

# Soberanía por diseño en espacios de datos agrarios

Roberto García

Universitat de Lleida

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

# Introducción

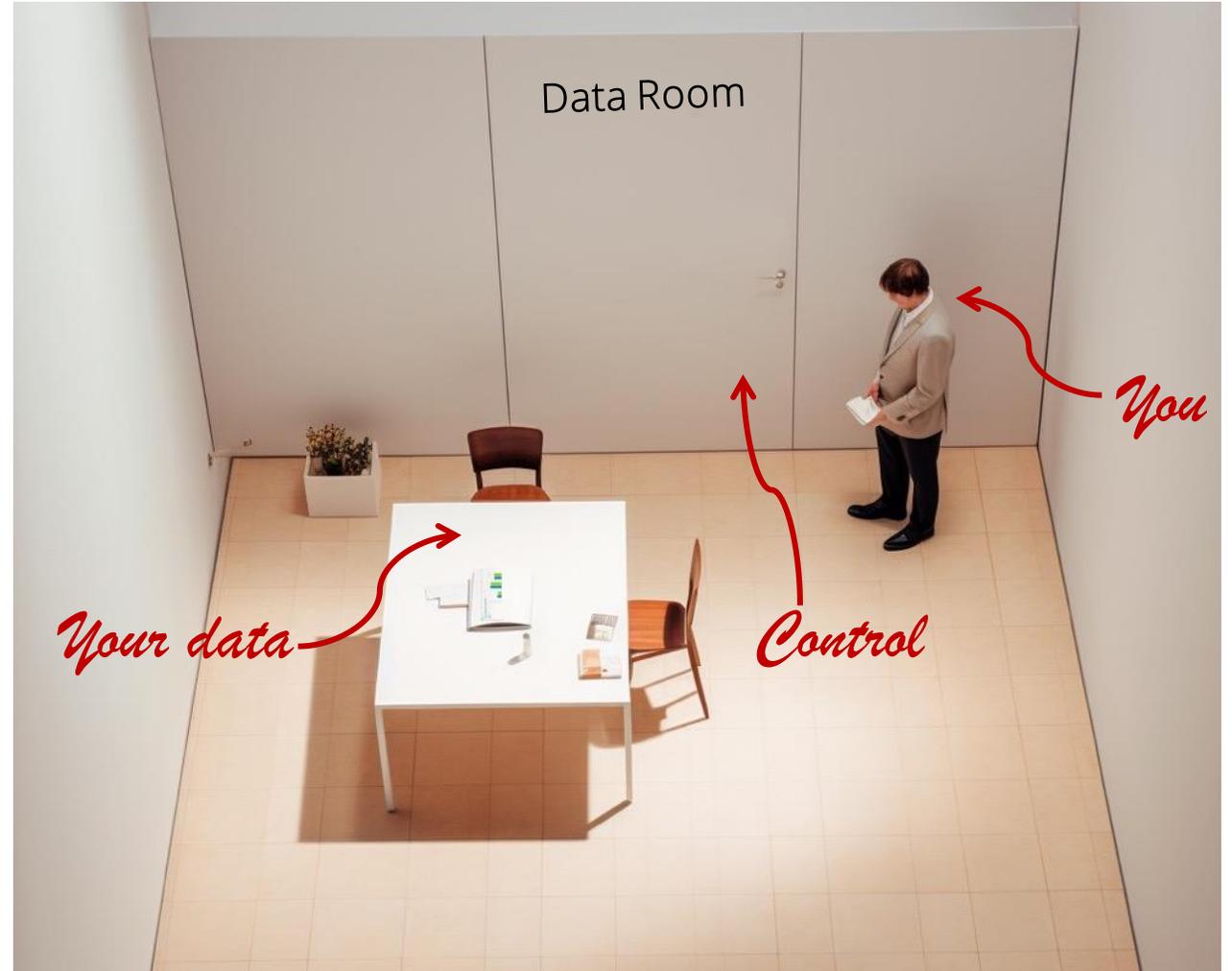
- Importancia creciente de la **digitalización** y los **datos** para la transformación del sector agrario
- Compartir **más allá** de los datos abiertos:
  - Complicado con datos personales, sensibles, estratégicos...
- **Espacios de Datos**: una infraestructura descentralizada y abierta para el intercambio **soberano** de datos
  - Los participantes **controlan** los datos que se producen y consumen
  - No hay un almacén central de datos, **intercambios directos** entre los participantes

