

# Posicionamiento sobre los Espacios de Datos

Enero 2021

Iniciativa Interplataforma en Big Data e Inteligencia Artificial



## Edición y coordinación

Sáez-Domingo, Daniel; Costa-Soria, Cristóbal; Beltrán-Blanco, Liliana - INSTITUTO TECNOLÓGICO DE INFORMÁTICA  
Gil, Guillermo. TECNALIA

## Contribuidores

### PLANETIC

Bernabeu-Aubán, José; Sáez-Domingo, Daniel; Arjona-Aroca, Jordi; Costa-Soria, Cristóbal; Alonso-Román, Daniel; Beltrán-Blanco, Liliana; ITI - INSTITUTO TECNOLÓGICO DE INFORMÁTICA  
Gil, Guillermo. TECNALIA  
Corcho, Oscar. UNIVERSIDAD POLITÉCNICA DE MADRID  
Angulo, Cecilio. TASK FORCE DE MOVILIDAD URBANA  
Ordoñez, Dolores. ANYSOLUTION

### Plataforma Española de Innovación en Tecnologías Sanitarias.

Muñoz, Sergio.  
Fuentes, Jorge. FENIN – FEDERACIÓN ESPAÑOLA DE EMPRESAS DE TECNOLOGÍA SANITARIA

### Plataforma de Medicamentos Innovadores

Martín, Amelia; Montes, Fátima. FARMAINDUSTRIA - ASOCIACIÓN NACIONAL EMPRESARIAL DE LA INDUSTRIA FARMACÉUTICA

### MANU-KET, Plataforma Tecnológica Española de Fabricación Avanzada

Anzaldi, Gabriel. EURECAT – CENTRO TECNOLÓGICO DE CATALUÑA

### PACKNET, Plataforma Tecnológica Española de Envase y Embalaje

García, Belén.

### PTEA, Plataforma Tecnológica Española del Agua

Sobreira, Juan Luis; Brandón, Oscar; Garabato, Lucía. INSTITUTO TECNOLÓGICO DE GALICIA  
Hernández, Jose Luis. CENTRO TECNOLÓGICO CARTIF  
Dirección General del Agua - MITERD

### PTEPA, Plataforma Tecnológica Española de la Pesca y la Acuicultura

Orden, Cristina.

### Vet+i, Plataforma Tecnológica Española de Sanidad Animal

Fernández, Patricia

### Thinktur, Plataforma Tecnológica del Turismo

Fernández, Pedro. ITH – INSTITUTO TECNOLÓGICO HOTELERO

### Connected Mobility Hub

Carballo, Luis; Díez, Sergio.

### IDSA Hub Spain

Lázaro Oscar. INNOVALIA  
Castelvi, Silvia. International Data Spaces Association (IDSA)

## Contenido

0	Prólogo .....	5
1	La Economía del Dato .....	6
1.1	Los datos y la digitalización .....	7
1.2	El Ciclo del Dato .....	7
1.3	Europa y los DataSpace .....	8
1.4	Políticas e instrumentos relacionados con el fomento de los DSES .....	9
2	Espacios de Datos como elementos clave para la explotación de datos .....	16
3	Servicios habilitados por los Espacios de Compartición y Explotación Datos .....	19
4	Casos de uso de un Espacio de Compartición y Explotación de Datos .....	21
4.1	Casos de uso simples .....	21
4.2	Posibles ejemplos o situaciones de uso .....	22
5	Retos legales, políticos y regulatorios; de mercado y técnicos de los Espacios de Datos .....	25
5.1	Los retos legales, políticos y regulatorios .....	25
5.2	Los retos de mercado/financieros .....	26
5.3	Los retos técnicos .....	28
6	Modelos y casos de éxito de espacios de Datos .....	29
6.1	Modelos de referencia .....	29
6.2	Casos de éxito .....	37
7	Recomendaciones para el despliegue de DSES en España .....	44
7.1	Actores en los ecosistemas y DSES .....	44
7.2	Del Piloto a la Explotación. ....	46
7.3	Recomendaciones finales .....	70
8	Referencias .....	72
9	Anexos .....	73
9.1	Caso GAIA-X .....	73
9.2	Caso International Data Spaces (IDS) .....	77

