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

  ...

Data Room

Your data

Control

You
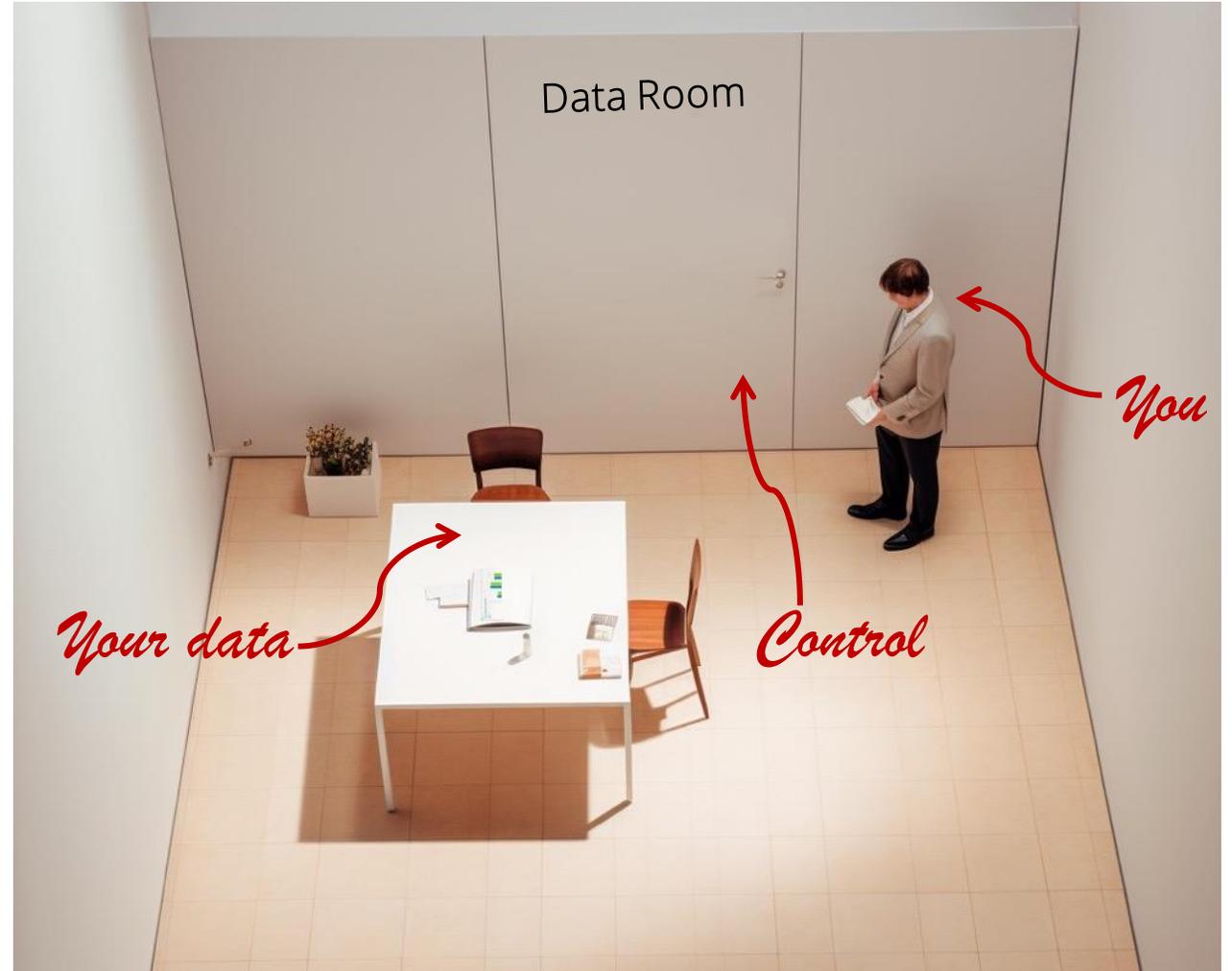
daro1995_a_bird-view_of_the_inside_of_a_whole_room_without_wind_1b2e5c91-5513-471d-9673-6c6493a76c11