# ¿Cómo motivar compartir datos?

- Monetización (y tokenización) como incentivo
- Entorno confiable para compartirlas sin perder su control
  - "tragedia" de lo digital, tan fácil copiar y distribuir...
    - Ejemplo: música digital, Sistemas Protección Copyright (DRM),... "streaming"
  - Oportunidad en el caso de los datos:  
¿Soberanía de Datos "por diseño"? Alternativas:
    - Conectores de Espacios de Datos (tipo punto-a-punto)
      - Difícil garantizar aplicación control datos en entornos sin relaciones de confianza sólidas previas (similar al caso del DRM)
    - "Habitaciones de datos" o "Data Rooms" con Computación-a-Datos?

## Soberanía de Datos por diseño

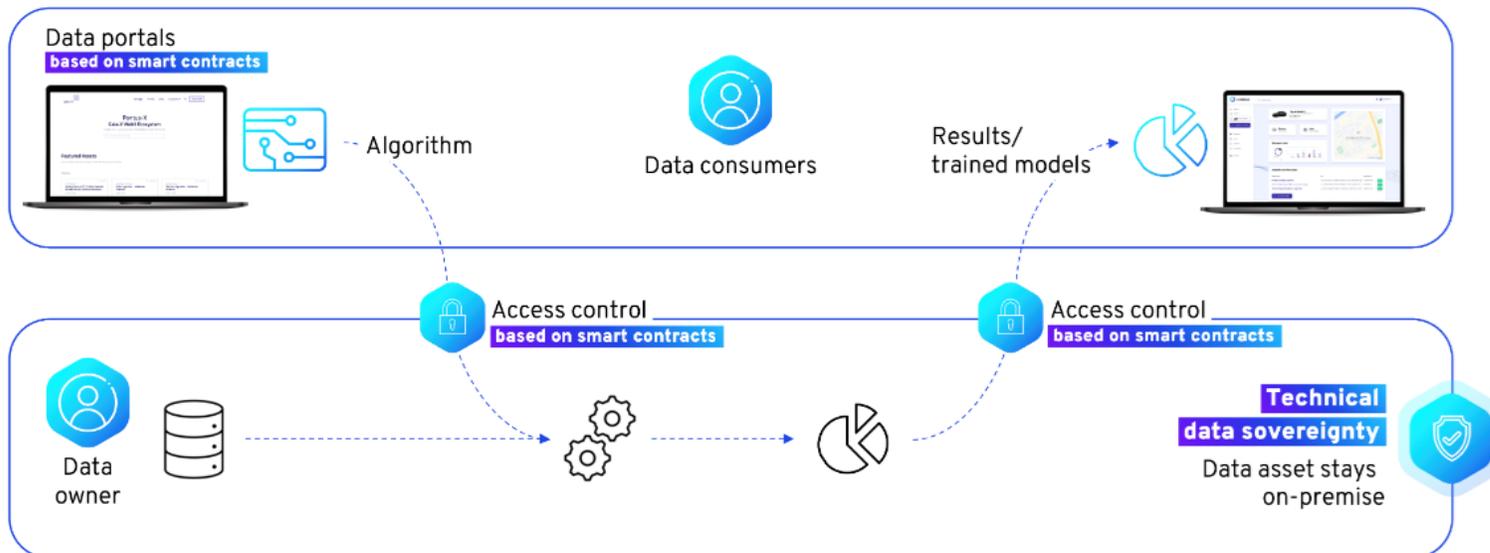
- Sus datos:  
Nunca "dejan la habitación"
- Usted:  
controla quién "entra" y qué "se llevan" con ellos
  - ✗ Copias completas o parciales
  - ✓ Datos agregados, estadísticos
  - ✓ Modelos IA entrenados
  - ✗ Datos personales
  - ...



