation** for agri-food innovation
 - Innovation and **digitalization** for sustainable and efficient food production
 - Network of **physical** and **digital** facilities across Europe
 - Services to evaluate and **validate third-party AI and robotics solutions** in real-world conditions
 - €60M funding (50% EU / 50% EM)
- **Nodes (10M€):**
 - Italy
 - France
- **Satellites (5M€)**
 - Belgium
 - Holland
 - Poland
 - Sweden
 - Austria
 - Spain
- **Affiliated:**
 - Germany



<https://www.agrifoodtef.eu>

Thank you for your time Questions?

Roberto García

Universitat de Lleida

<https://rhizomik.net/~roberto>

Acknowledgements

[AI4Pork](#) (TED2021-130829B-100) funded by MCIN/ AEI/10.13039/501100011033 and by the European Union NextGenerationEU/PRTR

[ANGLIRU](#) (PID2020-117912RB-C22) funded by MCIN/AEI/10.13039/501100011033

[AgriDataSpace](#) (101083401) funded by the European Union

[AgrifoodTEF](#) (101100622) cofunded by the Spanish Ministry of Agriculture, Fisheries and Food together with the European Union