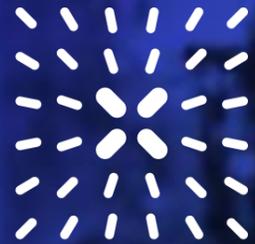


Gaia X España

Introducción a los espacios de datos

Joaquín Salvachúa (Universidad Politécnica de Madrid)

Joaquin.salvachua@upm.es @jsalvachua



gaia-x

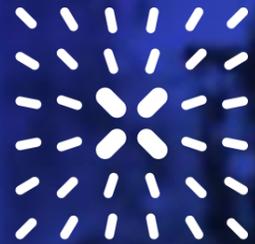


Hub España



UNIVERSIDAD
POLITÉCNICA
DE MADRID

¿Qué es un espacio de datos y como he podido vivir sin conocerlos ?



gaia-x



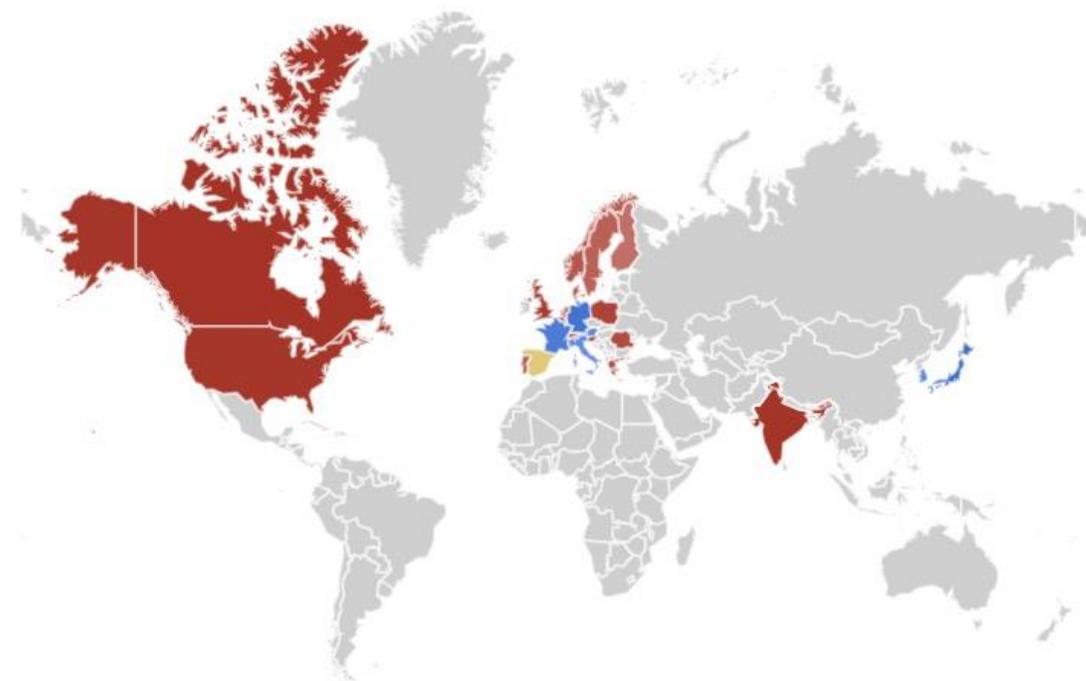
Hub España

«No hay nada más poderoso
que una idea a la que le ha
llegado su tiempo»
Víctor Hugo

// 01 Espacios de datos

Búsquedas en Google trends

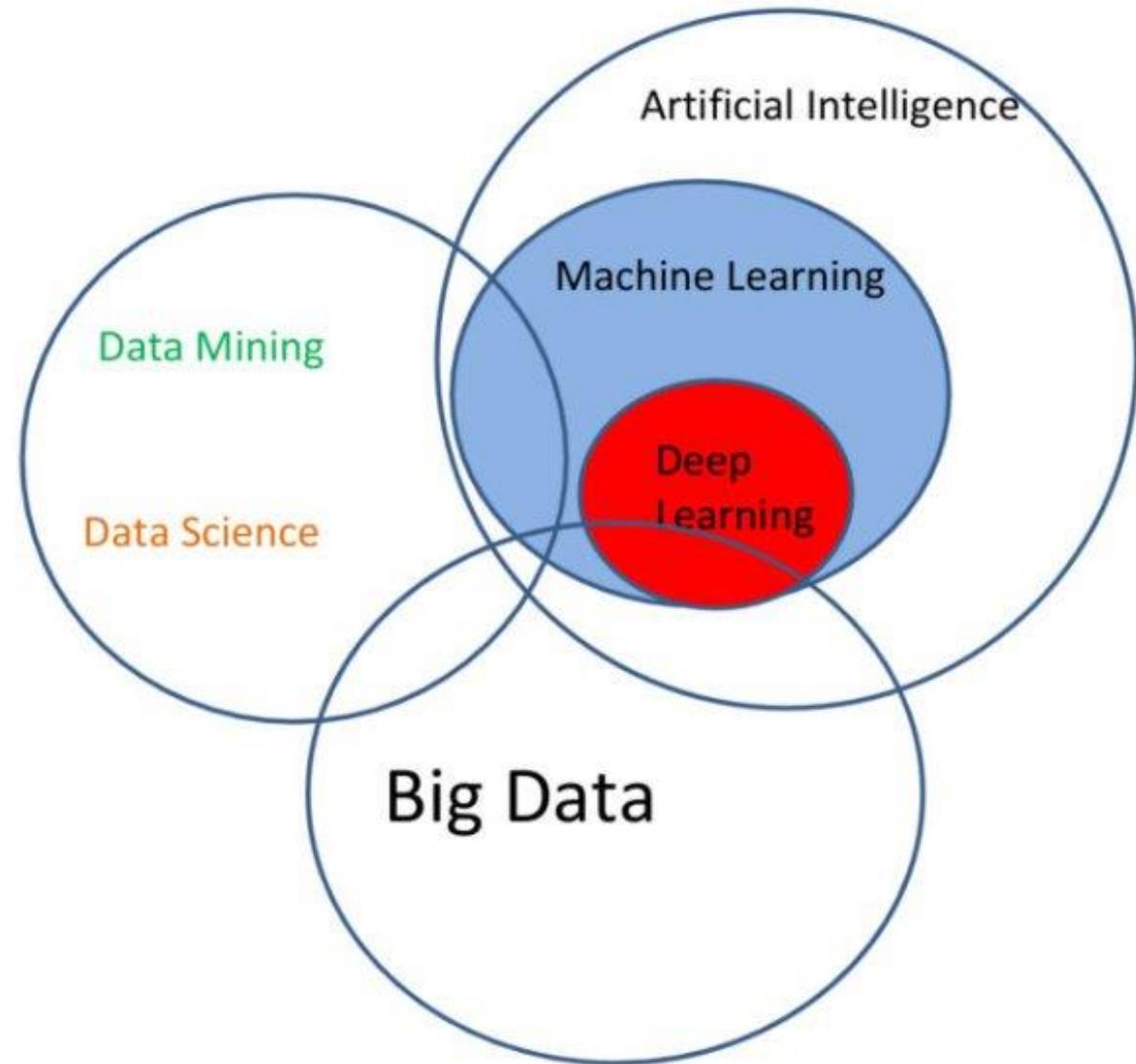
● gaia-x ● data spaces ● espacios de datos



¿Qué es un espacio de datos ?

- Definición Wikipedia :
 - Un espacio de datos 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, 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.
- Definición en Gaia-X :
 - Permite una infraestructura de datos federada y segura, en la que los datos se comparten, y los usuarios conservan el control sobre el acceso y el uso de sus datos.
 - Permite la creación de vínculos entre muchos proveedores de servicios en la nube en un ecosistema más amplio, transparente y justo para impulsar la economía europea de los datos del mañana.

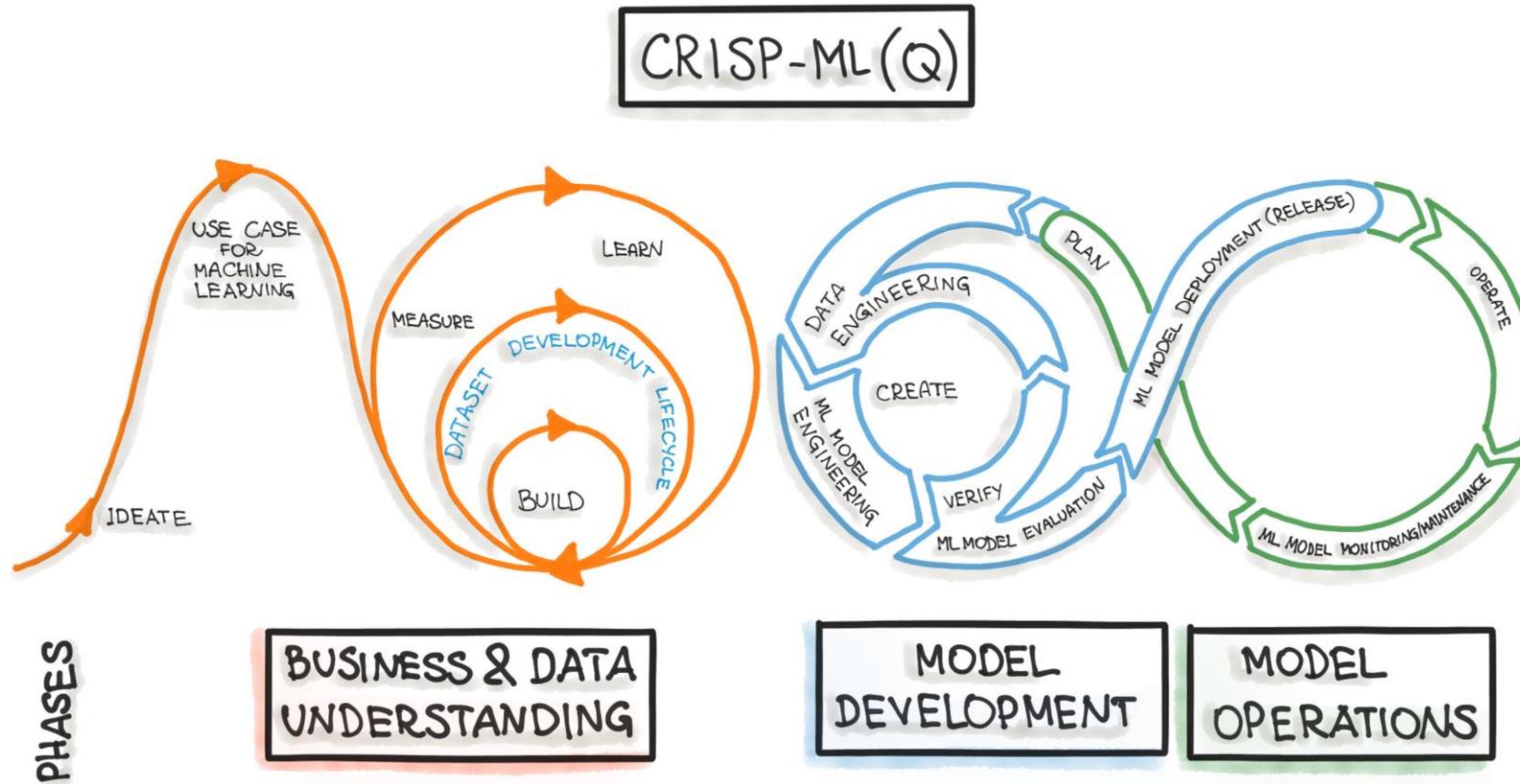
Usos de los datos



Principales ramas de la Inteligencia Artificial

- Generative AI: producir nuevas ideas y contenidos
- ML: Machine Learning
- LLM: Large Languages Models
- Deep Learning: subconjunto de ML, basado en redes neuronales
- NLP, NLG: natural language processing / generation
- VDA: Virtual Digital Assistants, Chatbots
- Sistemas de Recomendación
- Análisis Predictivo