dar01995\_a\_bird-view\_of\_the\_inside\_of\_a\_whole\_room\_without\_wind\_1b2e5c91-5513-471d-9673-6c6493a76c11

# Computación-a-Datos

- Facilita la privacidad y la soberanía de datos por defecto
- Los algoritmos se ejecutan donde se almacenan los datos
  - Los originadores de datos mantienen el control, no se copian en ningún otro lugar



# Ecosistema Pontus-X de Gaia-X



- Gaia-X, principal iniciativa europea para el desarrollo de espacios de datos:
  - Establecer un **ecosistema** para **compartir datos** en un entorno de confianza
  - Devolver el **control** a los usuarios, **soberanía** sobre sus datos
  - Sistema **federado** y **transparente** de servicios en la nube que impulsará la **economía europea de datos**

# Caso de uso AI4Pork

Proveedor de Computación-a-Datos  
Universitat de Lleida



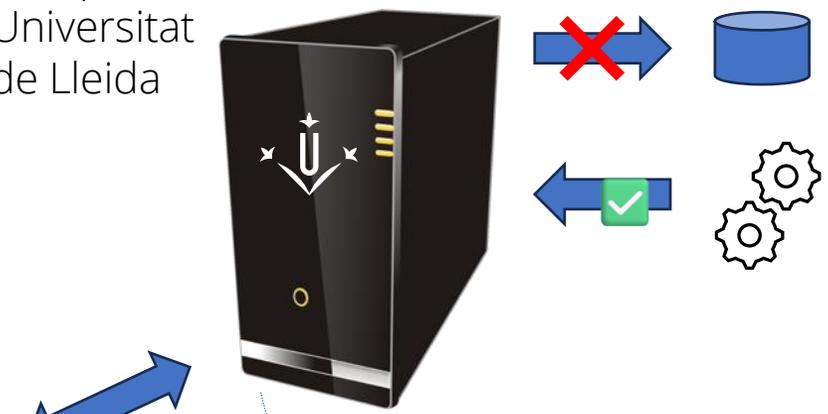
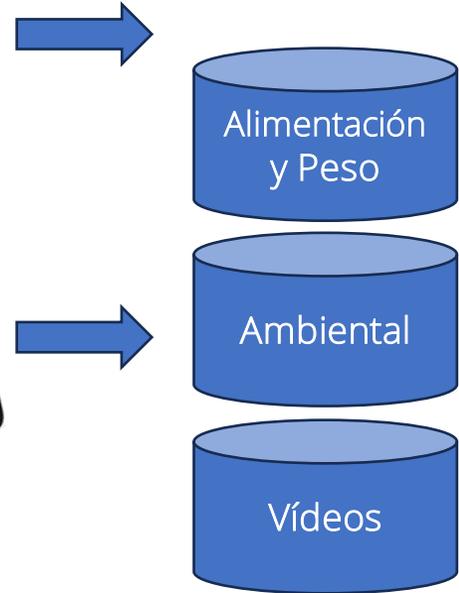
Alimentación de precisión



Temp., Hum., CO<sub>2</sub>, Amonio,...



Cámaras



# Portal del Espacio de Datos

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

CATALOGUE PUBLISH VERIFY LOG ECOSYSTEM
GEN-X Testnet 1631
Ox6207...6822

## 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**  
 Ox38f8...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)**  
 Ox3BB6...BE40  
 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

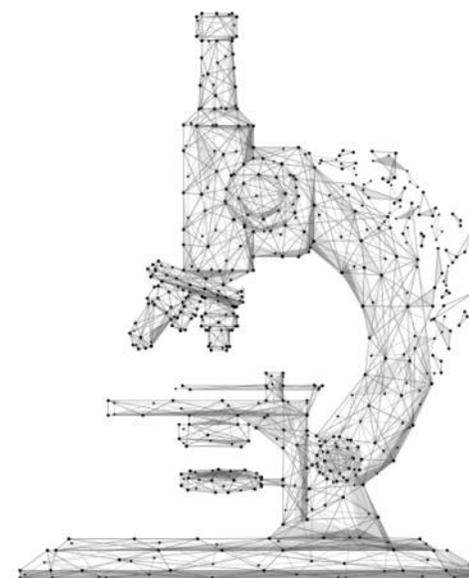
### Precision Pig Feeding Semantic Data Integration and Sovereign Data Pooling

COMPUTE ALGORITHM CEP-RML  
**CEP's CSV Data Mapper and Semantic Data Pooler**  
 Ox6207...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**  
 Ox38f8...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**  
 Ox6207...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.