## RESUMEN EJECUTIVO

---

Hoy en día, se genera y captura cada vez un mayor volumen de datos. Al mismo tiempo, se pone en evidencia el potencial existente en su explotación y extracción de valor, como clave para la digitalización de la sociedad y la competitividad europea. Sin embargo, su nivel de aprovechamiento aún es bajo, en gran medida motivado por la falta de estructuras adecuadas para el almacenamiento, consulta, compartición y explotación de estos datos, así como el desconocimiento de mecanismos para su tratamiento.

Como consecuencia, en Europa ha surgido la necesidad de habilitar Espacios de Datos e instalaciones de prueba y experimentación en las que se posibilite tanto el uso de datos como su privacidad. De esta forma, se busca que los Espacios de Datos promuevan un ecosistema en el que se compartan datos y se generen nuevos servicios y productos basados en ellos. A este respecto, los estamentos europeos han empezado a impulsar diferentes iniciativas con el fin de fomentar el desarrollo y la interconexión de estos Espacios de Datos favoreciendo así la explotación cruzada de los datos disponibles en ellos, extrayendo valor de los mismos mediante la aplicación de técnicas de Big Data, analítica avanzada e Inteligencia Artificial.

Este documento tiene el objetivo de describir qué son y qué utilidad tienen, tanto los **Espacios de Datos (DataSpaces)** como los **Espacios de experimentación y explotación de datos (Data Innovation Spaces)**, como eje fundamental para generar una economía competitiva, así como la combinación de ambos, en lo que podemos llamar **Espacios de Compartición y Explotación de Datos (DSES)**.

En la UE destacan el Programa “Digital Europe” y la Estrategia Europea de Datos, publicada en febrero de 2020. Por su parte, en España destacan iniciativas como la nueva Agenda Digital Española, el Plan España Digital 2025, el Plan de Recuperación, Transformación y Resiliencia y la nueva Estrategia Nacional de Inteligencia Artificial publicada en diciembre de 2020.

Como ejemplos de aplicabilidad se abordan diferentes escenarios y casos de uso de forma general, para luego enumerar una serie de situaciones concretas que ejemplifican algunos casos de utilización. Además, también se presenta un análisis de los principales retos a nivel legal, tecnológico y de mercado para el despliegue de los Espacios de Datos y Espacios de Compartición y Explotación de Datos.

También se realiza una revisión de los principales modelos de referencia de Espacios de Datos europeos y un resumen de los casos de éxito desarrollados en el país.

Finalmente, se exponen una serie de recomendaciones y pasos para el desarrollo de Espacios de Compartición y Explotación de Datos en España, junto con un repaso a las características de diferentes sectores relevantes para España en los que existe masa crítica y ejemplos de Espacios de Datos Sectoriales (Manufactura, Salud, Pesquero y Acuícola, Sanidad Animal, Envase & Embalaje, Turismo, Agua y Transporte). Para cada uno de ellos, se describen aspectos como los principales beneficios para el sector, aplicaciones, plataformas de referencia, datos abiertos existentes, legislaciones, así como barreras y retos para su creación.



## 0 Prólogo

Los datos son el “nuevo petróleo”; son nuevos activos valiosos para los productos y servicios emergentes de los próximos años. Quienes dominen su procesamiento inteligente ejercerán un enorme control sobre el mercado y marcarán el futuro en un contexto cada vez más interconectado, automatizado y complejo.

Actualmente, unas pocas grandes plataformas tecnológicas (GAFAM (Google, Amazon Facebook, Apple y Microsoft)) ostentan un enorme control sobre la gran mayoría los datos existentes de naturaleza B2C. Existe sin embargo una gran cantidad de datos de carácter industrial y no personal, que seguirá creciendo con el despliegue de redes y tecnologías IoT. Estos datos serán determinantes tanto para la mejora en productividad y competitividad de las empresas, como para la resolución de los grandes desafíos relacionados con el bienestar o el medioambiente, retos que no podrán superarse sin una gestión inteligente de los datos.