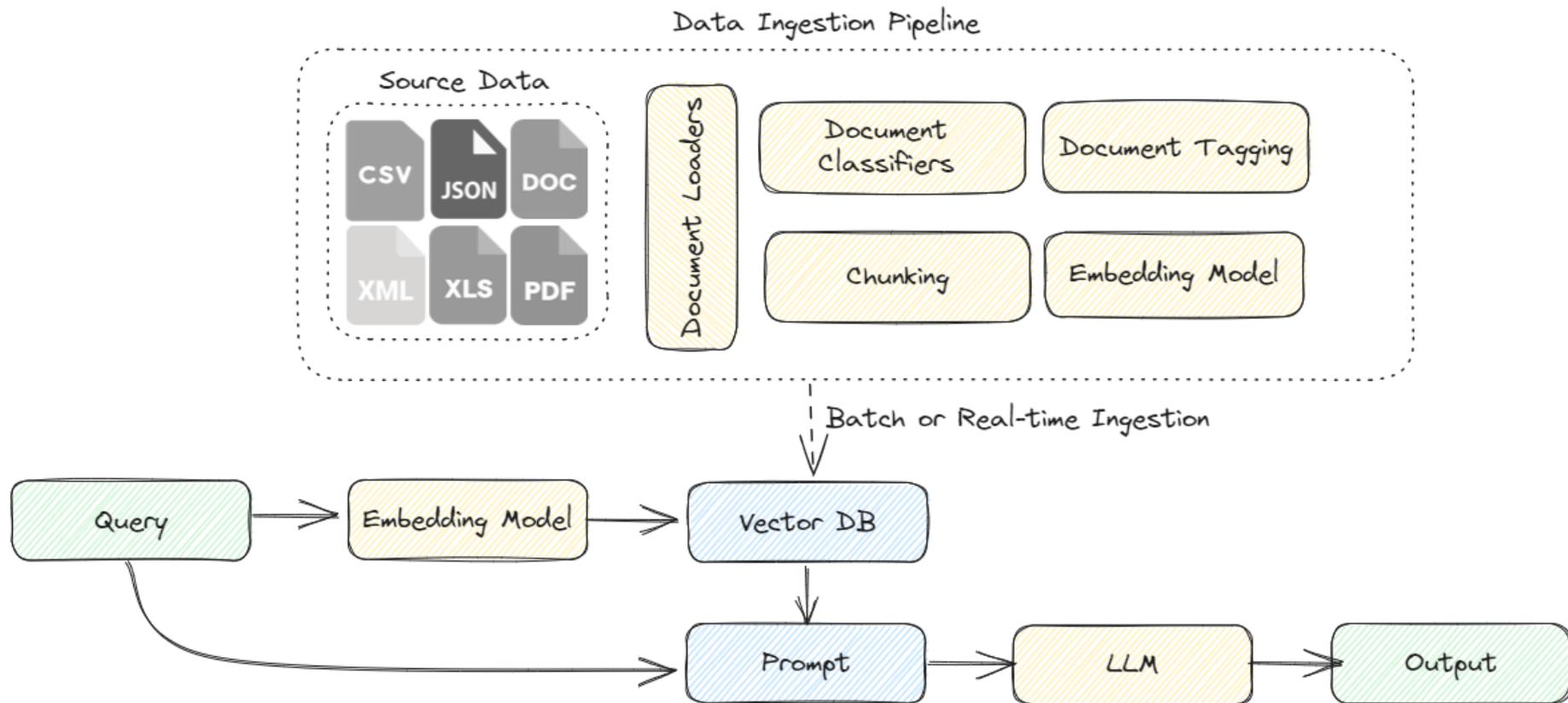
¿Y cómo hacerlo ? ML-OPS



@visonger

ML && LLM

- Arquitectura RAG (Retrieval augmented generation) para LLM también necesita infraestructura de datos



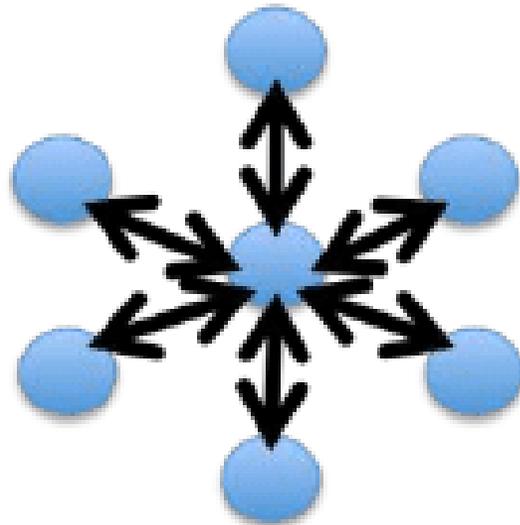
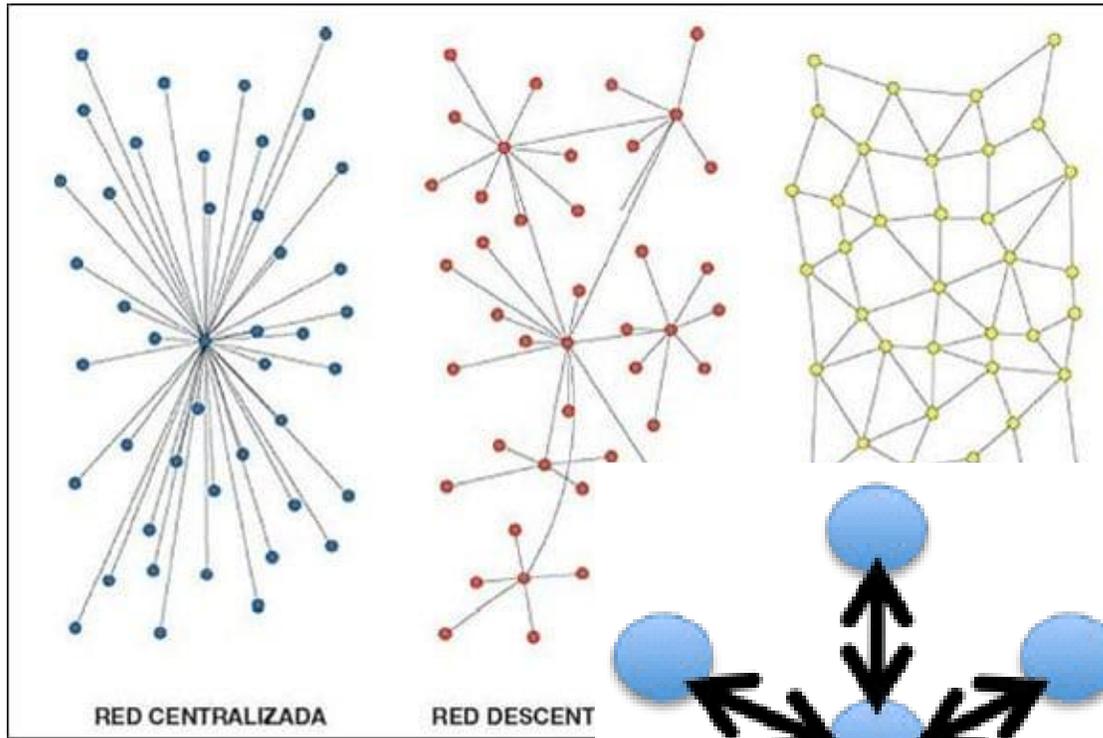
Alimentados por datos de calidad



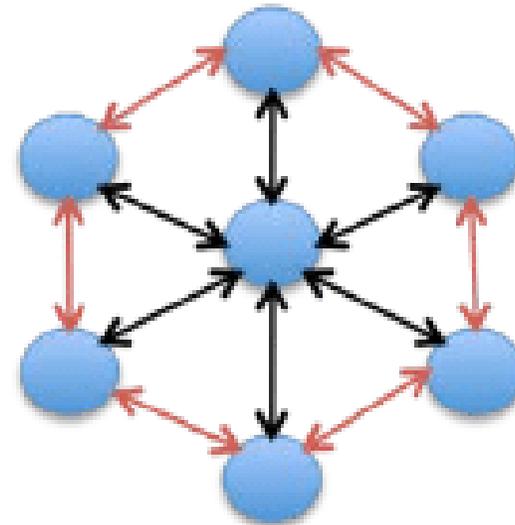
Entonces ¿Qué es un espacio de datos?

Se utiliza a menudo para describir un espacio de datos como una infraestructura descentralizada para el intercambio y el intercambio de datos confiables en ecosistemas de datos basados en principios comúnmente acordados (Nagel y Lycklama 2021).