# Caso de uso AI4Pork

Proveedor de Computación-a-Datos  
Universitat de Lleida



Alimentación de precisión



Temp., Hum., CO<sub>2</sub>, Amonio,...



Cámaras





# Pig pen images sequence for animal well-being assessment

GEN-X Testnet

Owned by 0x38f8...840d  
Accessed with GXAT

COMPUTE DATASET Published about 2 months ago

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 up of the Diputació de Lleida, the Regional Council of La Noguera, the Torrelameu Town Hall and the Universitat de Lleida.

The images can be used by animal well-being assessment algorithms available from the compute section on the right. These algorithms perform automatic image segmentation and tracking to identify and track pig movements in the sequence of images. Additionally, it is also possible to monitor the visits of pigs to defined areas of interest like the automatic feeding machine or the waterer bowl. This allows for the automatic generation of metrics that can be used for animal well-being assessment.

Data sovereignty is guaranteed by design through a **Data Room** implemented using "Compute-to-Data". The algorithm visits the image sequence inside the data room, where they are analysed, and just the computed metrics leave the room. Consequently, there is no leakage of any image from inside the farm. They are just copied to the data room and destroyed after the computation without leaving it.

The available animal well-being assessment algorithm computes the following metrics from sequences of images:

### Movement metric per pig

Example:

```
{
  "0": 2438.4733276550605,
  "1": 533.9017133622191,
  "2": 1411.6580468122302,
  "3": 1481.5184106820652,
  "4": 1863.6384070993606,
  "5": 2910.0410553493325,
  "6": 665.6211681307876,
  "7": 702.4427292321657,
  "8": 1212.4275062016434,
  "9": 1331.9712414388393
}
```

SAMPLE DATA

DOWNLOAD SAMPLE

### Time spent on the automatic feeding machine area per pig

Example:

```
{
  "0": { "frames": 5, "time": 0.2 },
  "1": { "frames": 0, "time": 0 },
  "2": { "frames": 0, "time": 0 },
  "3": { "frames": 0, "time": 0 },
  "4": { "frames": 1, "time": 0.04 },
  "5": { "frames": 8, "time": 0.32 },
  "6": { "frames": 0, "time": 0 },
  "7": { "frames": 0, "time": 0 },
  "8": { "frames": 0, "time": 0 },
  "9": { "frames": 12, "time": 0.48 }
}
```

Select a Compute Environment

**UdL Data Room**

CPU | GPU | max duration: 1 hour  
0.01 OCEAN / minute

Select an algorithm to start a compute job

Search by title, datatoken, or DID...

**CIDAI pig farm Mask R-CNN segmentation & Tracking (CUDA) (UDL)**

GXAT | did:op:5375665bed1605ab035e0fcacf3a440e740...

Free

You will pay **0.6 OCEAN**

**ORDER COMPUTE JOB**

You bought this dataset already allowing you to use it without paying again. You already bought the selected algorithm, allowing you to use it without paying again. In order to start the job you also need to pay the fees for renting the c2d resources. Please note that network gas fees still apply, even when using free assets.

I agree to the [Terms and Conditions](#)

Access allowed

Your Compute Jobs		
STATUS	ACTIONS	FINISHED
JOB FINISHED	<a href="#">SHOW DETAILS</a>	about 2 months ago
DATA PROVISIONING ...	<a href="#">SHOW DETAILS</a>	about 2 months ago
JOB FINISHED	<a href="#">SHOW DETAILS</a>	about 2 months ago

- Cálculo de las métricas de bienestar animal sin revelar las imágenes de la granja
  - Imágenes y algoritmo se cargan en la habitación de los datos
  - Sólo se pueden descargar las métricas
- Monetización:
  - datos, algoritmo y computación

