



# Espacios de datos: qué son y cómo generar valor con ellos

**Pablo Manuel García Corzo**

Expert Sales Data/AI T-Systems Iberia

Coordinador Grupos de Trabajo Dama España



**Diputació  
Barcelona**

Àrea de Bon Govern, Assistència  
Local i Cohesió Territorial

**#OpenGovWeek**



## Agenda

---

1. El valor del dato
2. El contexto europeo
3. ¿Qué son los espacios de datos?
4. Productos de datos
5. Sandboxes (Segittur)
6. Federated Learning
7. Casos de uso
8. El rol de las AAPP



## El valor de los datos

---



# El valor de los datos

---





Los espacios de datos son **ecosistemas** para el desarrollo de la **economía del dato**



## ¿A quién pertenecen los datos?

---



### **Derechos individuales y protección de la privacidad.**

- El individuo tiene un control significativo sobre sus datos.
- RGPD (Reglamento General de Protección de Datos).



### **Intereses empresariales y seguridad nacional.**

- Se prioriza la innovación y la seguridad, con un enfoque más laxo en la privacidad individual a nivel federal.
- Mosaico de leyes estatales (ej: CCPA en California) y leyes federales sectoriales.



### **Control estatal y seguridad nacional.**

- El Estado tiene una autoridad predominante sobre los datos, considerándolos un recurso estratégico nacional.
- LPIP (Ley de Protección de la Información Personal) y DSL (Ley de Seguridad de Datos).

## ¿Es más complicado ser competitivo en Europa?

---

El **Reglamento General de Protección de Datos (RGPD)** es uno de los marcos de privacidad de datos más estrictos del mundo.

Consentimiento Explícito y Derechos del Individuo  
Transferencias Transfronterizas de Datos

Mayor Confianza del Consumidor  
Innovación en Soluciones de Privacidad

### **Estandarización de Calidad:**

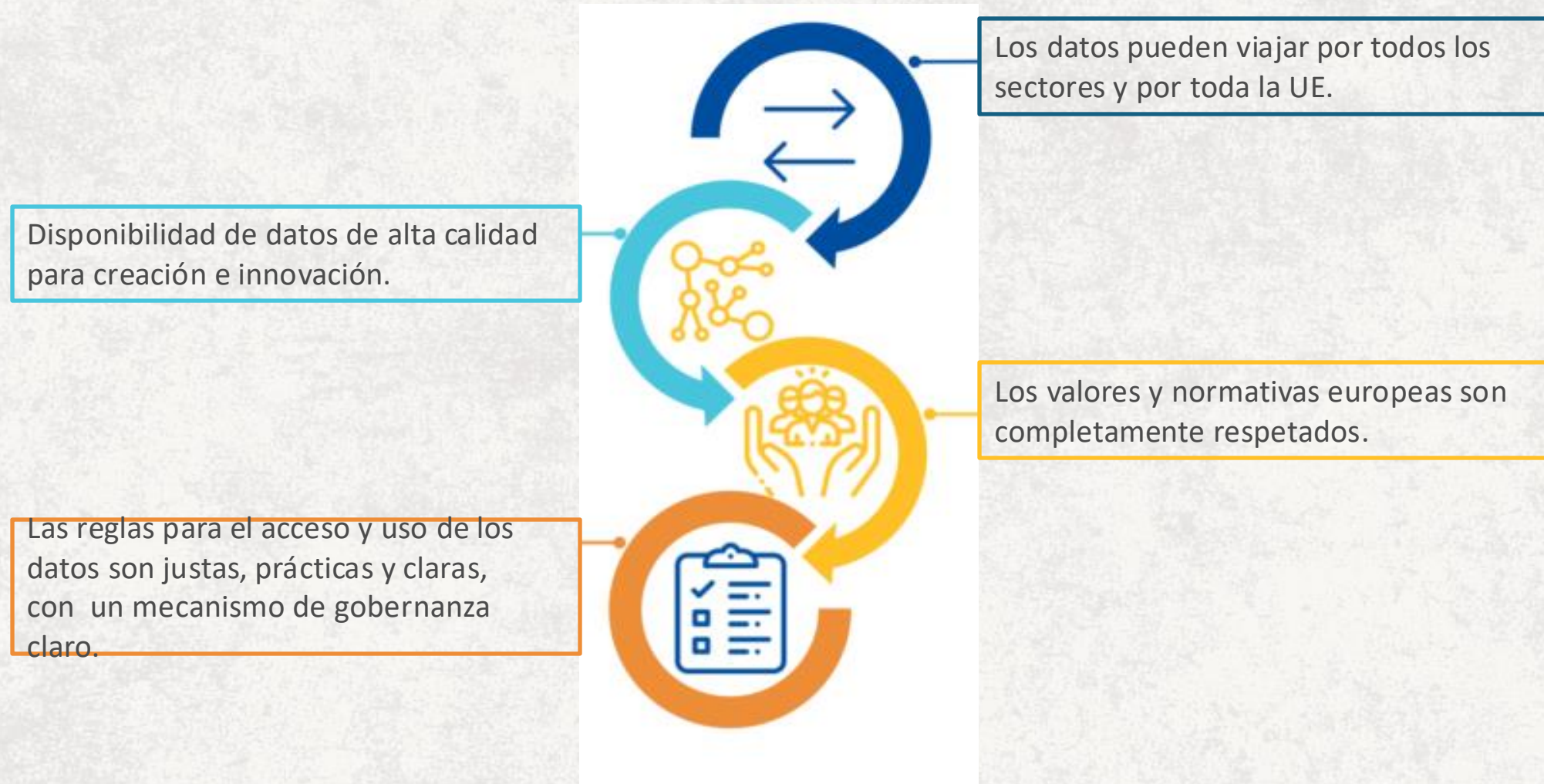
Las exigencias del RGPD a menudo llevan a una mejor gobernanza del dato, a una mayor calidad de los datos y a una gestión más transparente, lo que a la larga puede facilitar la generación de valor de forma más efectiva y sostenible.

El marco europeo impone **una barrera de entrada más alta** en términos de cumplimiento. Las empresas que logran gestionarlo de manera efectiva están **mejor posicionadas** para construir una relación de confianza con sus clientes y **desarrollar modelos de negocio** más resilientes y éticos **en el largo plazo**.

## ¿A quién pertenecen los datos?

---

Una serie de Data Spaces comunes con el objetivo de disponer de un mercado único de datos



## Espacio de datos

3 idiomas

Artículo [Discusión](#)

[Leer](#) [Editar](#) [Ver historial](#) [Herramientas](#)

Un espacio de datos<sup>1</sup> es un **ecosistema** donde materializar la **compartición voluntaria** de los datos de sus participantes dentro de un entorno de **soberanía, confianza y seguridad**, establecido mediante mecanismos integrados de gobernanza, organizativos, normativos y técnicos. El espacio de datos es pieza clave para el desarrollo de la **economía del dato**<sup>2</sup>, al habilitar su acceso, intercambio y reutilización legítima, posicionando al dato como **recurso no rival**, cuya utilidad crece según su uso se generaliza en un claro ejemplo de efecto red.

### INTERNATIONAL DATA SPACES ASSOCIATION



- Control **Soberano** de los Datos
- **Interoperabilidad** y estandarización
- Aspectos técnicos para **compartición segura**

#OpenGovWeek










Unión Europea

- Espacios de datos **sectoriales** para fomentar la innovación y el crecimiento económico.
- **Mercado único** de datos, garantizando la **libre circulación de datos** dentro de la UE
- Equilibrar la compartición de datos con la **protección de los derechos y libertades** de los individuos.



- Infraestructura Europea de Datos **soberana y federada**
- Transparencia, **interoperabilidad** y portabilidad
- Espacios de datos **sectoriales** y transfronterizos

## REQUISITOS TÉCNICOS DE LOS ESPACIOS DE DATOS

-  Solución basada en una **implementación europea de referencia** [Gaia-X, IDSA (**ODRL**), *Simpl*]
-  Plan de evolución futura de **adaptación al Trust Framework de Gaia-X**
-  Existencia de mecanismos para **garantizar la seguridad y confianza** (identificación, autenticación y autorización) basados en **identificadores descentralizados** (Self-Sovereign Identity – SSI)
-  Vigilancia de la **interoperabilidad** en el seno del espacio de datos
-  El Sistema debe ser **portable a diferentes entornos**
-  Garantía de la **seguridad y privacidad del espacio de datos**
-  Existencia de mecanismos de aseguramiento de la **transparencia de los datos** (trazabilidad y linaje de los datos)