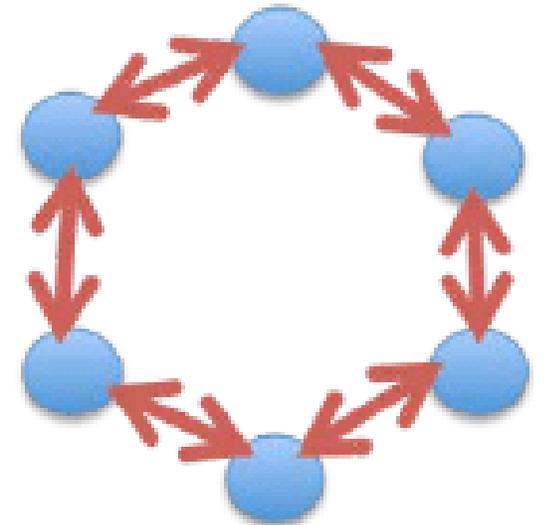
Centralizado vs distribuido y federado



a) Centralized

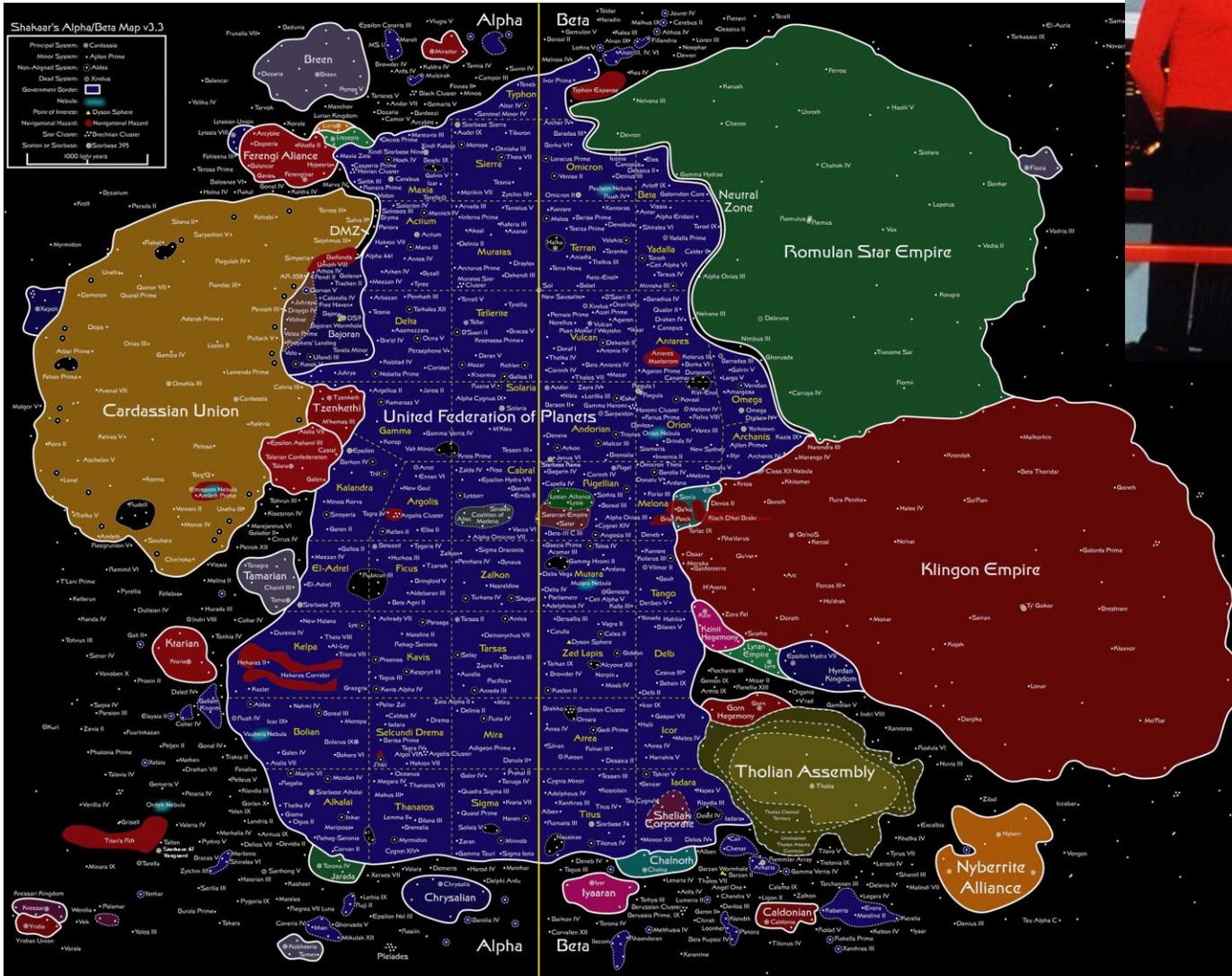


b) Federated



c) Decentralized

A que suena federación



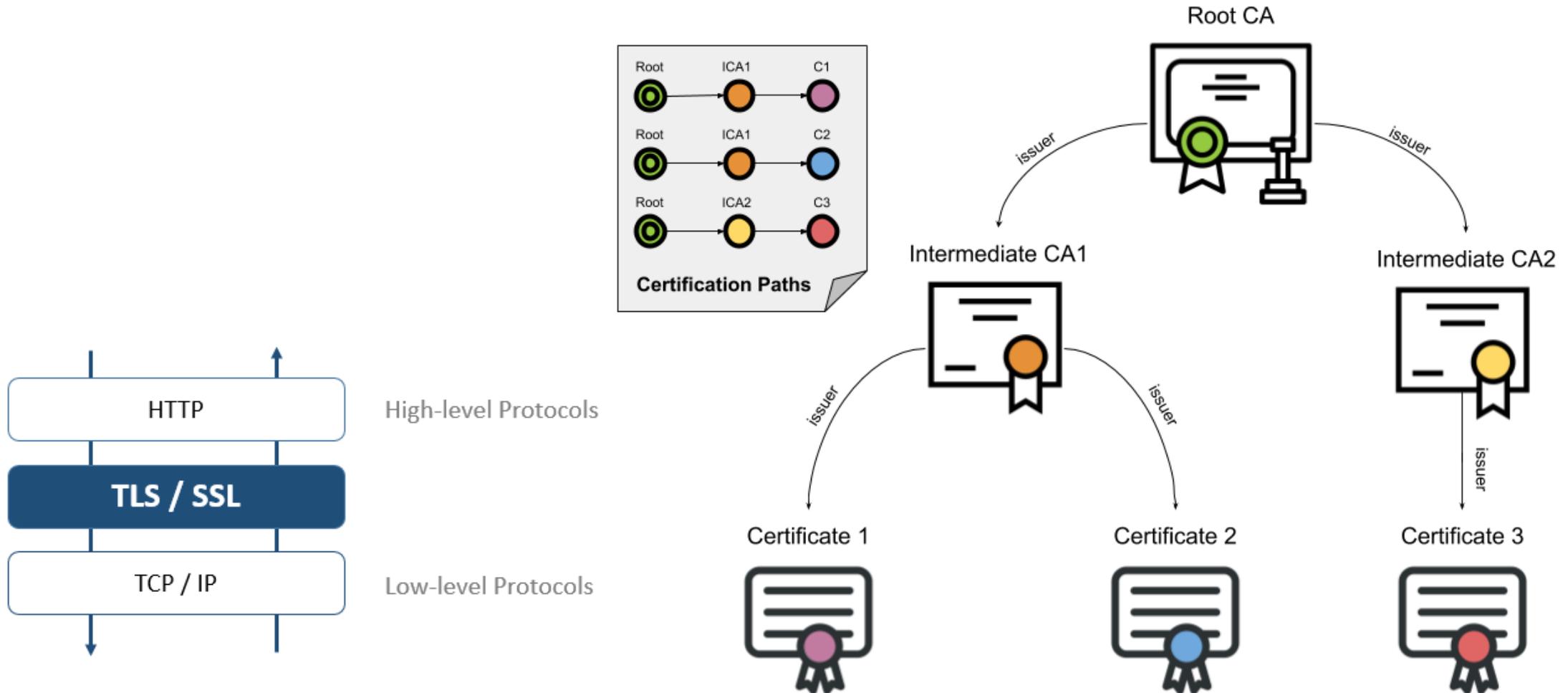
Federación de sistemas

- Los sistemas federados colaboran entre ellos, pero cada participante conserva el control y la gobernanza de sus componentes.
- La clave es la confianza distribuida entre ellos.
- Mecanismos de identidad distribuida o federada para la interacción entre los participantes.

Confianza en el modelo actual de internet

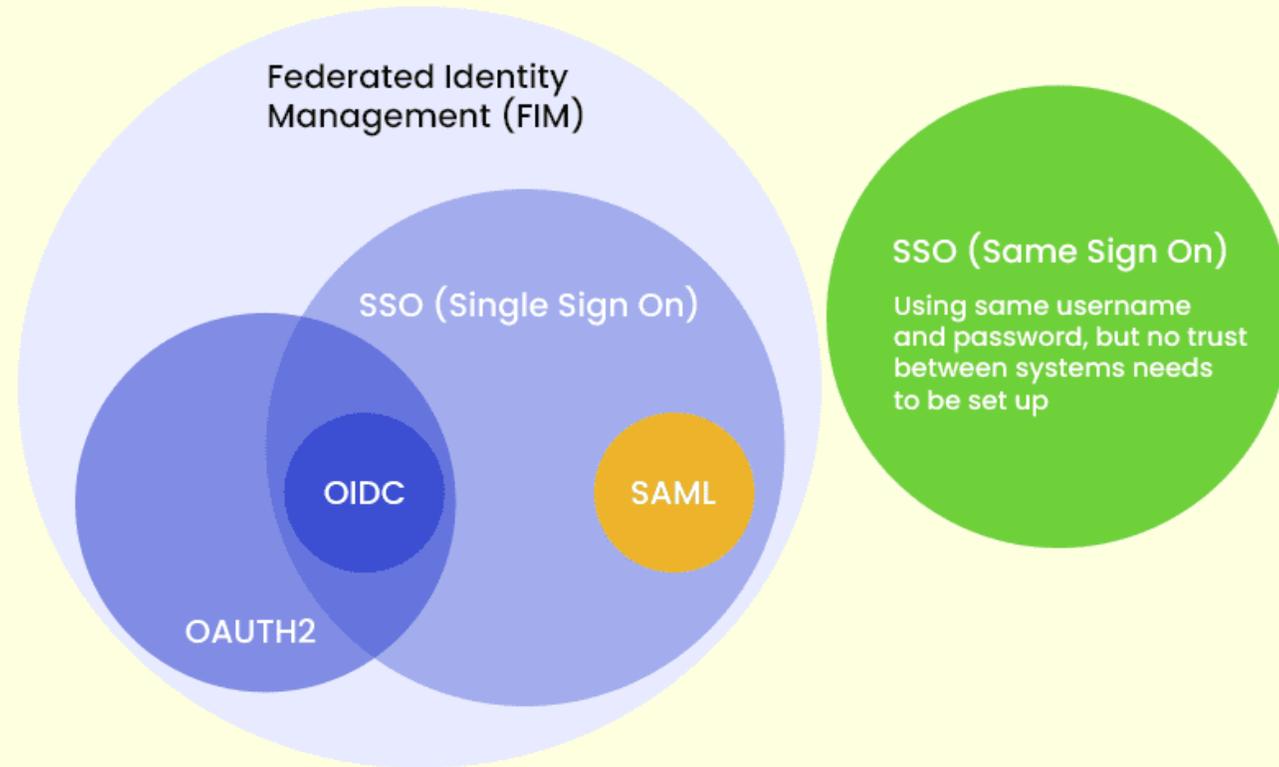
- Modelo de identidad y confianza inicial :
 - Internet solo se conectaban inicialmente empresas y universidades. Todos los participantes pertenecías a estas organizaciones y estaban identificados.
 - Identificación principal : correo electrónico alguien@dominio-internet.
 - Esto se rompió al poder instalarse cualquiera un servidor.
- Desconfianza, en general, en abierto : No puedo fiarme en Internet de alguien que no tenga identificado :
 - Personas
 - Sistemas