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
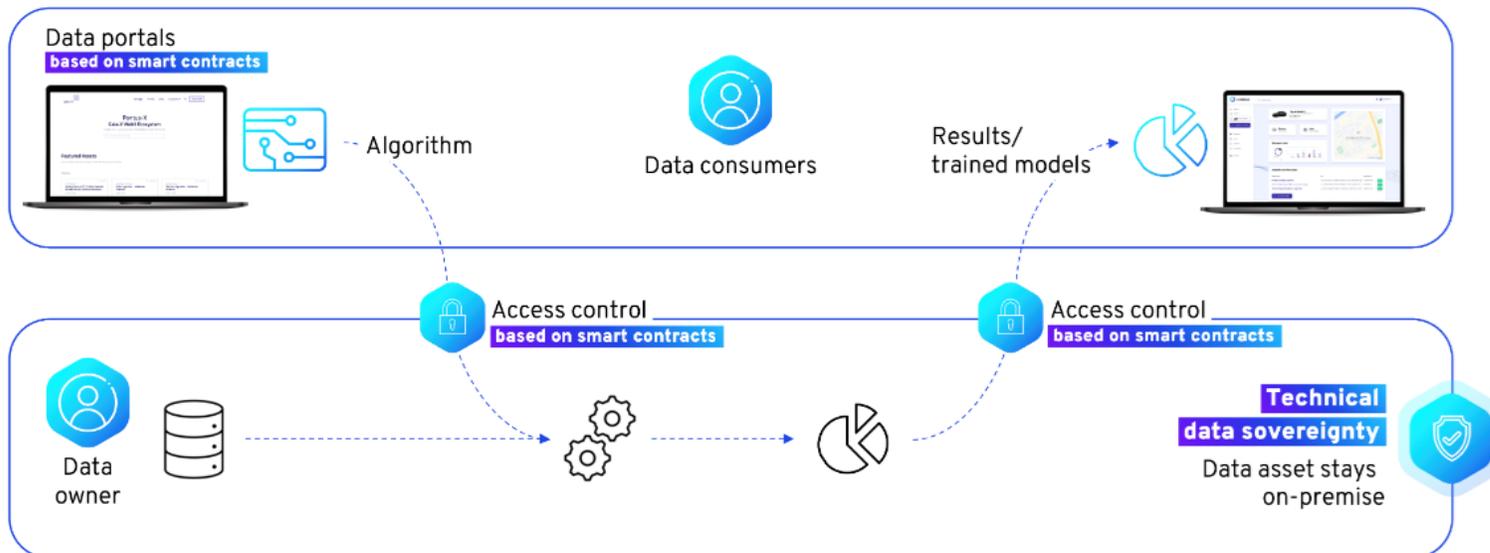
Ecosistema Pontus-X de Gaia-X

- 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

Alimentación de precisión

Temp., Hum., CO_2, Amonio,...

Cámaras

Alimentación y Peso

Ambiental

Vídeos

Proveedor de Computación-a-Datos Universitat de Lleida
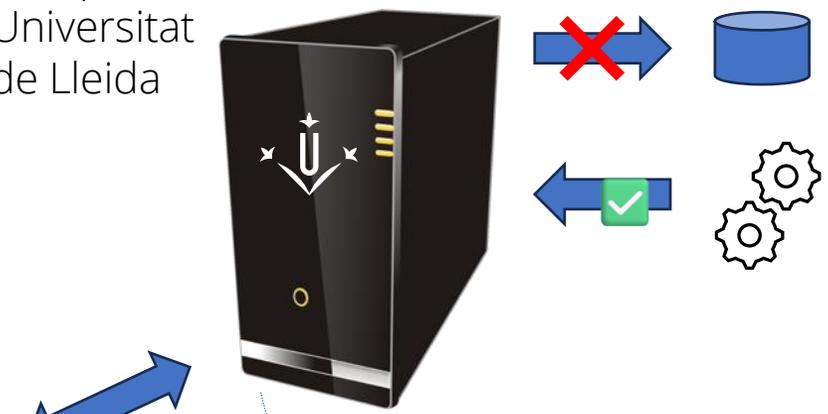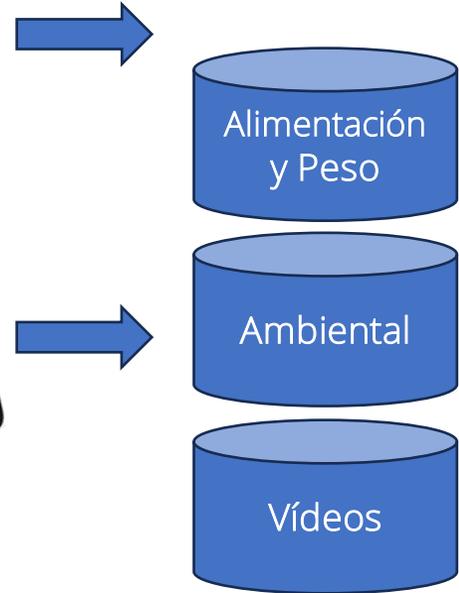
Portal del Espacio de Datos