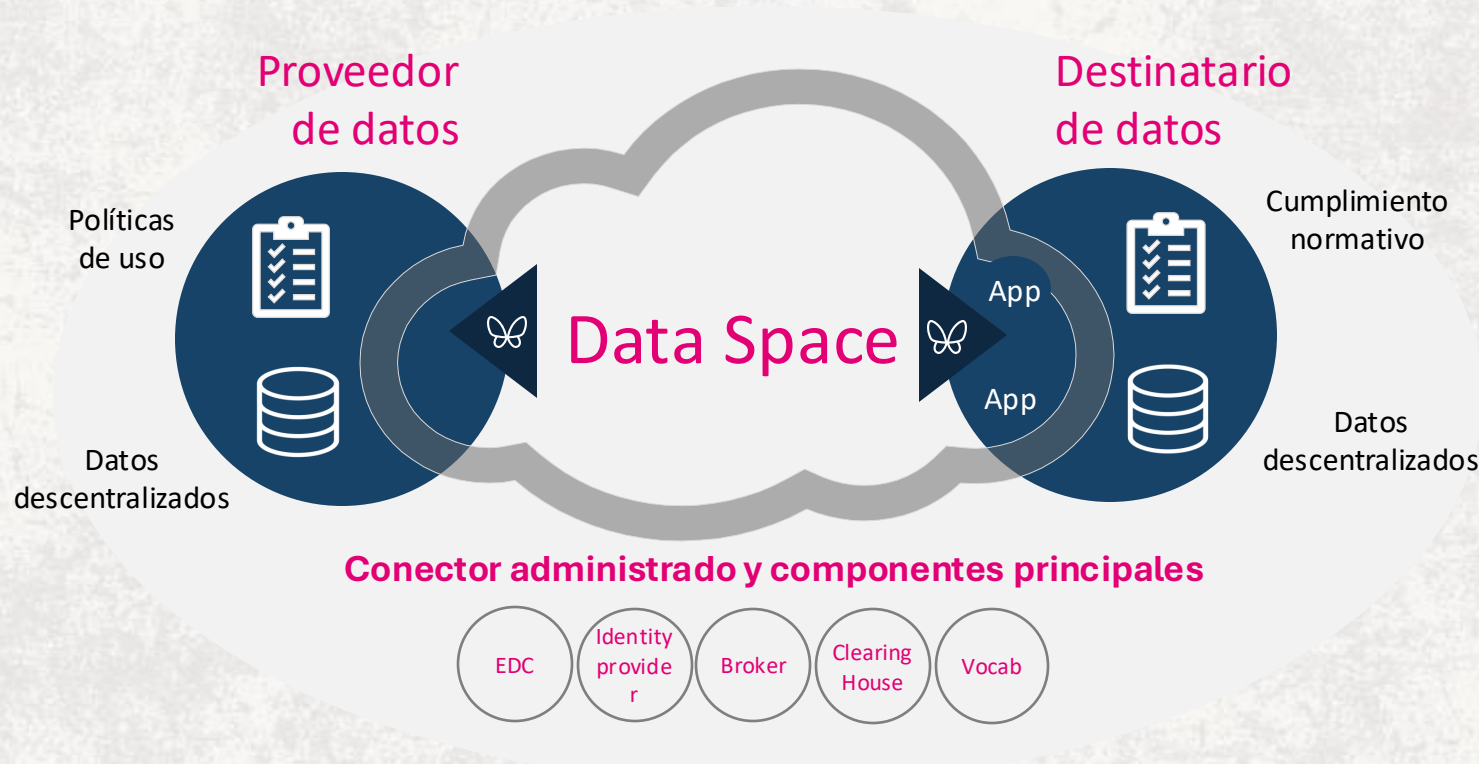


# ¿Qué son los espacios de datos?

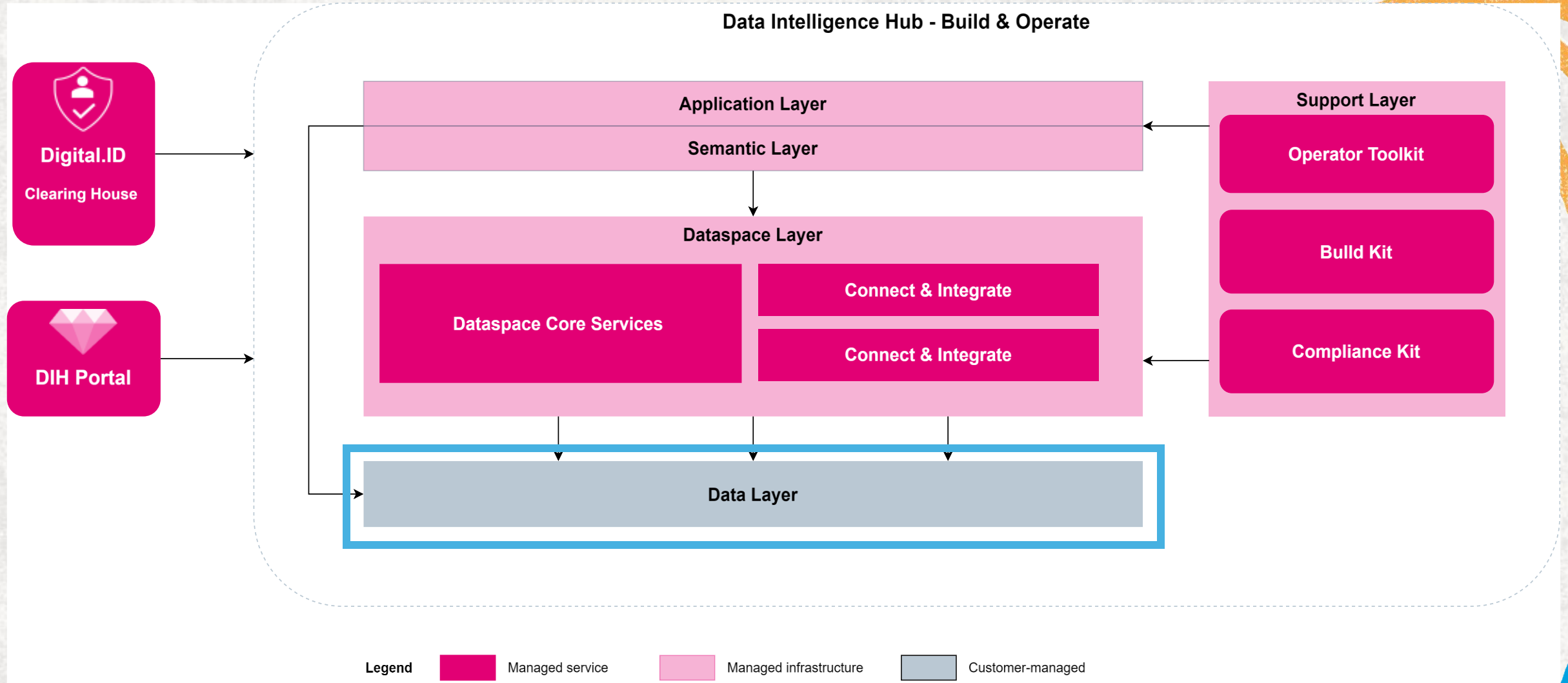


## PIEZAS FUNDAMENTALES

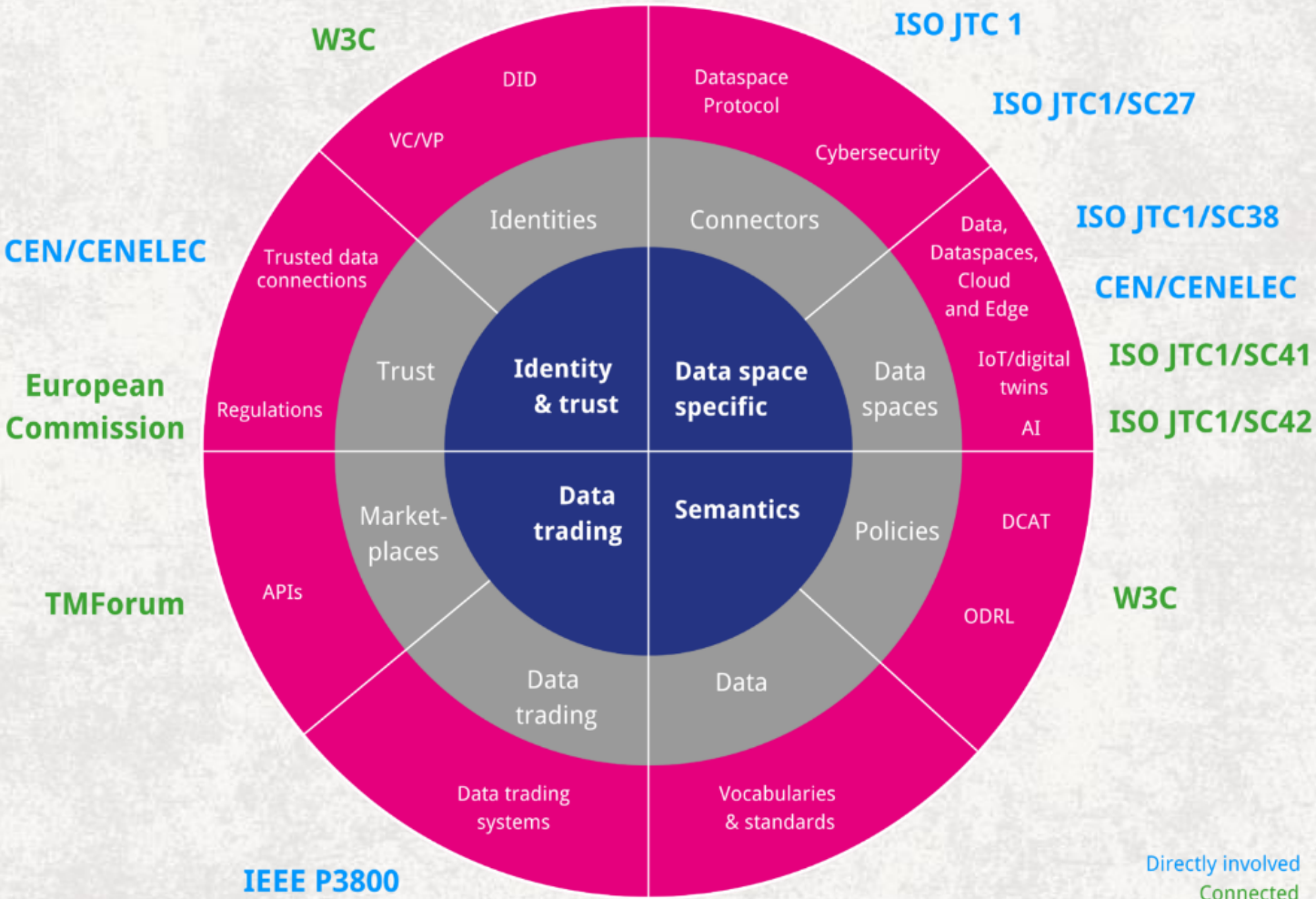
- EDC** Conector técnico de metadatos
- Identity provider** Garantiza identidades de los participantes
- Broker** Gestiona los mensajes entre participantes
- Clearing House** Confianza en el intercambio de datos
- Vocab** Lenguaje interoperable