Las empresas y organizaciones de todo tipo podrían beneficiarse de este potencial de mejora a través de un mercado abierto, seguro y confiable de datos, de una economía de datos, a través de espacios que faciliten el acceso, uso, interoperabilidad y control de los datos para los participantes.

***Los Espacios de Datos son estructuras que proporcionan confianza y seguridad para la compartición voluntaria de datos entre diversos agentes de manera homogénea a través de mecanismos combinados de gobernanza, organizativos, legales y técnicos. Estos espacios facilitan la interoperabilidad para acceder a los datos o transferirlos y posibilitan su reutilización eficiente y legítima en un contexto de soberanía y control para las partes sobre sus propios datos. Los Espacios de Datos permiten:***

- Crear valor para los participantes del Espacio de Datos al facilitar la búsqueda, el acceso, la interoperabilidad y la reutilización de los datos, independientemente de su ubicación física, en un entorno abierto, confiable y seguro.
- Garantizar la soberanía sobre los datos de cada partícipe.
- Proporcionar mecanismos de negocio relacionados con los activos de datos.
- Reducir los costes de transacción de datos de calidad.
- Potenciar el uso de activos valiosos, los datos, entre empresas o entre estas y las administraciones.
- Fortalecer y desarrollar la innovación vinculada a la industria de datos.

Por otro lado, la disponibilidad de Espacios de Datos facilita enormemente su explotación y la extracción de valor de los mismos a través de servicios basados en Inteligencia Artificial o Big Data, entre otras técnicas. Por ello, **es necesario visualizar los Espacios de Datos siempre en combinación con Infraestructuras preparadas para la explotación, experimentación e innovación con Datos, en lo que podemos llamar Espacios de experimentación y explotación de datos (Data Innovation Spaces).**

La combinación de Espacios de Datos y Espacios de Innovación con Datos deriva en lo que se puede citar como ***Espacios de Compartición y Explotación de Datos*** (DSES por sus siglas en inglés).

## 1 La Economía del Dato

Todos los avances científicos y técnicos a lo largo de la historia se han debido a la capacidad de plantear experimentos para obtener datos y a la habilidad para poder luego analizar los datos que se han derivado de ellos.

El desarrollo tecnológico actual, hace ya tiempo que ha posibilitado la captura y almacenamiento de cantidades masivas de datos de muy diversa índole. Datos obtenidos a partir de dispositivos que interactúan con el mundo real o virtual y que cada vez están más interconectados vía las redes de comunicación.

Esta tendencia no va a hacer más que aumentar, impulsada por el desarrollo de las redes de comunicación (el 5G va a posibilitar la interconexión de, prácticamente, todo dispositivo electrónico), de los procesadores (su miniaturización y eficiencia energética hace posible la inclusión de los mismos en prácticamente cualquier artefacto), y la sofisticación de diversos tipos de sensores que pueden acompañar los más diversos objetos.

Muchos de los datos capturados proceden de un ámbito público y son libremente accesibles, cayendo bajo lo que se ha venido en entender como “open data”. Aún otros muchos de los datos proceden de ámbitos sensibles que pueden desvelar aspectos individuales de personas y que se encuentran justamente protegidos por diversas legislaciones y códigos éticos.

Finalmente, en el ámbito empresarial, es posible la recogida de un amplio espectro de datos no de carácter personal y sujetos a propiedad, procedentes tanto de la actividad comercial de la empresa como de sus actividades internas, ya sean del desempeño de las cadenas de producción o, incluso, de la actividad de sus personas.

La explotación cabal de todas estas fuentes de datos es clave para la digitalización de la sociedad. Con las aplicaciones de macrodatos (o Big Data), analítica avanzada e Inteligencia Artificial será posible extraer claves y valor de estos datos de manera absolutamente inviable para las personas. En particular, ayudará a mejorar la posición competitiva de las empresas tanto industriales como de servicios, posibilitando la aparición de innovaciones resultantes de la explotación y análisis de los datos disponibles.