Confianza en internet : Sistemas de certificados digitales



Sistemas de federación de identidades

SSO is a Part of Federated Identity Management Architecture



SAML - Security Assertion Markup Language

OIDC - OpenID Connect

OAuth2 - Open Authorization2

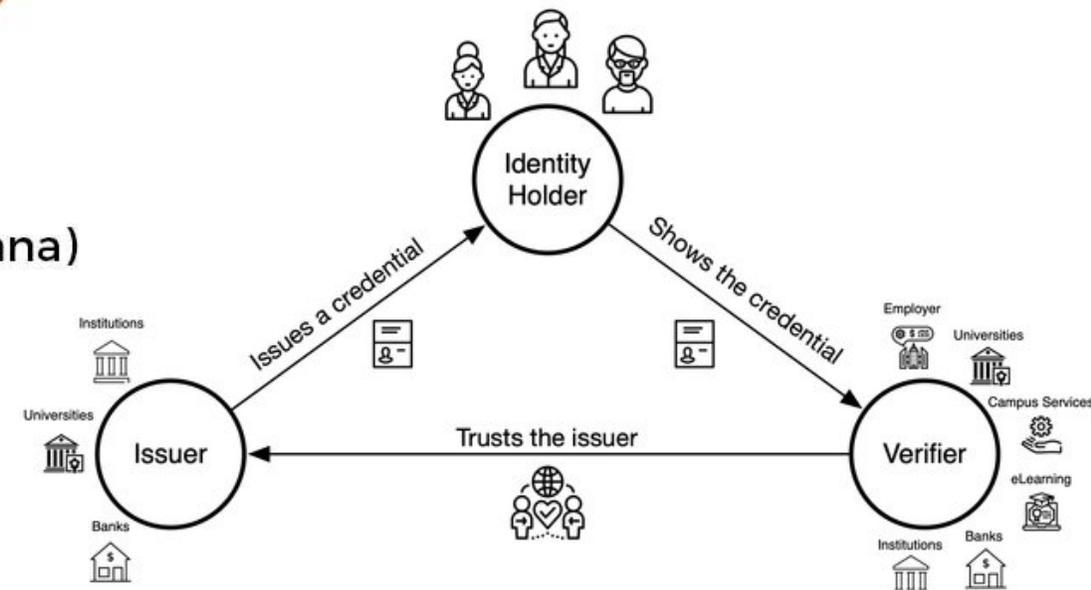
Confianza en identidad autogestionada (self sovereign identity SSI)



Modelo actual de Identidad

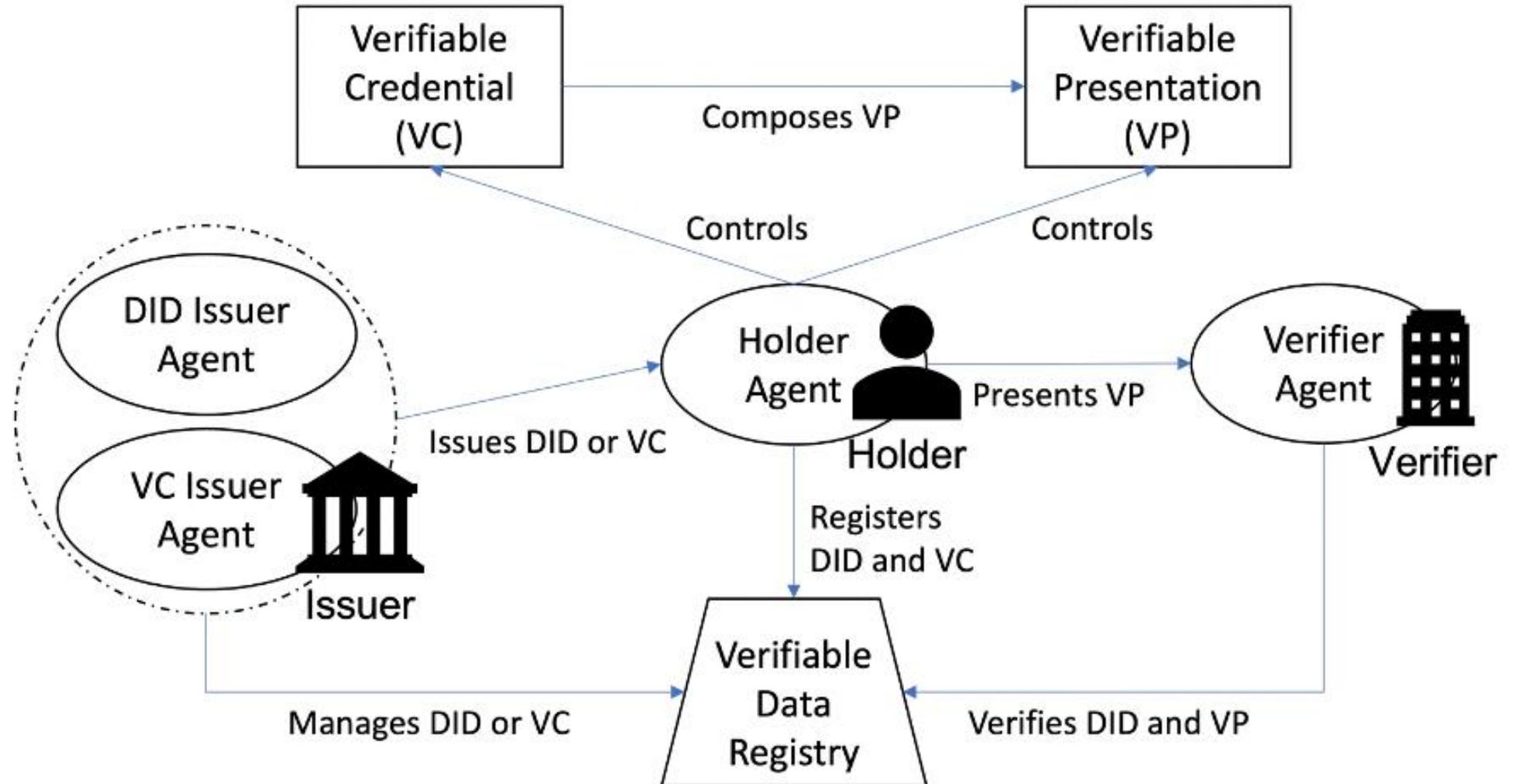


Modelo SSI (Auto-Identidad Soberana)



Necesidad de un punto inicial de confianza (Trust Anchor)

Modelo basado en DID (identificador descentralizado) v VC (Credenciales verificables)

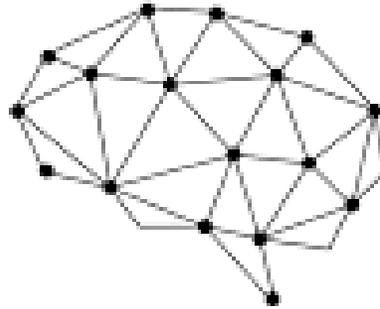
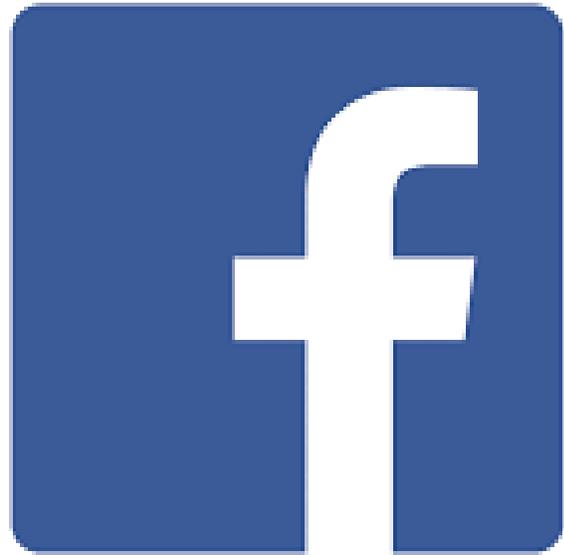


European Strategy for Data

A common European data space, a single market for data



Modelo de intercambios sin transparencia



Cambridge
Analytica



El malo de la película
(han existido más)

Ahora se llama
Meta



Datos abiertos

- Compartidos con todos sin contrapartidas.
- Diversos estándares para ayudar al descubrimiento de los datos.
- Existe una gran infraestructura :



Estandares de metadatos : Dcat (Dcat-ap)

- Etiqueta

- Lata

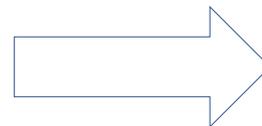
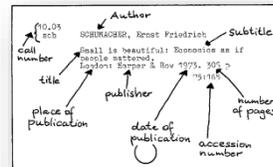
Provides metadata on



- Ficha del catálogo



- Libro



- Descripción de los datos

```
:weather1-7 a dcat:Dataset ;
dct:title "Measurements from weather stations 1-7" ;
dct:description "Data from seven weather stations
showing tempature, humidity,
wind direction and wind speed" ;
dct:modified "2013-07-01" ;
dct:publisher <http://myweather.com/id/myweather> ;
dcat:keyword "weather" ;
dcat:landingpage <http://myweather.com/stations1-7.html> ;
dcat:distribution :weatherdata-xlsx
.

:weatherdata1-7-xlsx a dcat:Distribution ;
dct:format <http://publications.europa.eu/resource/authority/file-type/XLSX> ;
dct:licence <http://creativecommons.org/licenses/by/4.0/> ;
dcat:downloadURL <http://myweather.com/stations1-7.xlsx>
.
```



- Datos

	Temp. °C	Humidity %	Wind direction	Wind speed km/h
Station 1	18.1	60	WSW	18
Station 2	17.5	59	WSW	20
Station 3	18.2	55	SW	22
Station 4	19.0	62	SW	18
Station 5	18.0	65	WSW	19
Station 6	18.2	63	SSW	21
Station 7	17.9	61	SW	22

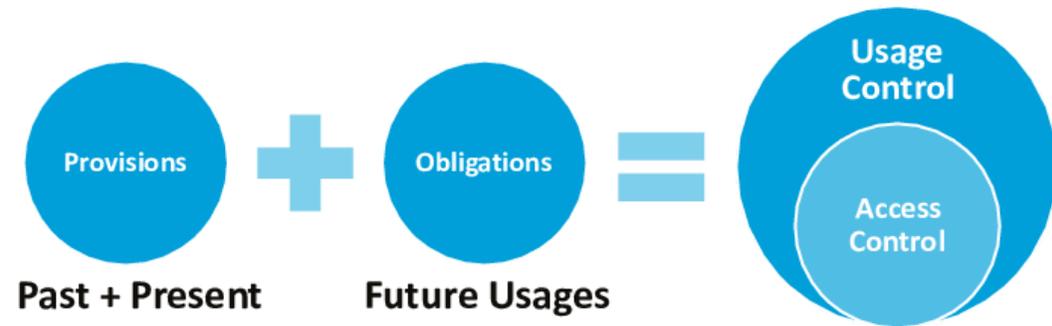
Control de acceso y de uso

- Control de acceso (Data Access Control):

- Especificar quién puede acceder a qué recurso
- También los derechos de acceso a la misma

- Control de uso (Data Usage Control):

- Garantiza la soberanía de los datos
- Regula lo que se permite que suceda con los datos (uso futuro).
- Relacionado con la ingesta y el procesamiento de datos
- Contexto de la protección de la propiedad intelectual, la protección de la privacidad, el cumplimiento de la normativa y la gestión de los derechos digitales



Modelo colaborativo

- Intercambio de datos transparente :
 - Acuerdo claro entre las partes (pueden existir múltiples participantes).
 - Control y acuerdo de lo que ocurre en todo momento.
- Control de Qué se usa y para qué en todo momento.
- Teoría de juegos : Equilibrio de Nash vs Pura competición
 - Se puede ganar más colaborando que en una pura competición.