# ¿Qué son los espacios de datos?



# ¿Qué son los espacios de datos?



La estandarización es crucial para el funcionamiento eficaz de los espacios de datos, ya que permite la **interoperabilidad** y el **intercambio seguro** de información entre diferentes participantes. Al establecer **protocolos y lenguajes comunes**, se facilita la comprensión y el uso compartido de los datos, lo que impulsa la innovación y la colaboración.

En este contexto, la International Data Spaces Association (**IDSA**) desempeña un papel fundamental al definir estándares como **ODRL** (lenguaje para la expresión de políticas de derechos y permisos), **IEEP3800** (modelo de identidad y confianza) y **DID** (identificadores descentralizados), que garantizan la seguridad, la privacidad y la soberanía de los datos en los ecosistemas digitales.

## ¿Qué son los espacios de datos?



### ODRL Information Model 2.2

W3C Recommendation 15 February 2018

Las reglas o políticas ODRL se usan para decir qué se puede y qué no se puede hacer con un producto de datos.

Por ejemplo, se puede permitir compartir un dataset o un modelo, pero prohibir modificarlo. También se pueden establecer **obligaciones**, como la de citar la fuente original.

Además, estas reglas pueden tener **límites**, como restricciones de **tiempo** (solo se permite usar el contenido durante un mes), de **lugar** (solo se permite acceder desde España) o de **propósito** (solo se permiten fines no comerciales).

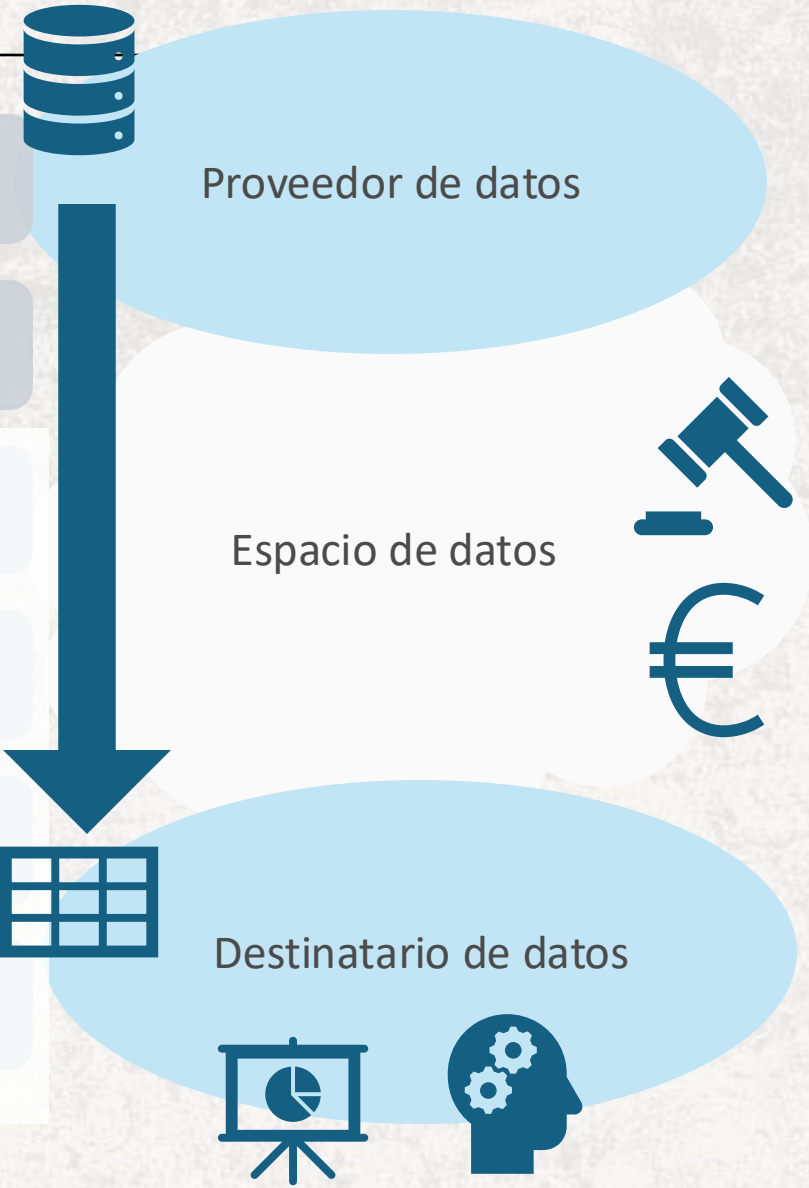
Finalmente, se pueden añadir **condiciones** a los permisos, como la obligación de pagar una cantidad determinada.

```
<odrl:Policy xmlns:odrl="http://www.w3.org/ns/odrl/2/">
  <odrl:Rule>
    <odrl:Prohibition>
      <odrl:target>mobility-dataset</odrl:target>
      <odrl:action odrl:type="http://www.w3.org/ns/odrl/2/function/use"/>
      <odrl:constraint>
        <odrl:LogicalConstraint odrl:operator="AND">
          <odrl:Scope>
            <odrl:SpatialConstraint odrl:type="http://www.w3.org/ns/odrl/2/spati
          </odrl:Scope>
          <odrl:Duty>
            <odrl:target>payment</odrl:target>
            <odrl:action odrl:type="http://www.w3.org/ns/odrl/2/function/pay"/>
            <odrl:amount>100</odrl:amount>
            <odrl:unit>EUR</odrl:unit>
            <odrl:periodicity>monthly</odrl:periodicity>
          </odrl:Duty>
        </odrl:LogicalConstraint>
      </odrl:constraint>
    </odrl:Prohibition>
  </odrl:Rule>
</odrl:Policy>
```