**Job finished**

Pig pen images sequence for animal well-being assessment

**Results**

- RESULTS (output.zip) - 328.1 MB
- ALGORITHM LOGS - 86.3 KB
- CONFIGURATION LOGS - 2.4 KB
- PUBLISH LOGS - 132 Bytes

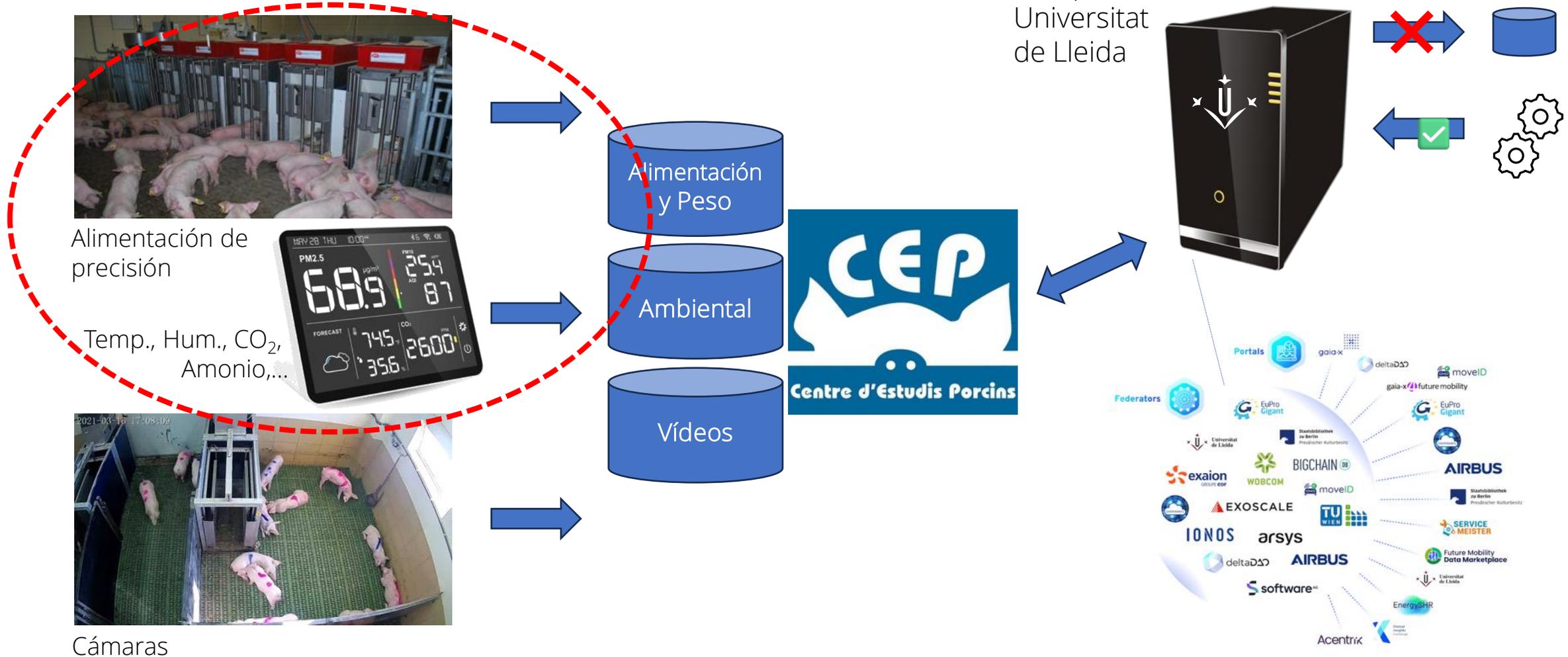
Results are stored for 30 days.

CREATED: about 2 months ago      FINISHED: about 2 months ago

JOB ID: 3ab788b14054e2988a98c093e44e47

# Caso de uso AI4Pork

Proveedor de Computación-a-Datos  
Universitat de Lleida



Alimentación de precisión

Temp., Hum., CO<sub>2</sub>, Amonio,...

Cámaras



# AI4Pork: Publicar Datos

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

## CEP - Automatic Pig Feeding - 2021 S1 - 982091062894196

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.