Confianza

- Definición RAE :

confianza. (De confiar).

1. Esperanza firme que se tiene de alguien o algo.
2. Seguridad que alguien tiene en sí mismo.
3. Presunción y vana opinión de sí mismo.
4. Animo, aliento, vigor para obrar.
5. Familiaridad (en el trato).
6. Familiaridad o libertad excesiva. Ú. m. en pl.
7. V. Abuso de confianza
8. Pacto o convenio hecho oculta y reservadamente entre dos o

más personas, particularmente si son tratantes o del comercio.

de confianza

1. Dicho de una persona: Con quien se tiene trato íntimo o familiar.
2. Dicho de una persona: En quien se puede confiar.
3. Dicho de una cosa: Que posee las cualidades recomendables

para el fin a que se destina.

en confianza

1. Confiadamente.
2. Con reserva e intimidad.

Seguridad

- Definición RAE :

- **seguridad.** (Del lat. securitas, -atis).

1. Cualidad de seguro.
2. Certeza (conocimiento seguro y claro de algo).
3. Fianza u obligación de indemnidad a favor de alguien, regularmente en materia de intereses.

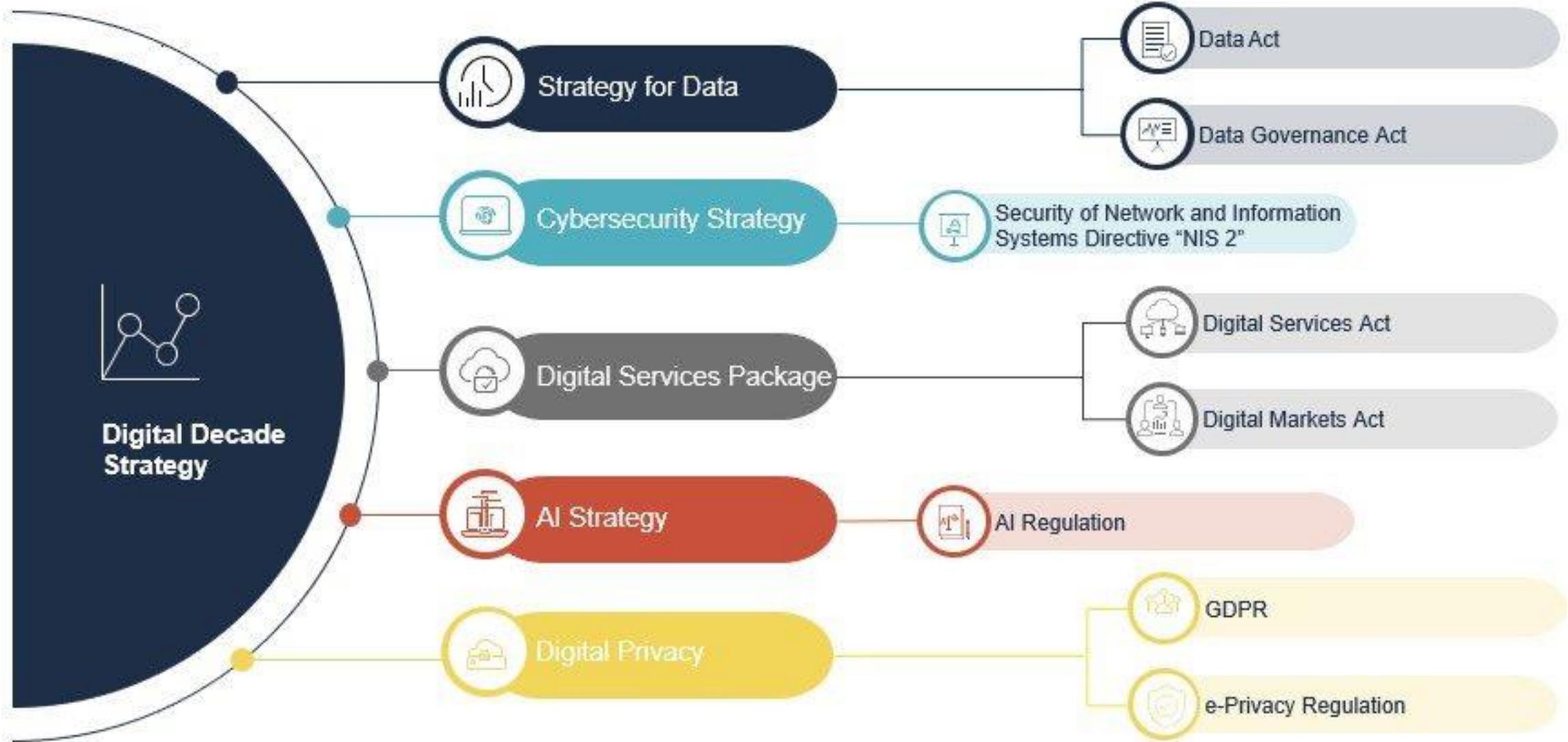
- **Seguridad jurídica.**

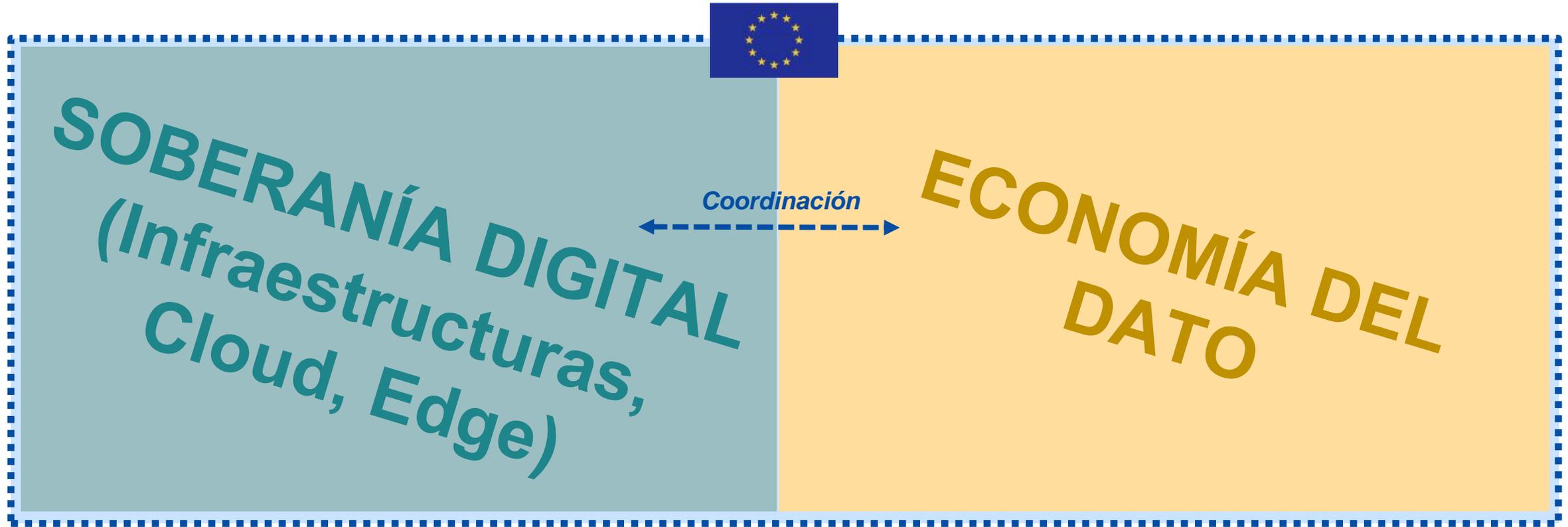
1. Cualidad del ordenamiento jurídico, que implica la certeza de sus normas y, consiguientemente, la previsibilidad de su aplicación. En España es un principio constitucional. **Seguridad social.**

- 1. Organización estatal que se ocupa de atender determinadas necesidades económicas y sanitarias de los ciudadanos.

- **de seguridad.**

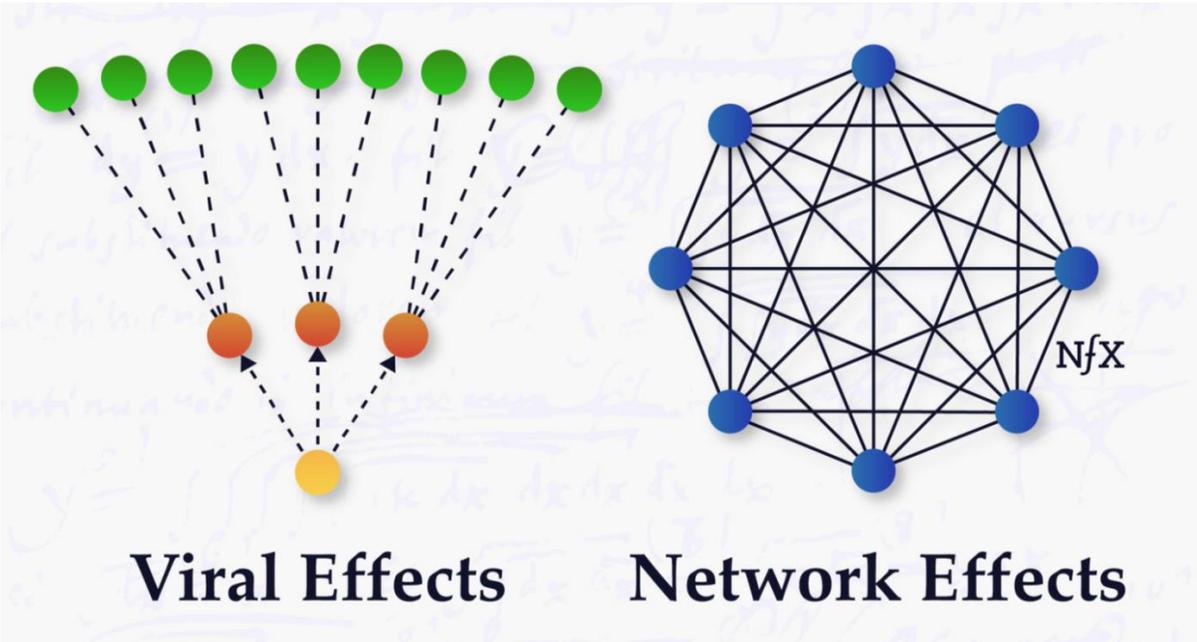
1. Dicho de un ramo de la Administración Pública: Cuyo fin es el de velar por la seguridad de los ciudadanos. Agente de seguridad.
2. Dicho de un mecanismo: Que asegura algún buen funcionamiento, precaviendo que este falle, se frustré o se viole. Muelle, cerradura de seguridad.