Caso de uso AI4Pork

Alimentación de precisión

Temp., Hum., CO_2, Amonio,...

Cámaras

Alimentación y Peso

Ambiental

Vídeos

Proveedor de Computación-a-Datos Universitat de Lleida

CATALOGUE   PUBLISH   VERIFY   LOG ↗   ONBOARDING ↗   ECOSYSTEM

GEN-X Testnet TEST   0x6207...6822

CIDAI

AI4 Pork

ANGLIRU

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
}
```

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 }
}
```

SAMPLE DATA
DOWNLOAD SAMPLE ↗

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	SHOW DETAILS	about 2 months ago
DATA PROVISIONING ...	SHOW DETAILS	about 2 months ago
JOB FINISHED	SHOW DETAILS	about 2 months ago

Job finished ✕

Pig pen images sequence for animal well-being assessment ↗
GXAT | did:op:1738490c806b4df0914c4b28572cc13a21872c8145ba624...

CIDAI pig farm Mask R-CNN segmentation & Tracking (CUDA) (UDL) ↗
GXAT | did:op:5375665bed1605ab035e0fcacf3a440e74092967891e

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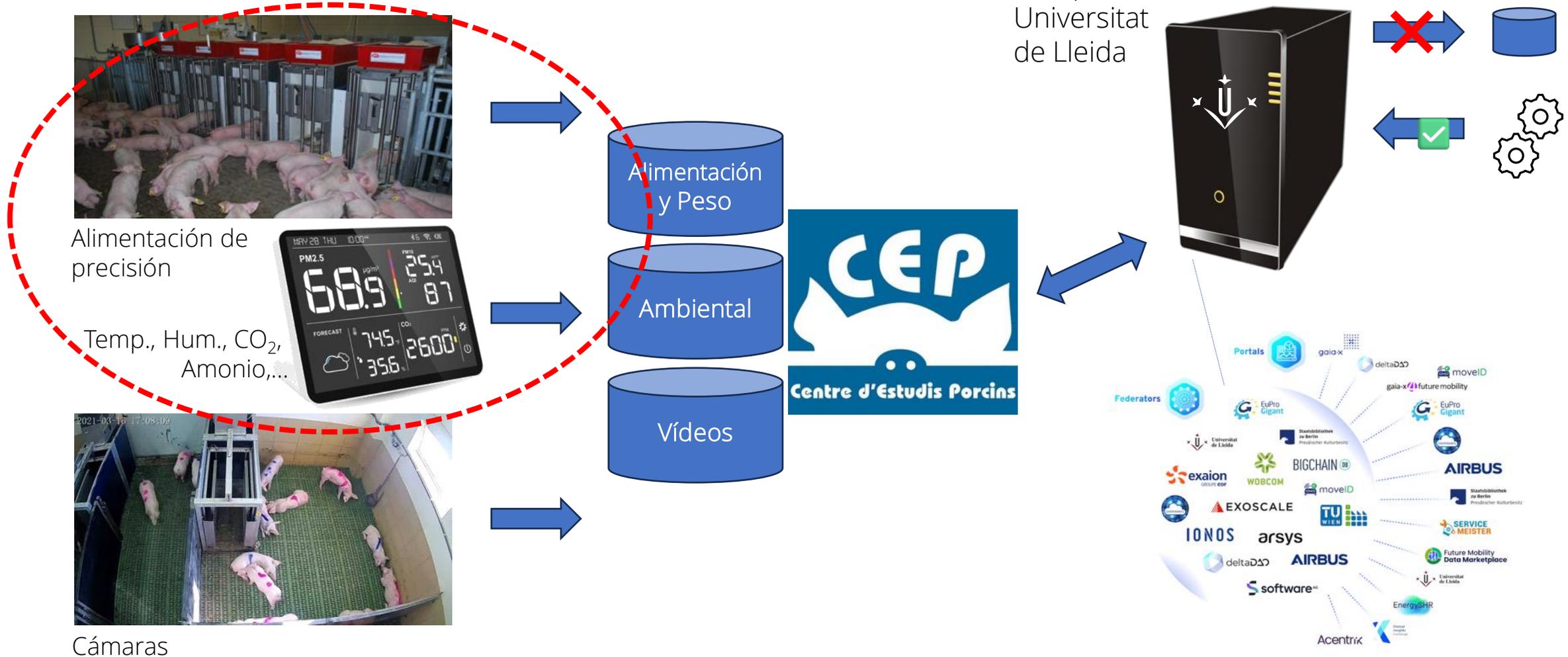- CONFIGURATION LOGS - 2.4 KB
- PUBLISH LOGS - 132 Bytes