Pese a este impacto potencial incuestionable, la compartición de datos entre empresas y organizaciones no se está generalizando lo suficiente. Esto se debe a la falta de incentivos económicos (incluido el miedo a perder una ventaja competitiva), a la falta de confianza entre los operadores económicos en cuanto a que los datos se utilicen en consonancia con los acuerdos contractuales, a los desequilibrios en el poder de negociación, al miedo a la apropiación indebida de los datos por parte de terceros, y a la falta de claridad jurídica sobre quién puede usar los datos y con qué fin (por ejemplo, en el caso de los datos creados conjuntamente, en particular los datos relativos al internet de las cosas)<sup>1</sup>. En definitiva, aspectos relacionados con la gobernanza y la soberanía del dato y que son cruciales para generar la verdadera economía del dato, en la que el dato es un activo con un elevado valor y tiene que tratarse como tal.

<sup>1</sup> European Commission. A European Strategy for Data <https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1593073685620&uri=CELEX%3A52020DC0066>

## 1.1 Los datos y la digitalización

Los datos son un elemento clave en la digitalización de la sociedad y, por supuesto, en la digitalización de las empresas. Esto siempre ha sido así, pues la implantación de cualquier sistema de información se basa en la explotación de datos con que se le alimenta, y se justifica por la mayor efectividad que ofrecen en múltiples ámbitos: desde la toma de decisiones hasta la automatización de la producción.

La diferencia con la situación actual estriba, básicamente, en la facilidad cada vez mayor de capturar datos, lo que consecuentemente conlleva un aumento considerable del tamaño de los mismos, que demanda un aumento de la capacidad de cómputo para su tratamiento. Afortunadamente, la evolución tecnológica actual hace posible proveer esa capacidad extra de cómputo con creces, tanto desde el punto de vista de la algoritmia disponible, como de la capacidad bruta que es posible disponer para su utilización.

Es un hecho ampliamente aceptado que la productividad y, por lo tanto, la competitividad de una organización varía en función de su grado de digitalización, es decir, del grado en que tal organización es capaz de utilizar, generar y analizar datos relevantes a su operativa.

Es pues imperativo que diseñemos estrategias que permitan acelerar el uso que de los datos hace nuestro sector productivo, permitiendo que puedan extraer valor de los mismos, así como generar nuevas oportunidades de negocio derivadas del valor implícito en los datos capturados. Es aumentar drásticamente la competitividad o morir.

## 1.2 El Ciclo del Dato

Hoy día, existen grandes cantidades de datos con los que trabajar y extraer valor, tanto públicos como privados. En el caso público, el volumen de datos que se genera en muchos sectores, como la salud, es considerable. Aunque es cierto que este tipo de datos están sujetos a restricciones para su explotación relacionados con la preservación de la privacidad, lo cierto es que en general estos datos están desaprovechados.

Los motivos de tal desaprovechamiento son múltiples, y entre ellos destacaríamos el de no disponer de unas estructuras adecuadas para su almacenamiento, consulta, compartición y explotación, o no entender cómo realizar un tratamiento adecuado de los mismos, acorde con preceptos éticos y legales.

Eventualmente, esto supone una pérdida continua de oportunidades de innovación, tanto en servicios al ciudadano, como en el ámbito empresarial. Por ello, es importante incrementar la actividad en tecnologías que permiten la **captación, acceso, comunicación, procesamiento, explotación y compartición de los datos** de forma robusta, segura y eficiente para ayudar a la toma de decisiones en múltiples dominios de aplicación.