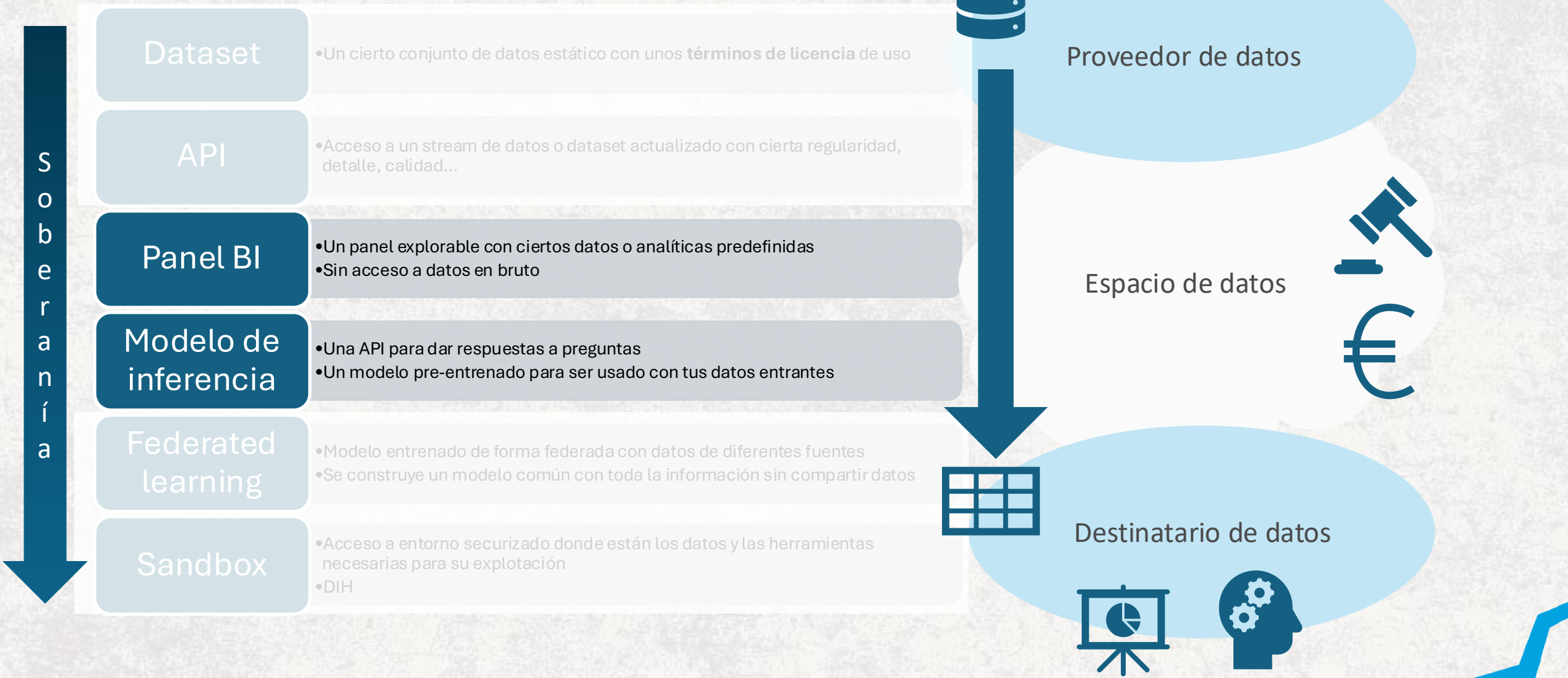
# Productos de datos

S  
o  
b  
e  
r  
a  
n  
í  
a

- Dataset**
  - Un cierto conjunto de datos estático con unos **términos de licencia** de uso
- API**
  - Acceso a un stream de datos o dataset actualizado con cierta regularidad, detalle, calidad...
- Panel BI**
  - Un panel explorable con ciertos datos o analíticas predefinidas
  - Sin acceso a datos en bruto
- Modelo de inferencia**
  - Una API para dar respuestas a preguntas
  - Un modelo pre-entrenado para ser usado con tus datos entrantes
- Federated learning**
  - Modelo entrenado de forma federada con datos de diferentes fuentes
  - Se construye un modelo común con toda la información sin compartir datos
- Sandbox**
  - Acceso a entorno securizado donde están los datos y las herramientas necesarias para su explotación
  - DIH