Results are stored for 30 days.

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

JOB ID: 3ab78bb14d454e2988a098c893e64e47

- **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

10

Caso de uso AI4Pork

Alimentación de precisión

Temp., Hum., CO_2, Amonio,...

Cámaras

Alimentación y Peso

Ambiental

Vídeos

Proveedor de Computación-a-Datos Universitat de Lleida

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

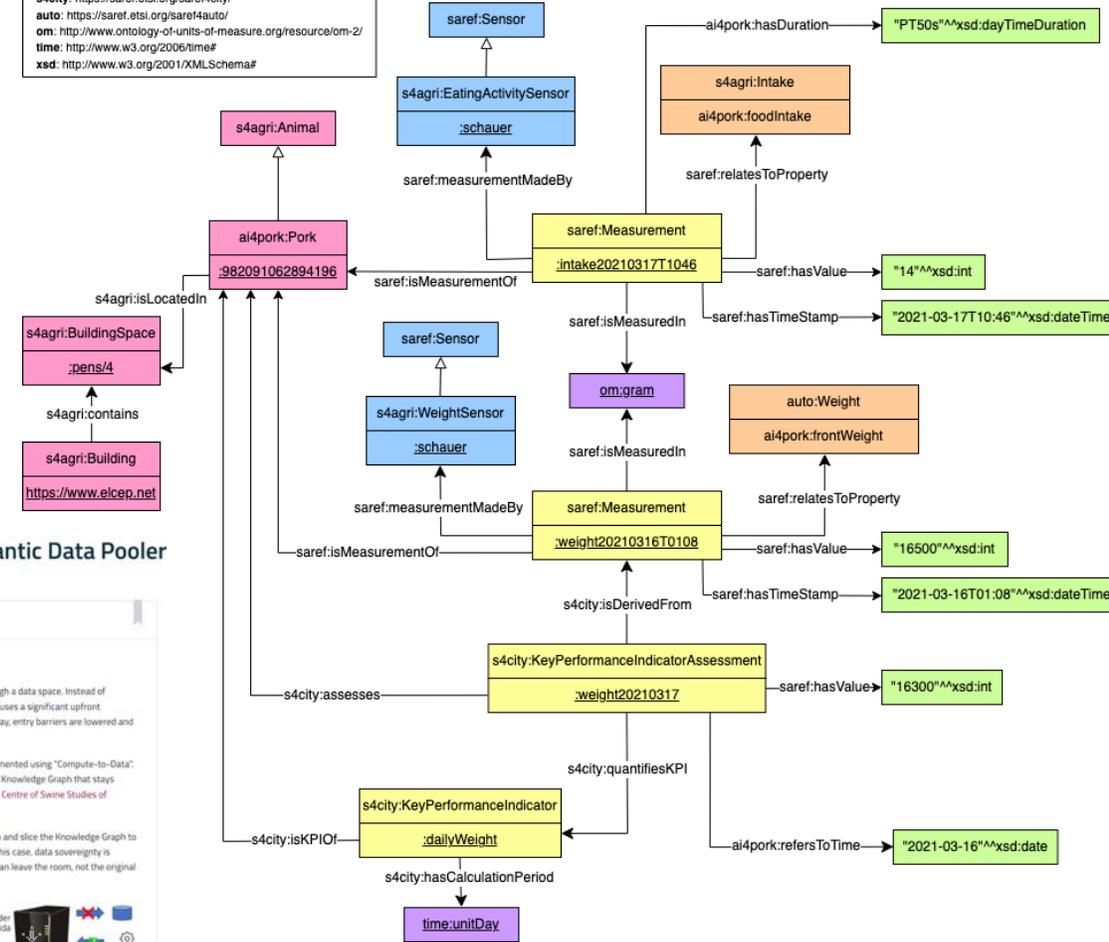# Integración semántica de datos

- 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**
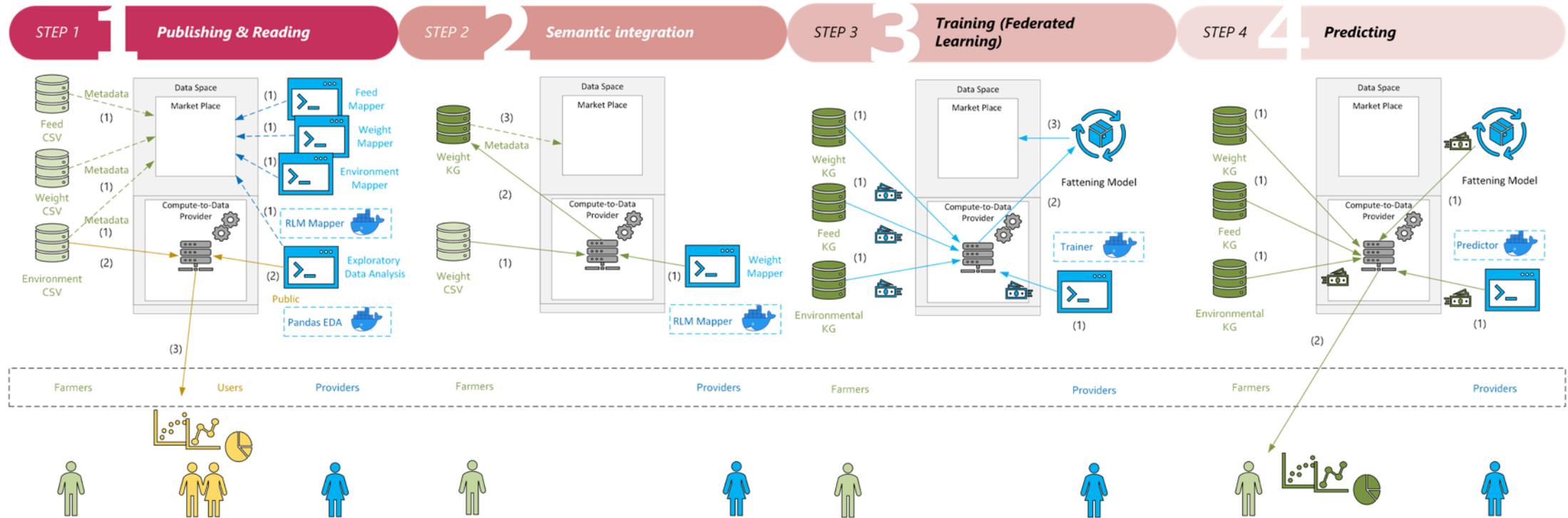- Integración como oportunidad para crear "**cooperativas de datos**"

AI4Pork: integración y explotación de datos

Integración con Gaia-X

¿Gracias por su atención, preguntas?

Roberto García (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