Estas áreas atacan problemas como la captación del dato - con precisión, fiabilidad y con densidad espacial y temporal adecuadas mediante dispositivos de fabricación escalable empleando nanotecnologías - (**Sensores Inteligentes, Sistemas Ciber Físicos**), su comunicación (**Ciber conectividad, IoT...**), su almacenamiento y procesamiento distribuido (**Cloud/Edge Computing, Blockchain...**) sobre dispositivos hardware nanoelectrónicos, hasta la analítica de esos datos en grandes cantidades (**Big Data Analytics**), potencialmente usando técnicas estadísticas y metaheurísticas (**Sistemas de Optimización**) o técnicas de aprendizaje automático (**Inteligencia Artificial**), con el fin de producir sistemas capaces de predecir comportamientos, simular entornos, o tomar decisiones de forma automática. Todo ello con una capa horizontal de buenas prácticas de **Ingeniería de Software** que aporta Calidad y Seguridad al producto desarrollado

Todas estas tecnologías, como las recogidas en general en el Plan España Digital 2025<sup>2</sup> (IoT y redes de comunicación 5G, las tecnologías para el tratamiento de datos masivos y bases de registro distribuido, la supercomputación, la computación difusa y en la nube, el edge computing, la realidad virtual, la ciberseguridad, la micro/nano electrónica o la propia Inteligencia Artificial) encajan de forma sinérgica en el ecosistema del dato, y su continuo desarrollo ha de asegurar la mejora constante en la generación, procesamiento y explotación de los datos. Como tecnologías digitales habilitadoras que son, su aplicación se extiende a múltiples dominios, tanto pertenecientes a los sectores productivo como a los sectores orientados a la sociedad.

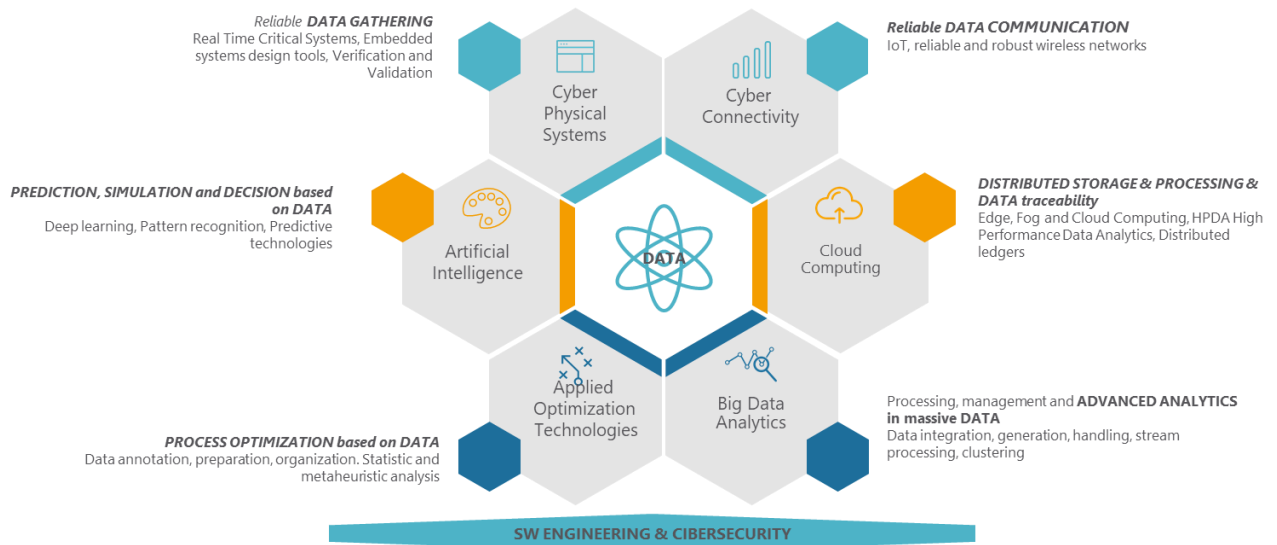


Figura 1. Habilitadores digitales que conforman el ciclo del Dato

### 1.3 Europa y los DataSpace

Los datos constituyen el centro de la digitalización. Siempre ha sido así, y así seguirá siendo. Desde siempre, las empresas que hoy en día se reconocen como “tecnológicas” se han centrado bien en la generación y captura de datos (sensores, dispositivos de entrada de datos, captura de frases de búsqueda, captura de visitas de páginas, etc.), bien en la construcción de herramientas software para gestionarlos (bases de datos, sistemas operativos, sistemas de comunicación, middleware, etc.), bien en la creación de software para explotarlos (ERPs, CRMs, visión artificial, robots, SCADA, software estadístico, *machine learning*, optimizadores, etc.). Es precisamente esta centralidad en los datos la que ha hecho posible que dichas “tecnológicas” capturen una buena parte de la capitalización de los mercados bursátiles.