Select an algorithm to start a compute job

Search by title, datatoken, or DID...

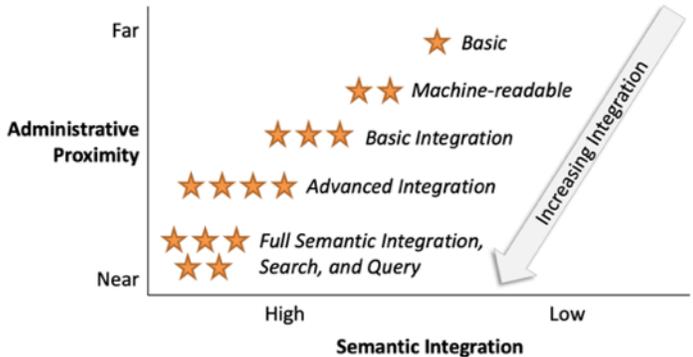
- Exploratory Data Analysis Free
- RML Mapper for Daily Pig Weight CSV Data Free

BUY COMPUTE JOB

Your Compute Jobs SHOW

2 sales

# Integración semántica de datos

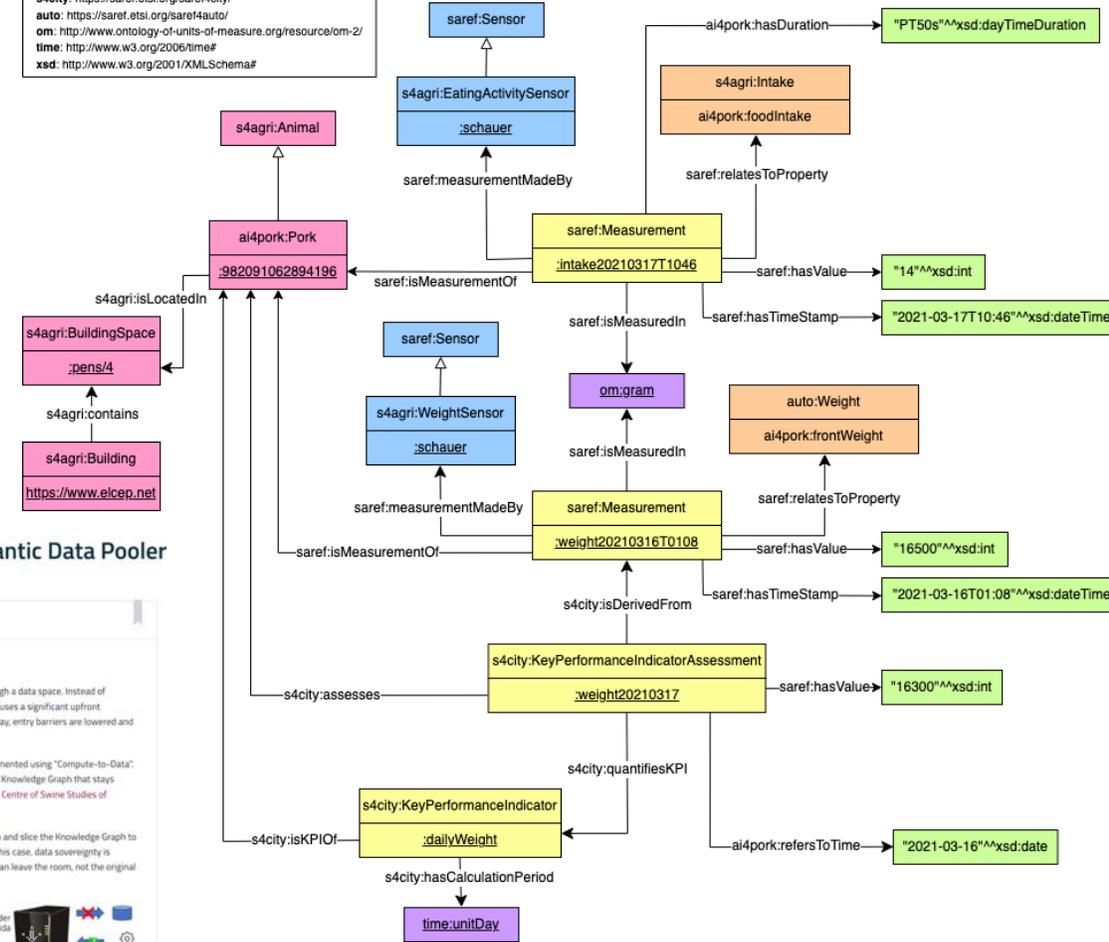


- Modelo "pay-as-you-go"
  - Demorar el esfuerzo de integración hasta que sea necesario
  - Reducir barreras iniciales para compartir datos

- Reutilización de vocabularios y ontologías existentes
- Algoritmos y reglas de mapeo forman parte del espacio de datos y sus incentivos