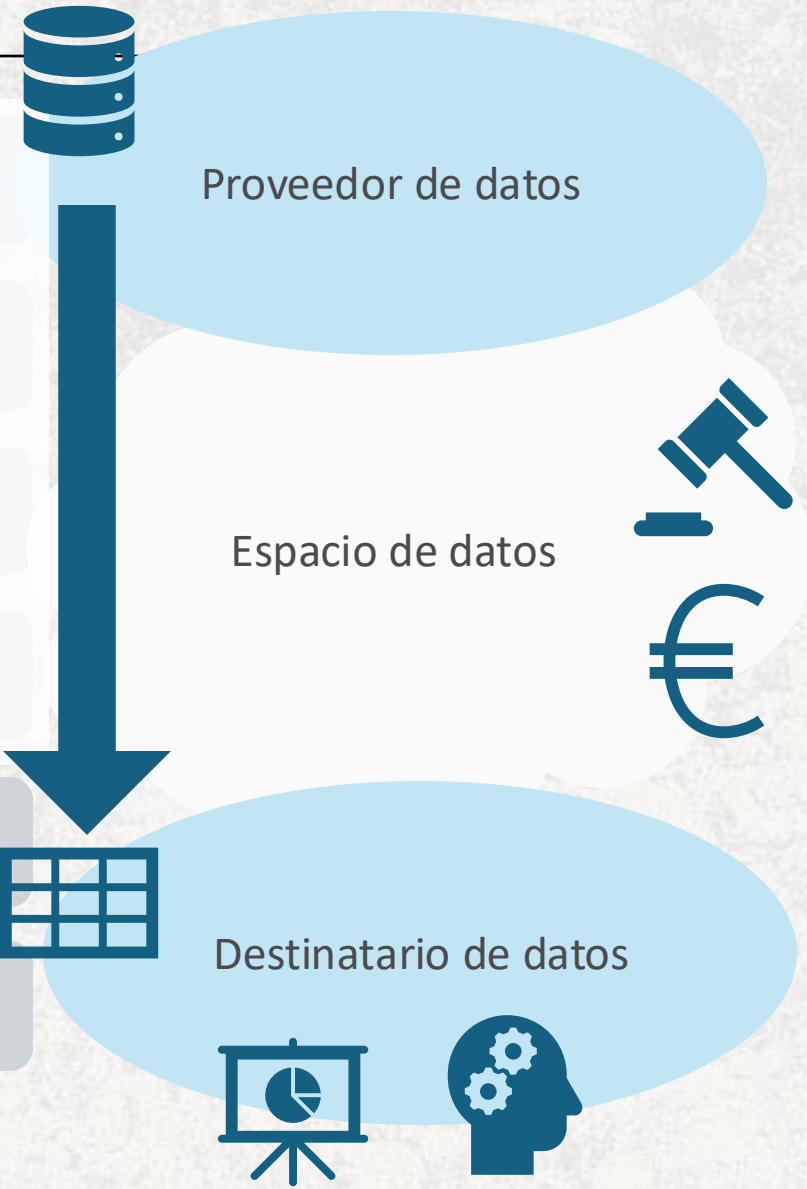
# Productos de datos



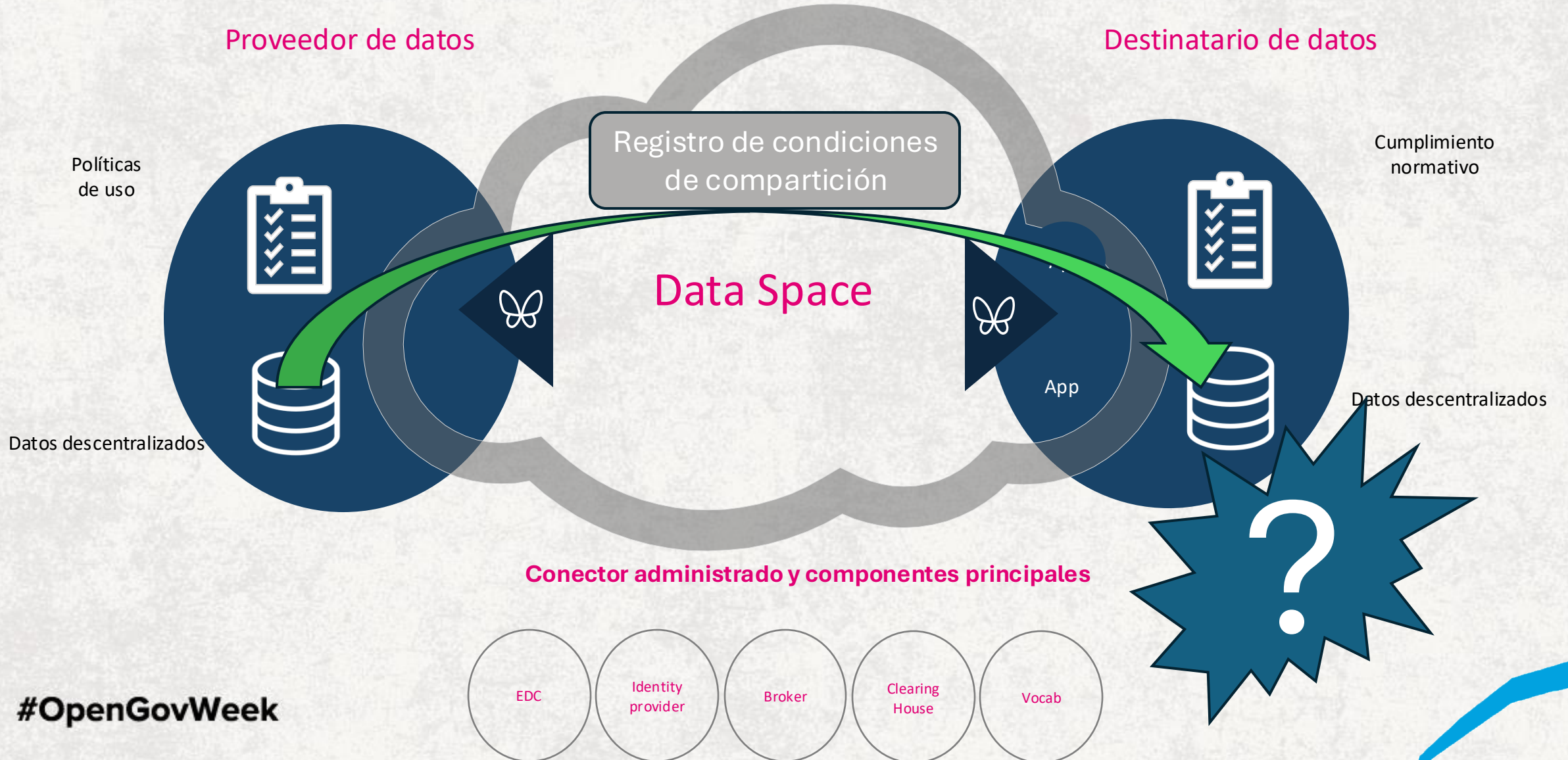
# Productos de datos

S  
o  
b  
e  
r  
a  
n  
í  
a

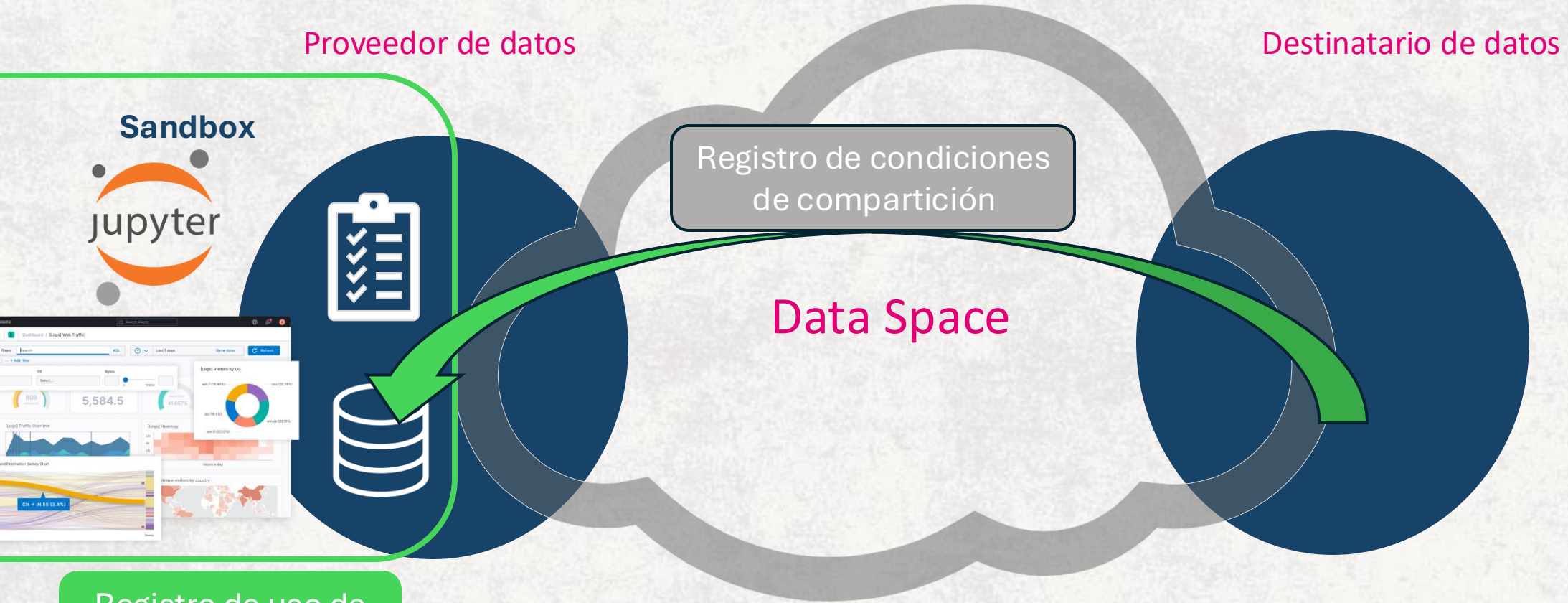
- Dataset**
  - Un cierto conjunto de datos estático con unos **términos de licencia** de uso
- API**
  - Acceso a un stream de datos o dataset actualizado con cierta regularidad, detalle, calidad...
- Panel BI**
  - Un panel explorable con ciertos datos o analíticas predefinidas
  - Sin acceso a datos en bruto
- Modelo de inferencia**
  - Una API para dar respuestas a preguntas
  - Un modelo pre-entrenado para ser usado con tus datos entrantes
- Federated learning**
  - Modelo entrenado de forma federada con datos de diferentes fuentes
  - Se construye un modelo común con toda la información sin compartir datos
- Sandbox**
  - Acceso a entorno securizado donde están los datos y las herramientas necesarias para su explotación
  - DIH



# Sandboxes



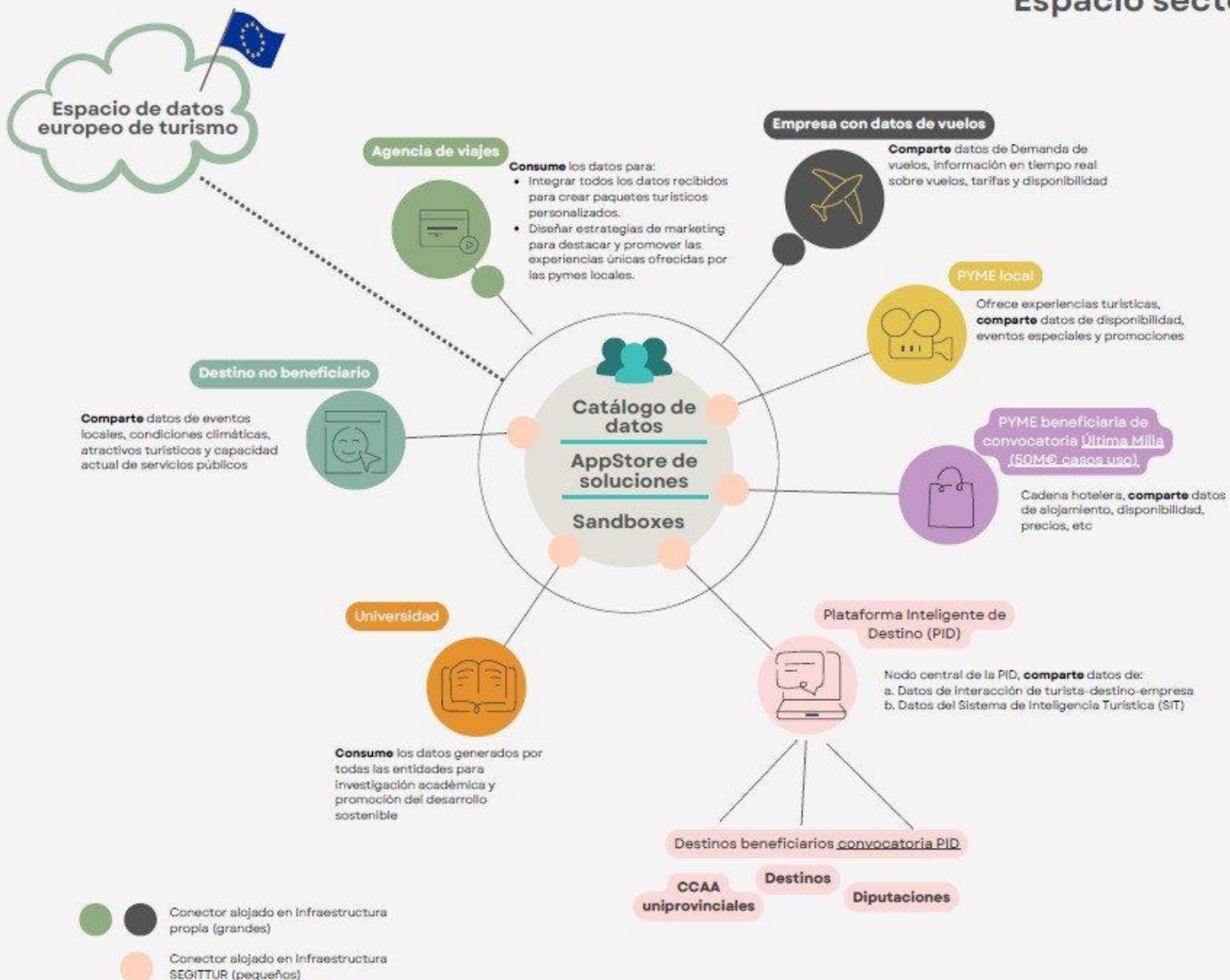
# Sandboxes



**Sandbox**  
jupyter

Registro de uso de los datos

- Conector administrado y componentes principales
- EDC
  - Identity provider
  - Broker
  - Clearing House
  - Vocab



## Resultado

- 1 Promoción Directa de Pymes y Cadena Hotelera:** Tanto las pymes locales como la cadena hotelera obtienen visibilidad y ventas a través de la agencia de viajes.
- 2 Diversificación de Experiencias:** Los visitantes pueden disfrutar de una gama más amplia de experiencias al combinar servicios de pymes y alojamientos de la cadena hotelera.
- 3 Colaboración Intersectorial:** La cadena hotelera contribuye al desarrollo local al trabajar en conjunto con pymes y el destino turístico.
- 4 Potenciación Económica Sostenible:** Se fomenta la economía local a través de la participación activa de actores turísticos diversos
- 5 Investigación Académica:** La universidad utiliza datos reales para investigar y proponer mejoras en la gestión turística y sostenibilidad..

# Federated Learning

---



Un hospital tiene diferentes fuentes de datos (de diversas áreas) que debe consolidar para poder construir un modelo sobre sus pacientes.