En el ámbito europeo, hace tiempo que se ha reconocido el potencial existente en la abundancia de datos que son capturados y que ***es posible llegar a capturar***, más allá de los datos tradicionales que han sido objeto de los sistemas de información que se han implantado en las empresas y que son clave para mejorar la competitividad económica de Europa a través de su explotación. Esta concienciación a nivel europeo se está traduciendo en iniciativas que fomentan el desarrollo de lo que se denominan **DataSpaces europeos (Espacios de Datos europeos)** o también llamados *Infraestructuras de datos*.

<sup>2</sup> Plan España Digital 2025, julio 2020; [https://www.lamoncloa.gob.es/presidente/actividades/Documents/2020/230720-Espa%C3%B1aDigital\\_2025.pdf](https://www.lamoncloa.gob.es/presidente/actividades/Documents/2020/230720-Espa%C3%B1aDigital_2025.pdf)



El concepto de un **Espacio de Datos** es, en su núcleo, el de un ecosistema de datos “valiosos” capturados a nivel “regional” o “sectorial”, disponibles de algún modo para su procesamiento y análisis.

La Comisión Europea realiza un llamamiento específico en la Estrategia Europea de Datos para desplegar DataSpaces en nueve sectores estratégicos. Actualmente hay un gran interés por conocer qué son los DataSpaces y cómo pueden crearse. En general, un **DataSpace** es un concepto de gestión o integración de datos que se puede definir por varios principios fundamentales:

- Los *DataSpaces* no requieren de integración física de datos, pero soportan almacenamientos de datos distribuidas. No requiere un almacenamiento centralizado de datos, como por ejemplo los “DataLakes” o las plataformas de datos desarrollados en los últimos años.
- Los *DataSpaces* no requieren un esquema común. La integración se consigue a nivel semántico.
- Los *DataSpaces* permiten que haya superposiciones, es decir los DataSpaces pueden ser parte de otros DataSpaces.

Desde los estamentos europeos se pretende, a través de varias iniciativas, fomentar primero e interconectar después estos Espacios de Datos (DataSpaces) para permitir la explotación cruzada de los datos disponibles en todos ellos con el fin de fomentar que la siguiente ola de digitalización permita que Europa no se quede rezagada. Destacan iniciativas como GAIA-X, IDSA o EOSC, que son descritas en el [apartado 6.1](#)

#### 1.4 Políticas e instrumentos relacionados con el fomento de los DSES

En la “**Estrategia Anual de Crecimiento Sostenible 2021**”<sup>3</sup>, la Comisión Europea insta encarecidamente a los Estados miembros a incluir en sus planes de recuperación y resiliencia inversiones y reformas en los ámbitos que se enumeran a continuación: Activación, Renovación, Recarga y repostaje, Conexión, Modernización, Expansión, y Reciclaje y Perfeccionamiento Profesional.

En concreto, con referencia al ámbito Expansión, señala que “**la transición digital de la UE depende del aumento de sus capacidades de datos para fines industriales en la nube y de la capacidad de desarrollar los procesadores más potentes, vanguardistas y sostenibles**”.

Por su parte, en “**Una Estrategia Europea de Datos**”<sup>4</sup>, de febrero de 2020, se propone la creación de un espacio único europeo de datos que sirva para aumentar el uso y la demanda de datos, y de servicios y productos basados en datos en todo el mercado único, garantizando su acceso y su uso responsable.

Mediante las normas y mecanismos asociados se pretende garantizar que:

- Los datos puedan fluir;
- Se respeten plenamente las normas y valores europeos;
- Las normas de acceso a los datos y su uso sean justas, prácticas y claras.