- Integración como oportunidad para crear "cooperativas de datos"

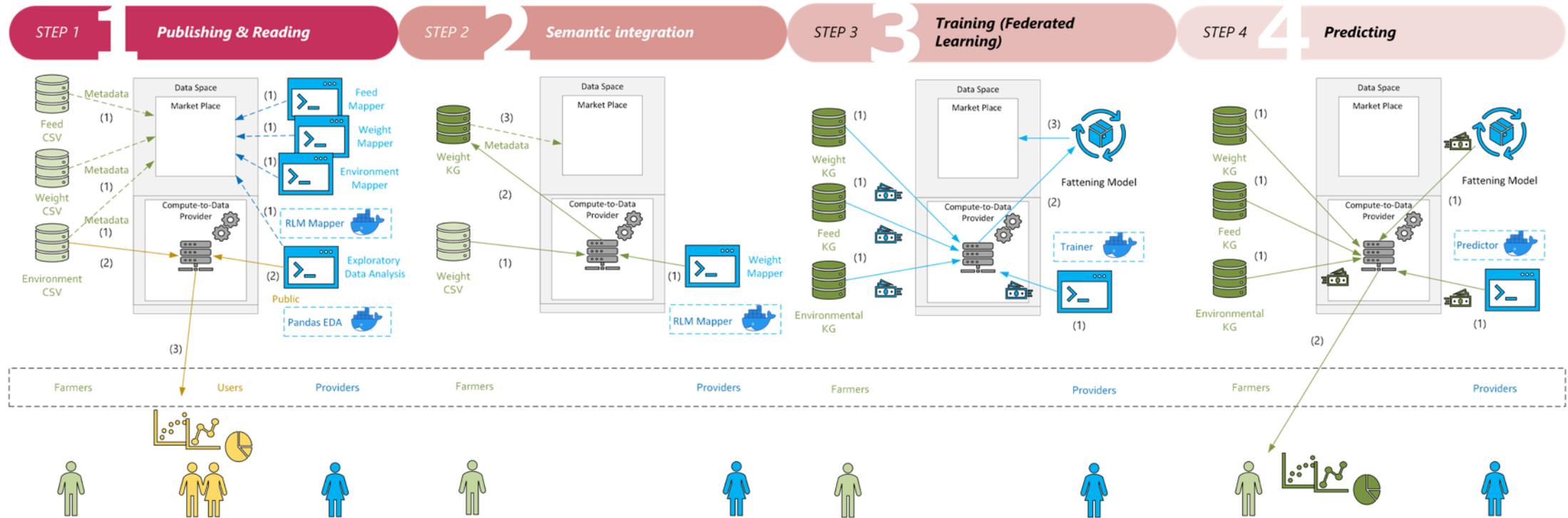
```

base: https://ai4pork.angliru.udl.cat/schauer/
ai4pork: https://ai4pork.angliru.udl.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#
    
```



## CEP's CSV Data Mapper and Semantic Data Pooler

# AI4Pork: integración y explotación de datos





# ¿Gracias por su atención, preguntas?

Roberto García ([roberto.garcia@udl.cat](mailto:roberto.garcia@udl.cat))

Universitat de Lleida

Con el apoyo de:

<https://agridataspace-csa.eu>

<https://agrifoodtef.eu>

<https://ai4pork.udl.cat>

<http://angliru.udl.cat>