Basado en: Comisión Europea

Efecto red : multiplicativo

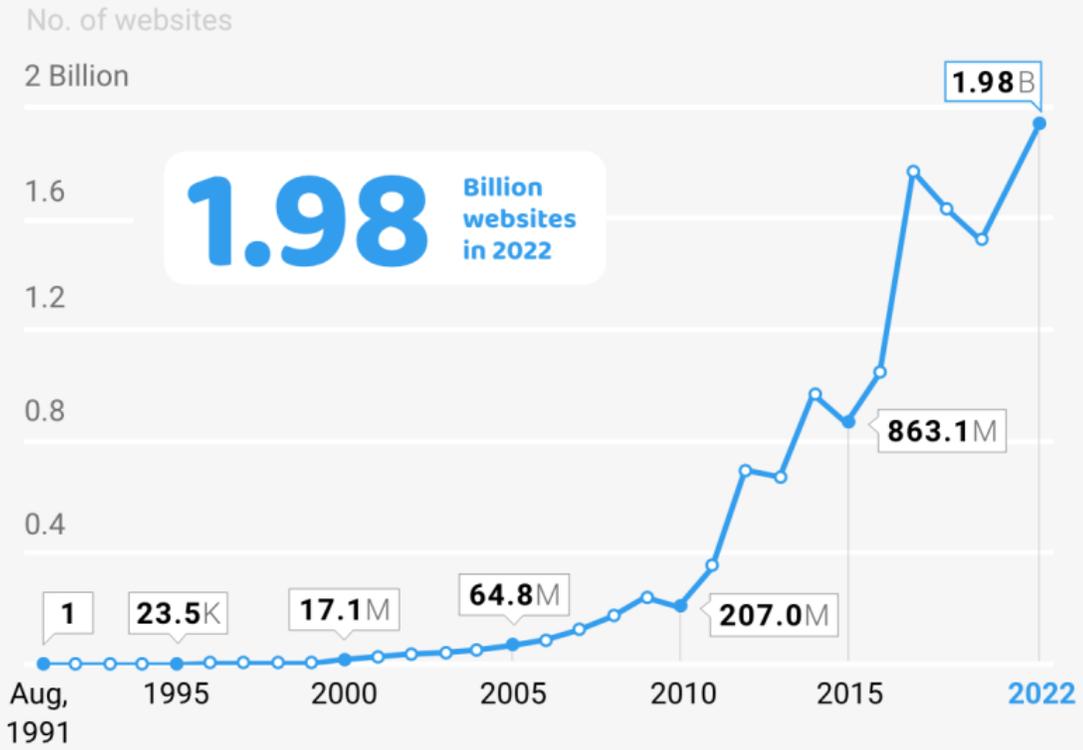


Adopción de la Web

Number of websites in the world



The global number of websites has more than doubled from 2015 to 2022. The website's growth rate from 1991 to 2022.



Source: internetlivestats.com

//
02

Gaia X

Definición Gaia X



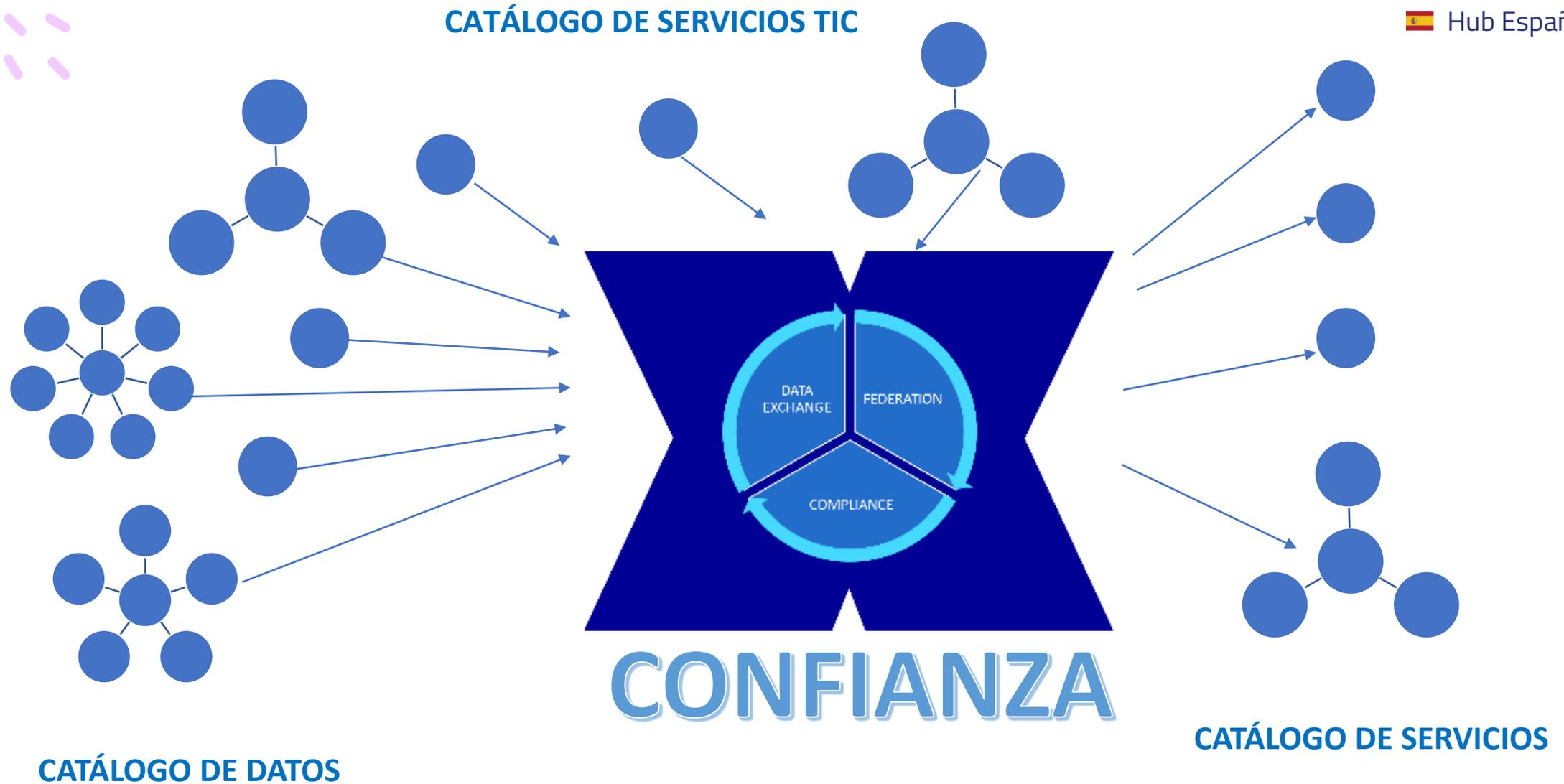
Gaia-X es una **iniciativa europea**, cuyo objetivo es la creación de una infraestructura de datos abierta y segura, cumpliendo con los más altos estándares de soberanía digital al tiempo que promueve la innovación.

Un ecosistema **digital, federado y transparente**, donde los datos y servicios pueden estar disponibles, recopilados y compartidos en un entorno de confianza.

Definición Gaia X



gaia-x
Hub España



Especificaciones de GAIA-X

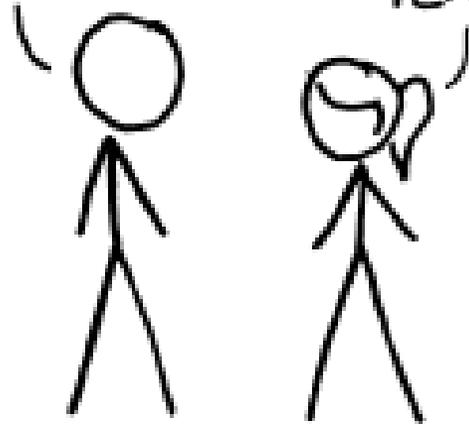
- **Especificaciones:** Los Grupos de Trabajo de la Asociación Gaia-X escriben documentación como:
 - El documento técnico de arquitectura para describir qué es una federación.
 - Trust Framework para describir el conjunto mínimo de atributos y reglas que los participantes deben proporcionar para cumplir con Gaia-X.
 - Reglas de políticas, un conjunto común de políticas y reglas.

HOW STANDARDS PROLIFERATE:

(SEE: A/C CHARGERS, CHARACTER ENCODINGS, INSTANT MESSAGING, ETC)

SITUATION:
THERE ARE
14 COMPETING
STANDARDS.

14?! RIDICULOUS!
WE NEED TO DEVELOP
ONE UNIVERSAL STANDARD
THAT COVERS EVERYONE'S
USE CASES.



SOON:

SITUATION:
THERE ARE
15 COMPETING
STANDARDS.

Data Spaces Business Alliance : DSBA & DSSC



gaia-x

 Hub España

Data Spaces Business Alliance

Unleashing the Data Economy



Our X-Model

Connecting Data & Infrastructure Ecosystems



Advanced Services

New (Cross-) Sector Innovations / Applications build from service composition.



Data Spaces / Federations

Interoperable & portable (Cross-) Sector data-sets and services.



Data Exchange

Anchored contract rules for access and data usage.



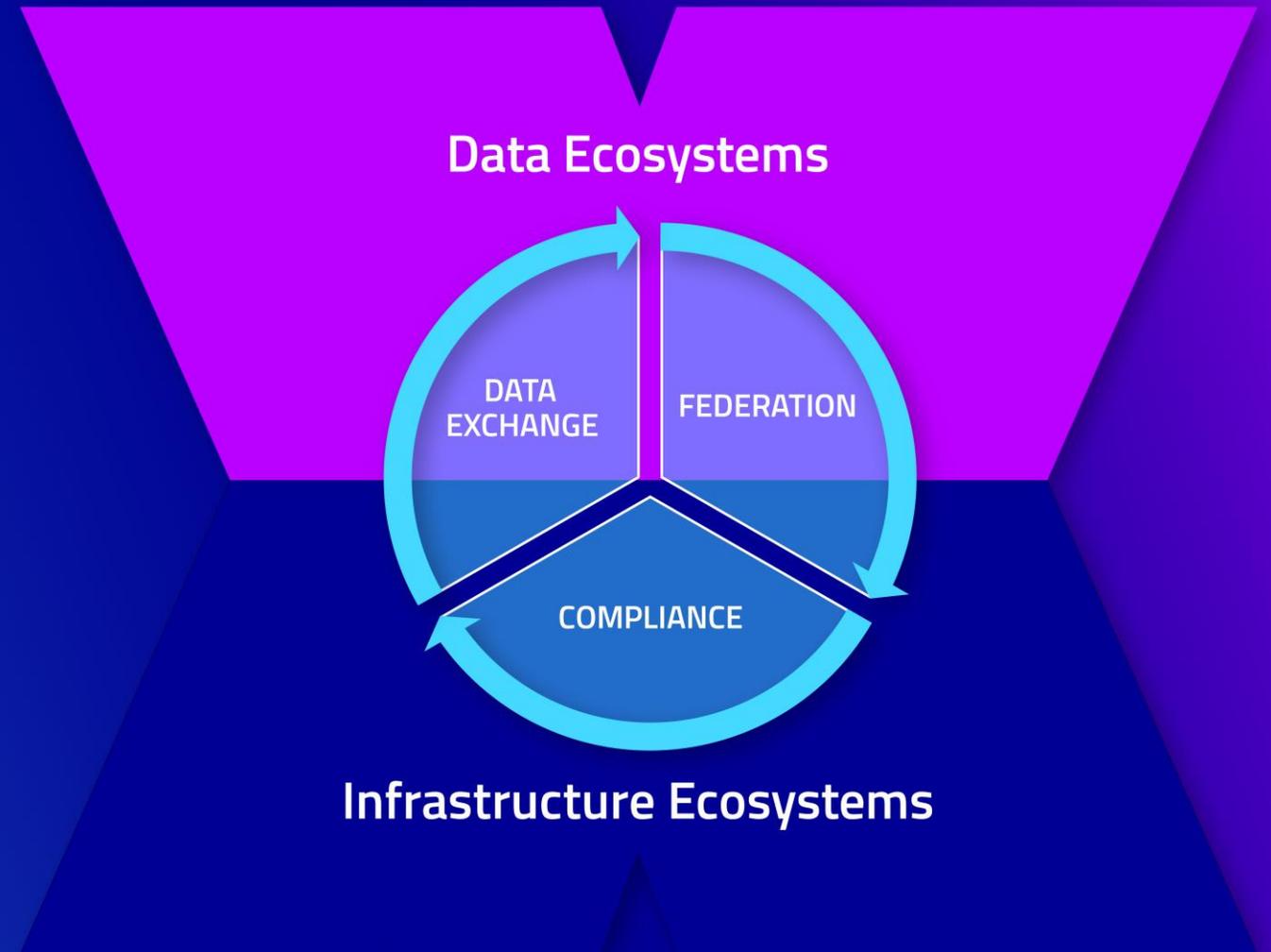
Gaia-X Compliance

Decentralized services to enable objective and measurable trust.