- Gobernanza
- Calidad del dato
- Seguridad
- Control

# Federated Learning

---

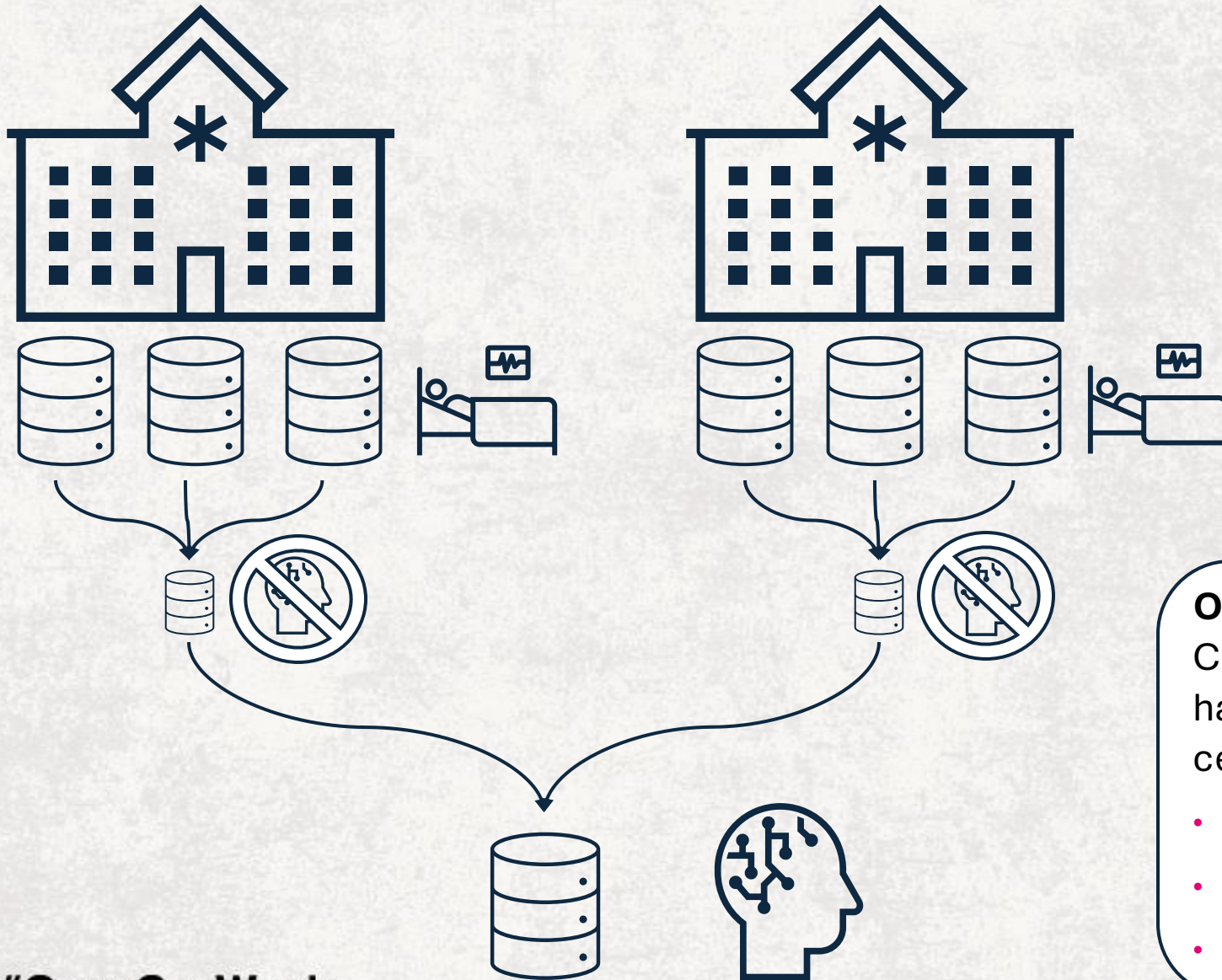


Un segundo hospital tiene diferentes fuentes de datos (de diversas áreas) que debe consolidar para poder construir un modelo diferente sobre un conjunto distinto de pacientes.

¿Cómo construir un modelo conjunto?



# Federated Learning



Un segundo hospital tiene diferentes fuentes de datos (de diversas áreas) que debe consolidar para poder construir un modelo común sobre un conjunto distinto de pacientes.

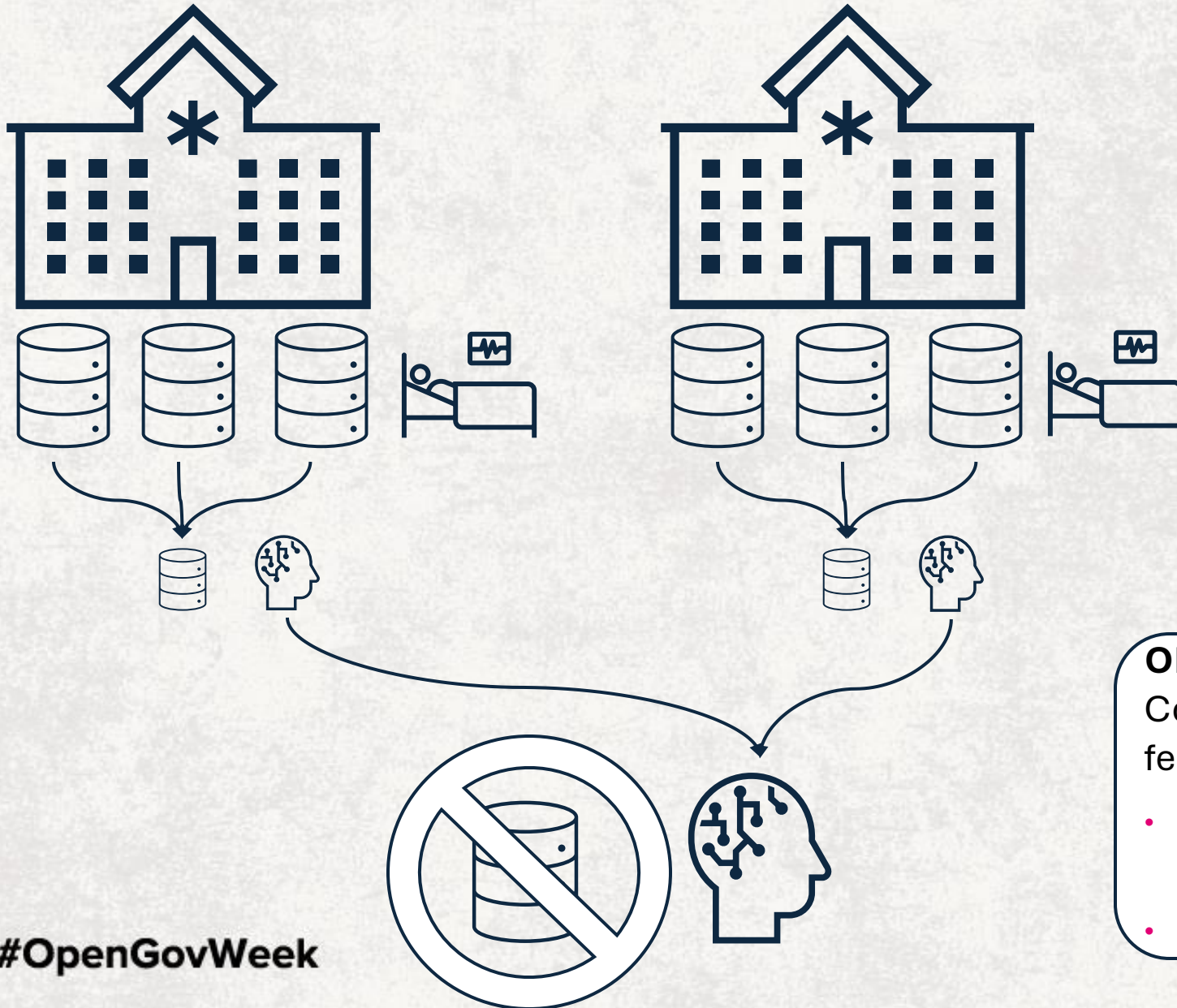
¿Cómo construir un modelo conjunto?

## OPCIÓN A:

Compartimos los datos hacia un repositorio centralizado

- Seguridad
- Compliance
- Almacenamiento

# Federated Learning



Un segundo hospital tiene diferentes fuentes de datos (de diversas áreas) que debe consolidar para poder construir un modelo común sobre un conjunto distinto de pacientes.

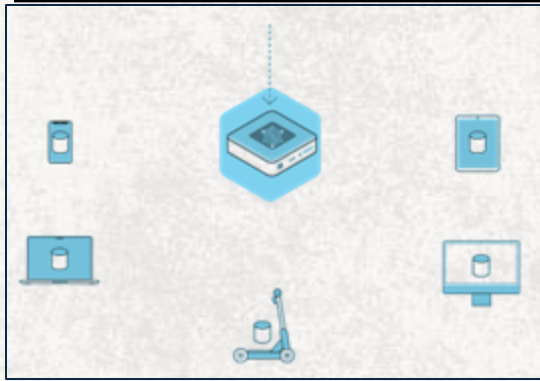
¿Cómo construir un modelo conjunto?