Label framework

Gaia-X and ecosystem specific Labels to ease market adoption through autonomy and self-determination.



Misión de Gaia-X:

«Creación de **ecosistemas federados, abiertos, seguros** y transparentes donde los **datos y los servicios asociados** cumplen un conjunto de **reglas comunes** que permiten compartirlos, crearlos y recopilarlos»



Ecosistemas federados
(federación)

Comunidades de proveedores y consumidores de servicios que compiten y colaboran en la provisión y consumo de servicios mediante relaciones de igual a igual y que actúan de forma autónoma.



Abiertos

Varios sentidos:

- Abiertos a todo tipo de participantes (Bajas barreras de entrada)
- Software open source
- Interoperables con otros ecosistemas, evitando silos



Seguros

Mecanismos para garantizar la confianza en las credenciales

Etiquetado del nivel de confianza de los servicios (labels)



Para servicios de datos y asociados

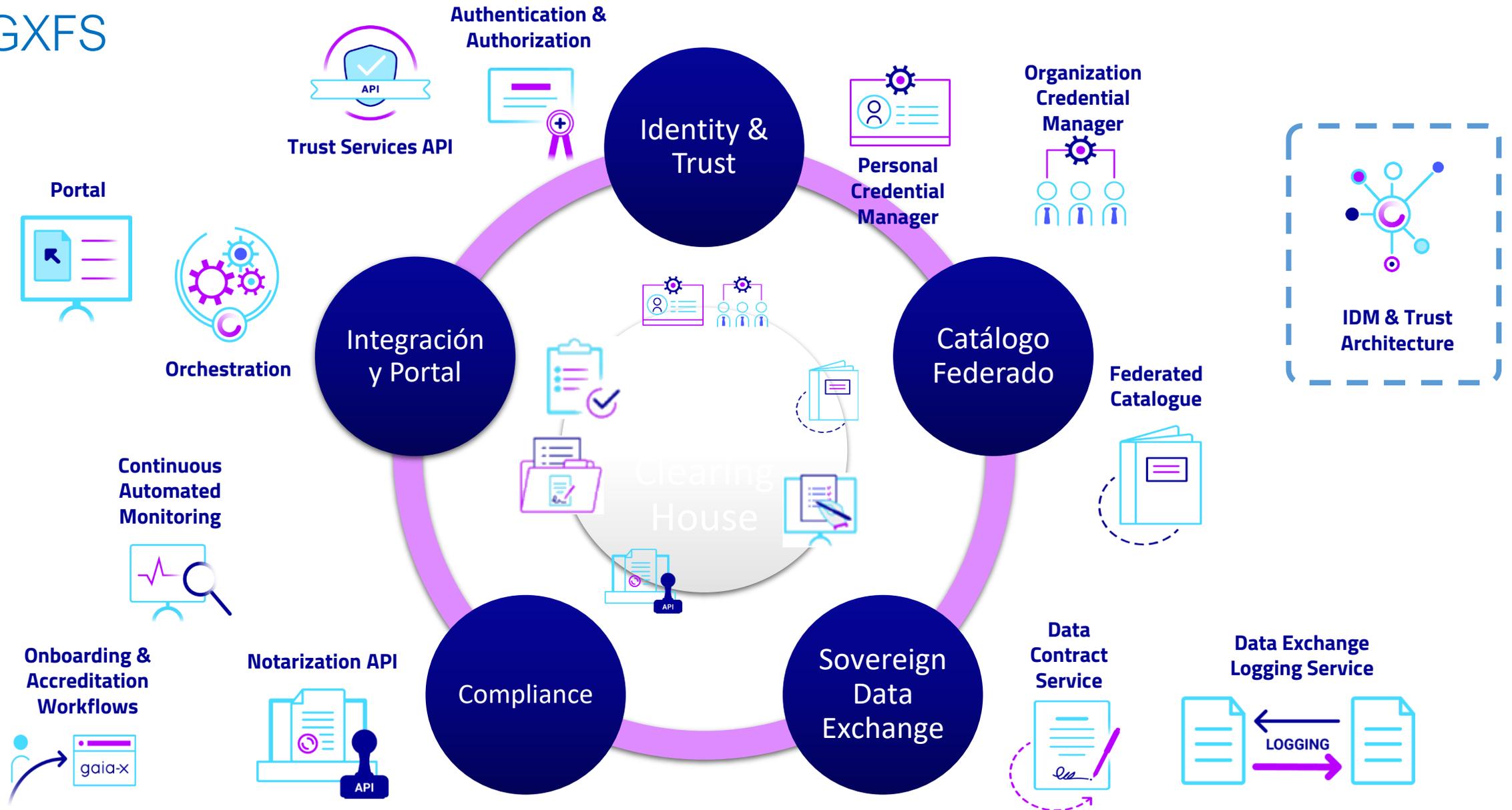
- Servicios de compartición de datos
- Aplicaciones para su explotación
- Recursos computacionales en la nube

Gaia-X denomina a los espacios de datos «**infraestructuras de datos**»

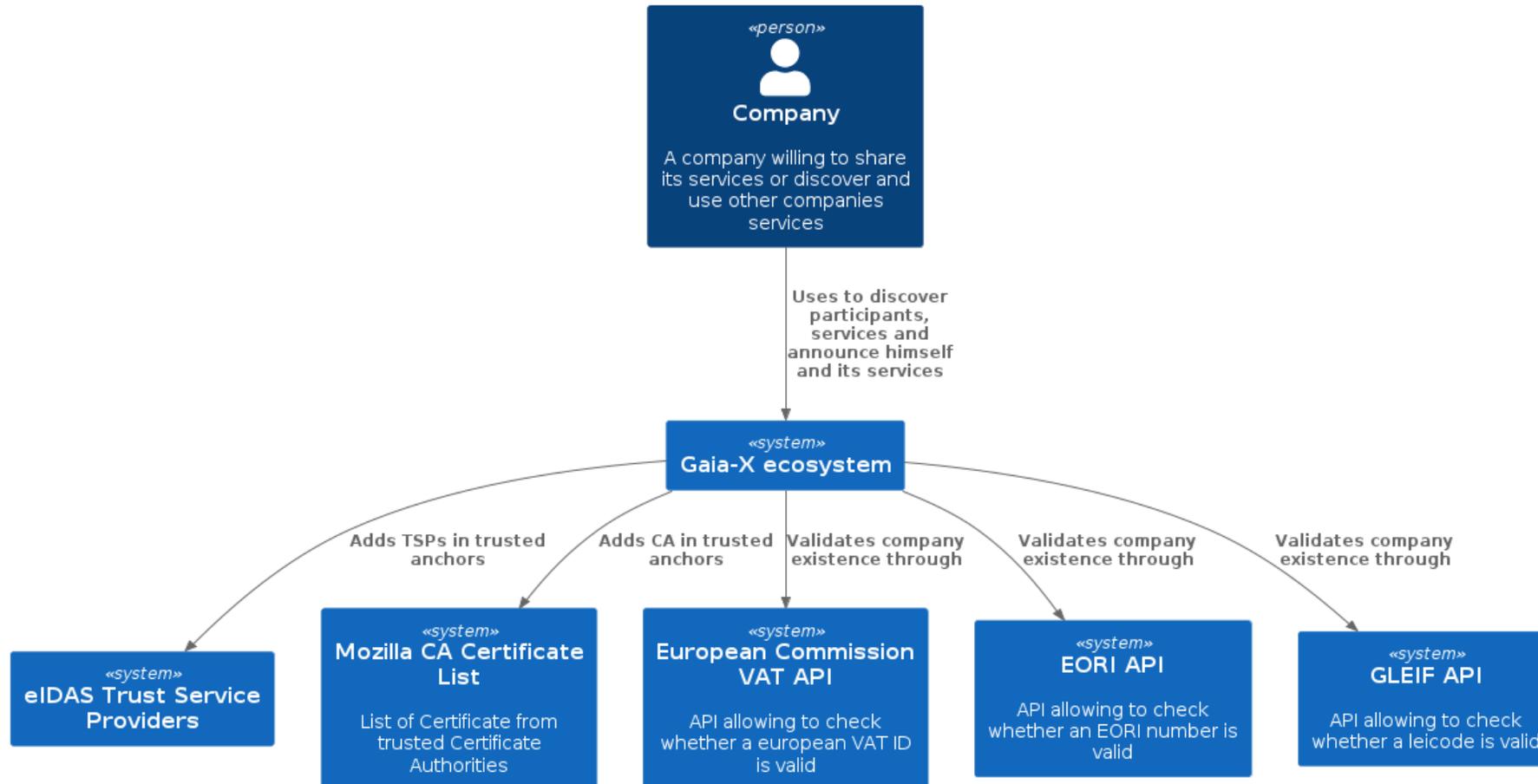


Con reglas comunes

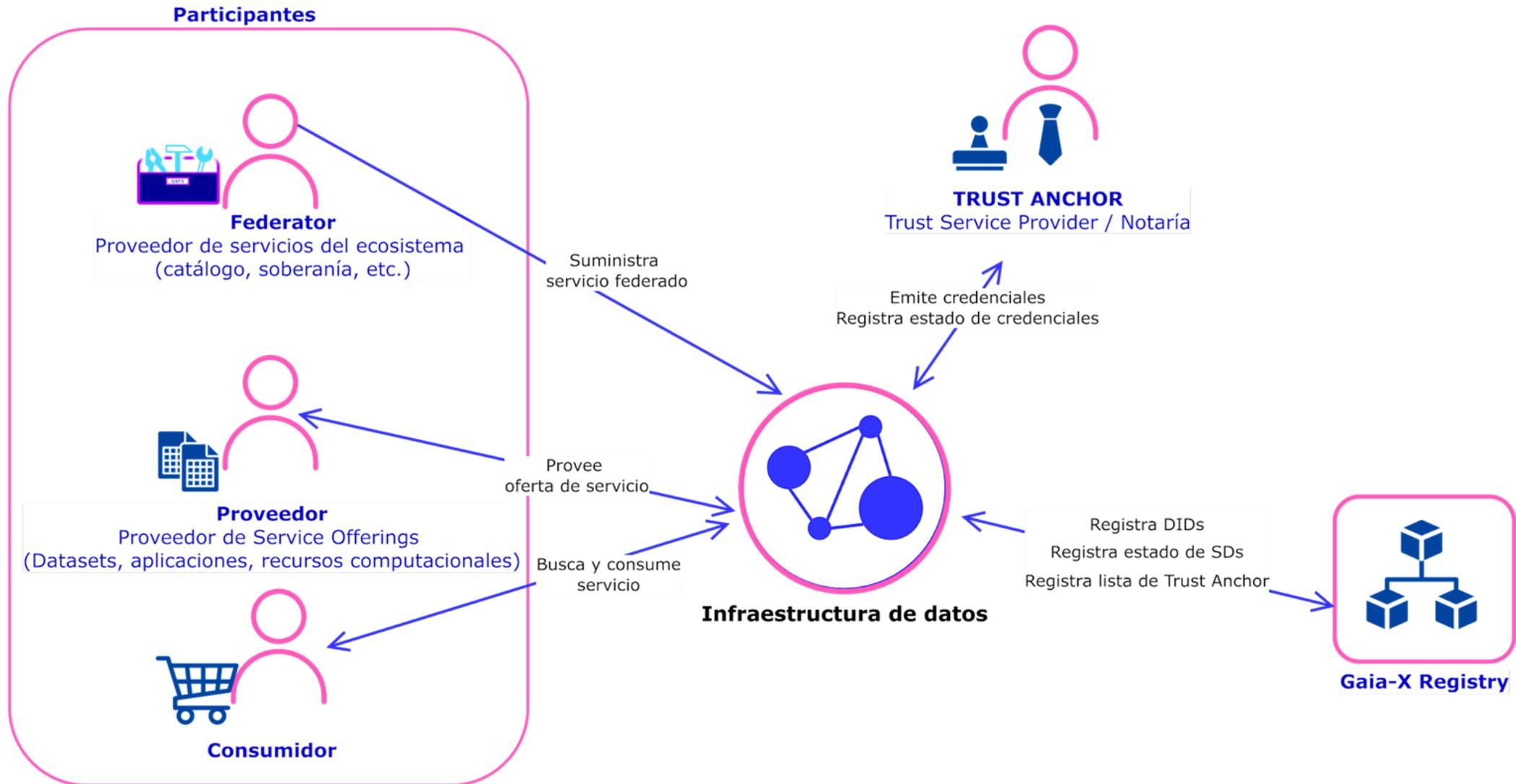
- Reglas de interoperabilidad basadas en especificaciones abiertas y estándares
- Legales
 - Organizativas
 - Semánticas
 - Técnicas



Arranque de la confianza : Gaia-X digital clearing house **GXDCH** (Cámara de compensación)



Arquitectura general de Gaia-X: roles



Servicios del GXDCH

Servicios obligatorios del GXDCH



Gaia-X Registry



Gaia-X Compliance

Servicios recomendados del GXDCH



Gaia-X Catalog



Gaia-X Credential Manager
(e-Wallet)

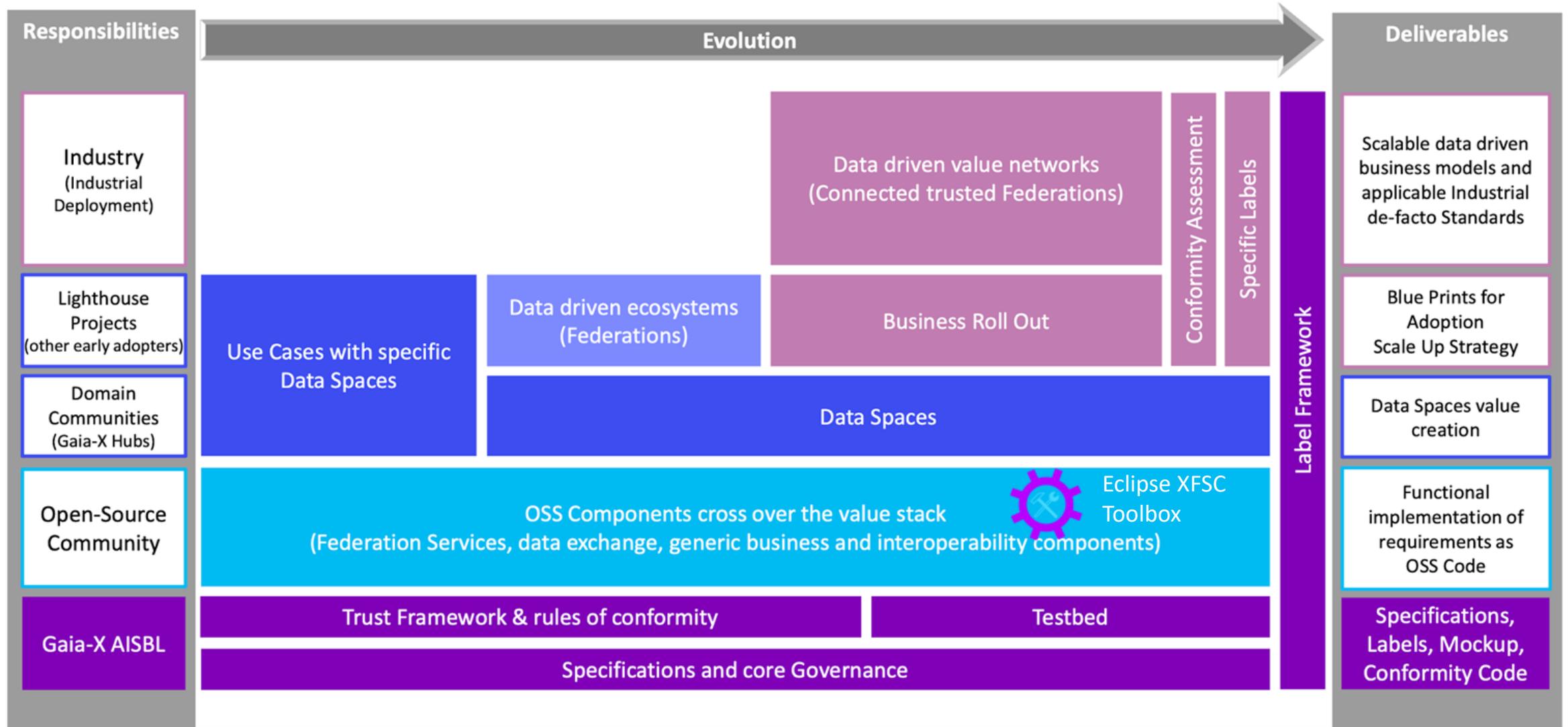


Wizard

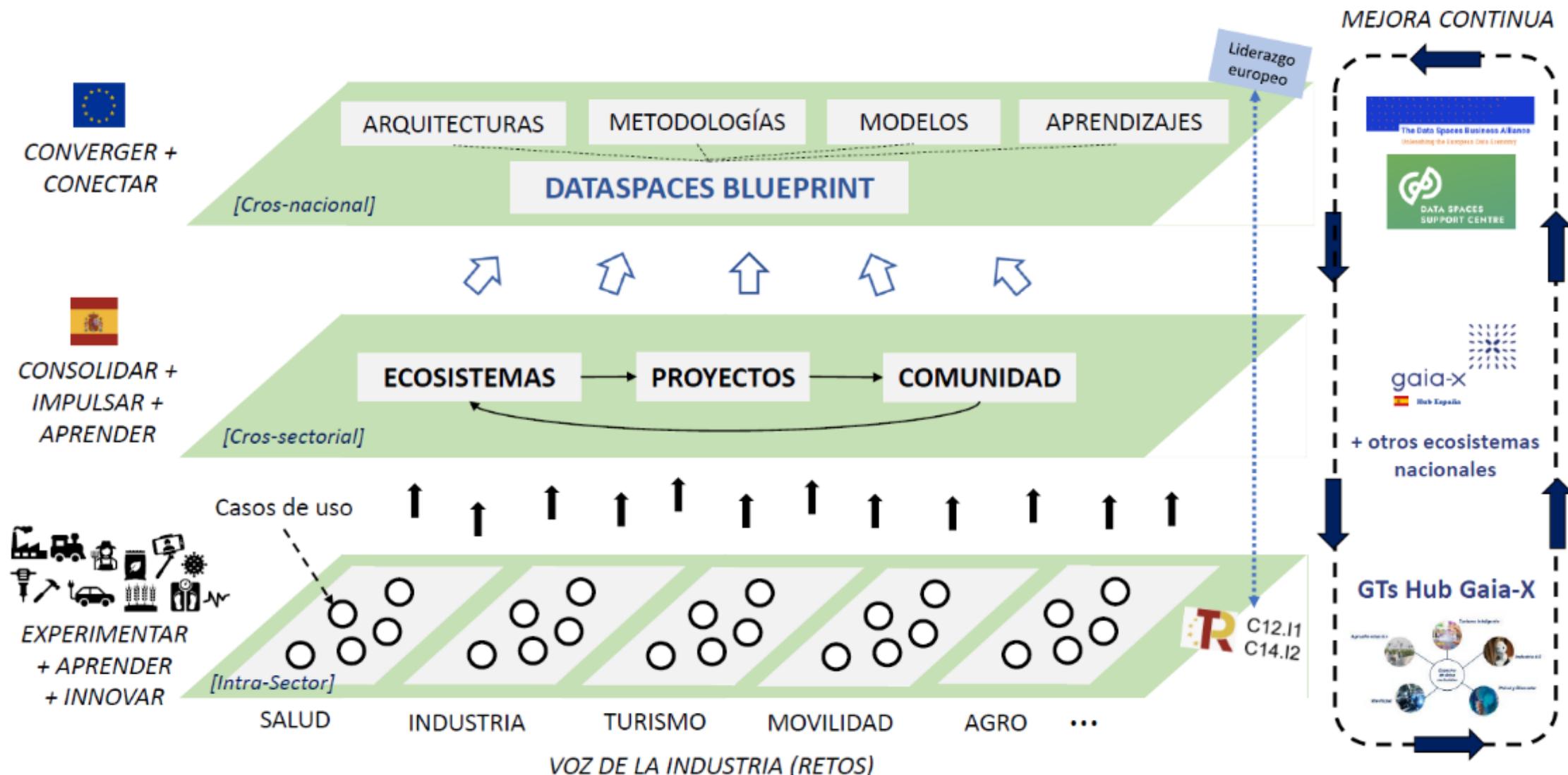


Notary

Gaia-X: Responsabilidades y entregas



Estructura Gaia X





Gracias

Si estás interesado en ser miembro de **Gaia-X España** ponte en contacto con nosotros escribiendo un correo electrónico a:

administracion@gaiax-spain.com