## OPCIÓN B:

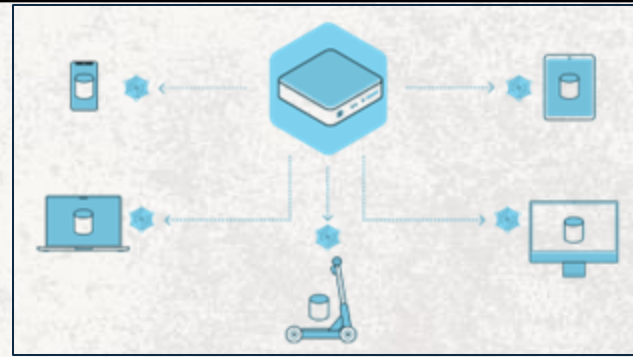
Construimos modelos federados

- Entrenamiento distribuido
- Modelos “caja negra”

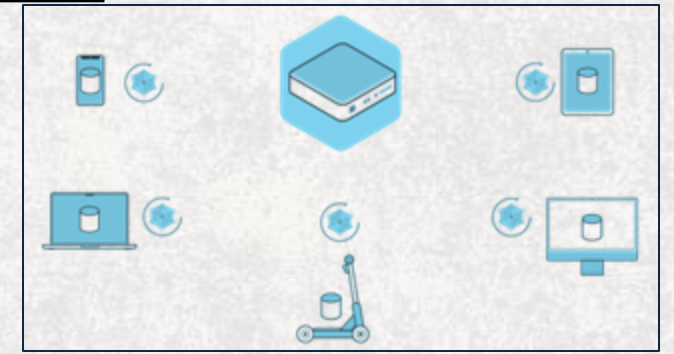
# Federated Learning



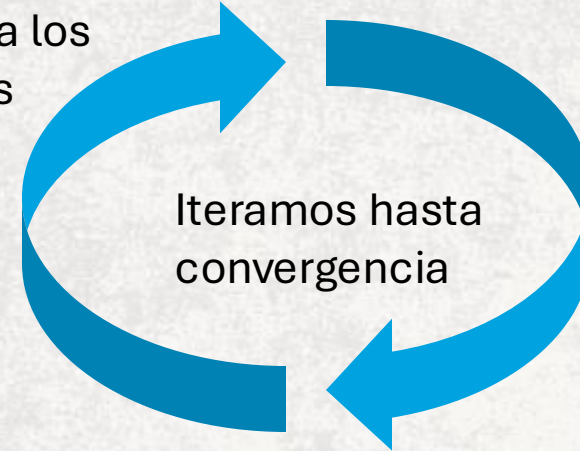
1- Inicialización del modelo inicial



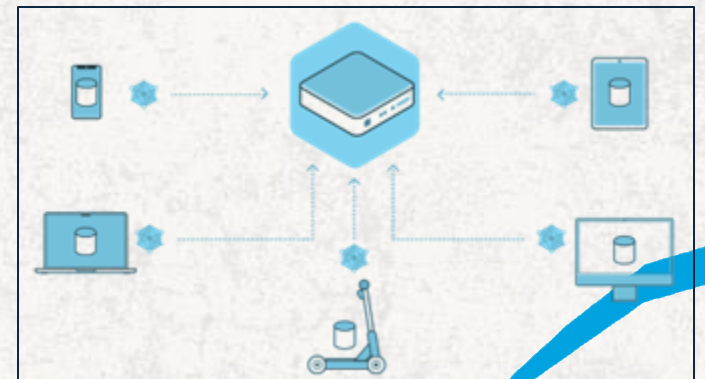
2- Llevo el modelo a los diferentes espacios



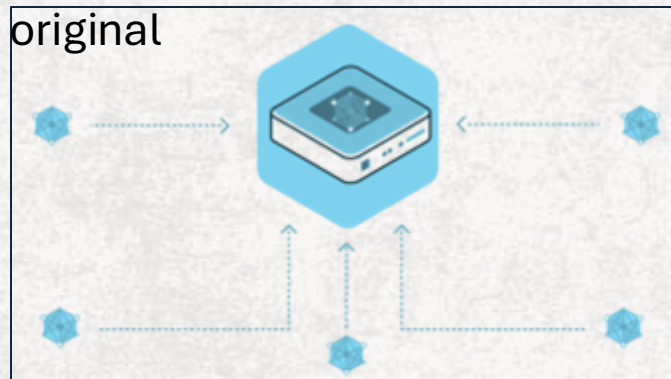
3- Entreno en los diferentes espacios



4- Traigo de vuelta los modelos



5- Consolido los modelos en el original

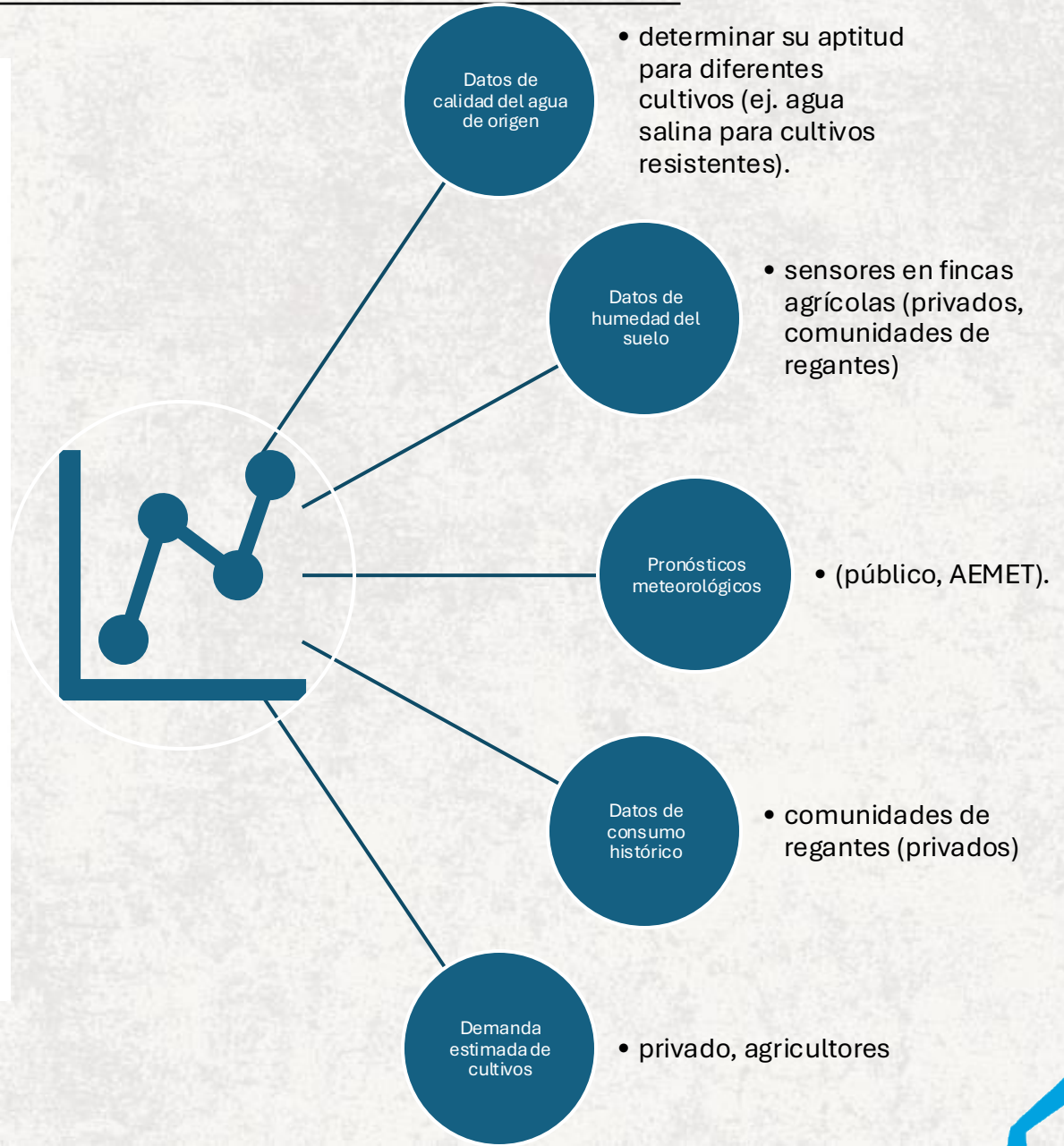


6- modelo listo para inferencia

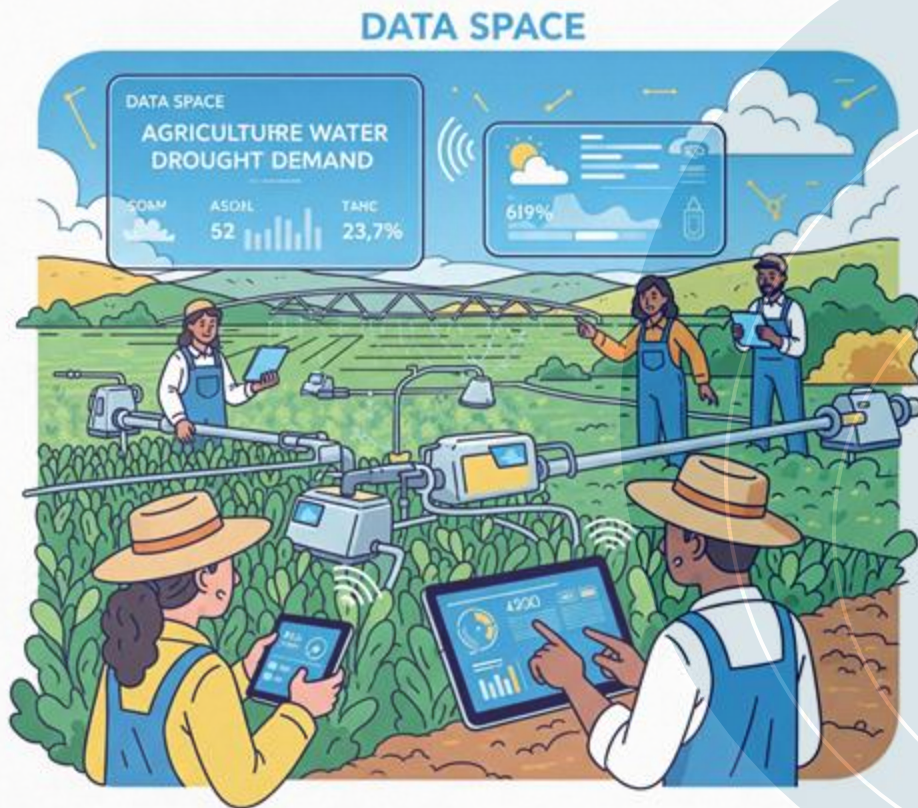


#OpenGovWeek

# Gestión de la Demanda Agrícola en Períodos de Sequía



# Gestión de la Demanda Agrícola en Períodos de Sequía



Datos de calidad del agua de origen

- determinar su aptitud para diferentes cultivos (ej. agua salina para cultivos resistentes).

Datos de humedad del suelo

- sensores en fincas agrícolas (privados, comunidades de regantes)

Pronósticos meteorológicos

- (público, AEMET).

Datos de consumo histórico

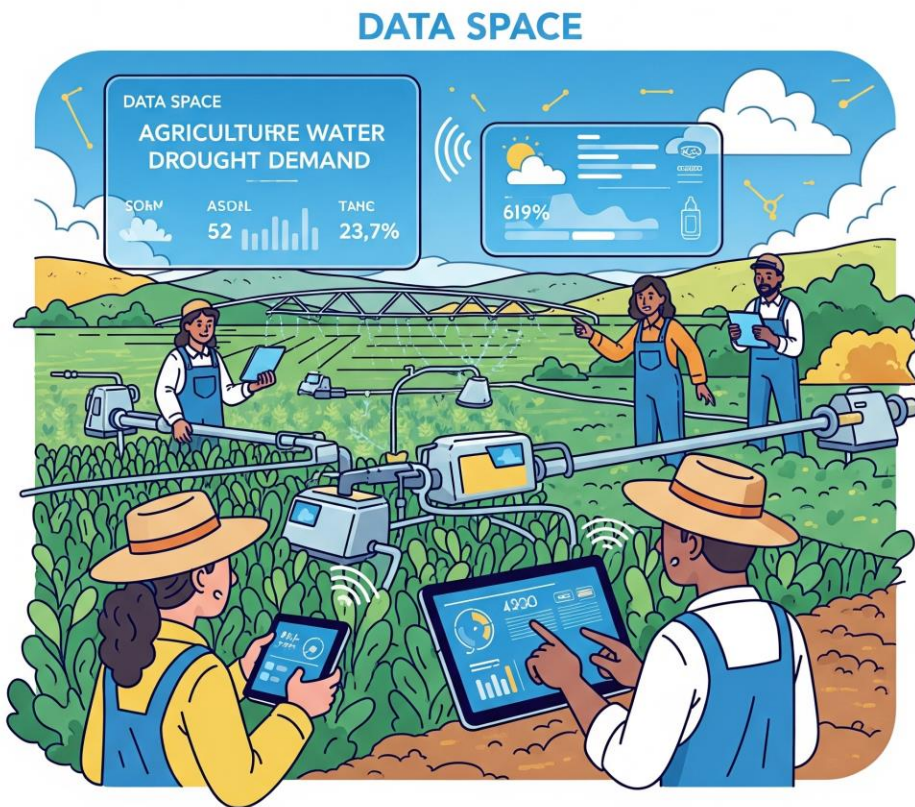
- comunidades de regantes (privados)

Demanda estimada de cultivos

- privado, agricultores

Actores Clave: Diputación de Barcelona, Agència Catalana de l'Aigua (ACA), Comunidades de Regantes, Empresas Agrícolas, Proveedores de Tecnología de Sensores Agrícolas

## Gestión de la Demanda Agrícola en Períodos de Sequía



La ACA o la Diputación podría tomar decisiones informadas sobre las restricciones de riego, priorizando cultivos de alto valor o áreas con mayor necesidad, minimizando pérdidas para los agricultores y gestionando eficazmente el recurso.

Los agricultores, a su vez, recibirían alertas y recomendaciones personalizadas sobre cuándo y cuánta agua aplicar, optimizando su uso y reduciendo costes.

Actores Clave:

Diputación de Barcelona, Agència Catalana de l'Aigua (ACA), Comunidades de Regantes, Empresas Agrícolas, Proveedores de Tecnología de Sensores Agrícolas.

# Gestión de la Demanda Agrícola en Períodos de Sequía

## El Papel del Agricultor

Su papel se transformaría de ser un mero receptor de restricciones a ser un **gestor proactivo y optimizador del riego**.

## Aportaría datos

### DATA SPACE

sensores en sus parcelas

tipo de cultivo

fase de crecimiento

tipo de suelo de su explotación

## Recomendaciones

disponibilidad global del recurso

humedad del suelo

pronósticos meteorológicos

necesidades específicas de su cultivo

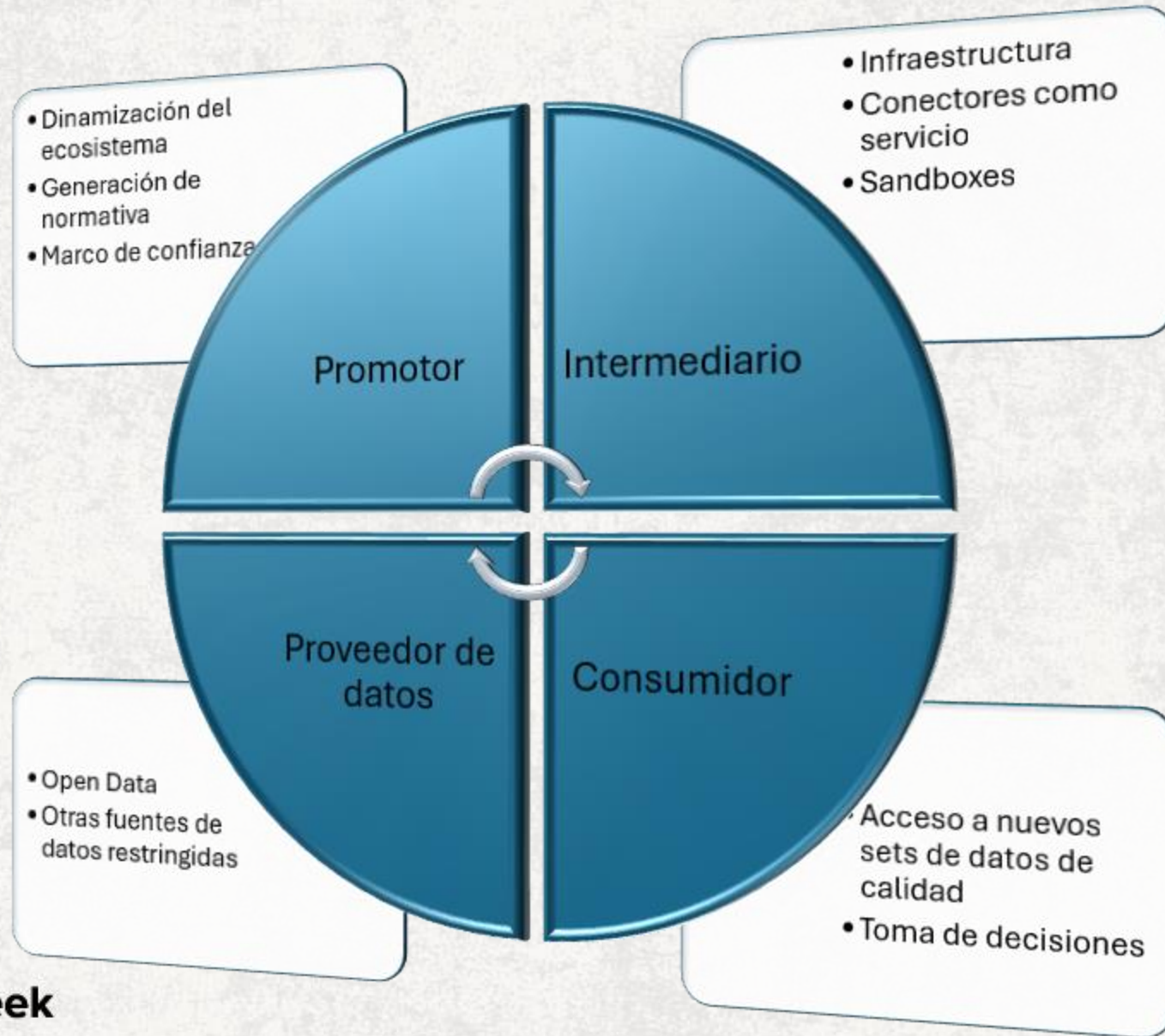
## Adaptaría prácticas

Ajustaría sus estrategias de riego según las alertas y recomendaciones

## Colaboraría

Compartir su experiencia y resultados con la Diputación o la Comunidad de Regantes, contribuyendo a mejorar la inteligencia colectiva del sistema.

# El rol de las Administraciones públicas





**Moltes gràcies**  
**Munches gracias**

**#OpenGovWeek**



**Diputació  
Barcelona**

**Àrea de Bon Govern, Assistència  
Local i Cohesió Territorial**



**OPENGOV  
WEEK**