<sup>3</sup> Estrategia anual de crecimiento sostenible 2021, septiembre 2020; <https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:52020DC0575&from=EN>

<sup>4</sup> “Una Estrategia Europea de Datos”, Comisión Europea, febrero 2020; <https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:52020DC0066&from=ES>

Los Espacios de Datos, centralizados o distribuidos, servirán para fomentar ecosistemas que creen nuevos productos y servicios basados en datos más accesibles. La Estrategia propone el desarrollo de Espacios de Datos en nueve ámbitos estratégicos. Los problemas a los que se dirige la estrategia son:

- Disponibilidad de datos: el valor de los datos reside en su uso y reutilización, sea G2B, B2B, B2G y G2G.
- Desequilibrios en el poder de mercado.
- Interoperabilidad de datos.
- Gobernanza de datos.
- Infraestructuras y tecnologías de datos.
- Empoderamiento de las personas para que ejerzan sus derechos.
- Cualificación y alfabetización en materia de datos.
- Ciberseguridad.

Las acciones de la Estrategia se basan en cuatro pilares:

- A. Marco de gobernanza intersectorial para el acceso a los datos y su utilización**, que sirva para el desarrollo de ecosistemas animados, dinámicos y vigorosos. Supone **(a)** poner en marcha un marco legislativo; **(b)** facilitar que se disponga de un mayor volumen de datos de calidad del sector público para su reutilización, en particular considerando su potencial para las pymes; **(c)** adoptar medidas legislativas sobre cuestiones que afecten a relaciones entre los actores en una economía ágil en el manejo de datos y **(d)** analizar medidas para establecer repositorios de datos a efectos del análisis de datos y el aprendizaje automático. Se contempla una norma relativa a los datos (2021) que sirva para fomentar, entre otras medidas, el intercambio de datos entre empresas y administraciones para el interés público. Principio: “tan abiertos como sea posible y tan cerrados como sea necesario”.
- B. Catalizadores: inversiones en datos y refuerzo de las capacidades e infraestructuras de Europa** para albergar, tratar y utilizar los datos, interoperabilidad. La Estrategia se basa en un ecosistema próspero de agentes privados para crear valor económico y social a partir de datos. Europa debe ofrecer un entorno que apoye la innovación basada en los datos y estimule la demanda de productos y servicios basados en los datos como un factor de producción importante.
  - Proyecto de gran impacto: creación de Espacios de Datos europeos comunes e interconexión de las infraestructuras en la nube, que incluyan: i) el despliegue de herramientas y plataformas de intercambio de datos; ii) la creación de marcos de confianza de datos; iii) la mejora de la disponibilidad, la calidad y la interoperabilidad de los datos. También abarcará el tratamiento de datos y las capacidades asociadas.
  - Permitir el acceso a unos servicios europeos en la nube competitivos, seguros y justos. Además, deben realizarse trabajos para conectar a las organizaciones activas en el lado de la demanda en los sectores público y privado con la nueva e innovadora oferta de servicios de tratamiento de datos (PaaS y SaaS).
  - Apoyo a los avances en materia de tecnologías de datos.
- C. Competencias: empoderar a las personas, invertir en cualificaciones y en pymes.**
  - Empoderamiento de las personas con relación a sus datos.
  - Inversiones en cualificaciones y alfabetización general en materia de datos.
  - Creación de capacidades específicas para las pymes.
- D. Espacios comunes europeos de datos en sectores estratégicos y en ámbitos de interés público**, es decir, en aquellos en los que el uso de datos tendrá un impacto sistémico en todo el ecosistema, y también en los ciudadanos. Esto debe conducir a la disponibilidad de grandes

repositorios de datos en tales sectores y ámbitos, junto con las herramientas y las infraestructuras técnicas necesarias para utilizar e intercambiar datos, así como unos mecanismos de gobernanza adecuados. Estos espacios deben complementarse con políticas que estimulen el uso de datos y la demanda de servicios enriquecidos con datos. La Comisión apoyará la creación de nueve espacios en varios ámbitos: fabricación, Pacto Verde, movilidad, salud, financiero, energía, agrario, administraciones públicas y cualificaciones.

La ambición de la Estrategia es que la UE se convierta en “la economía ágil en el manejo de los datos más atractiva, más segura y más dinámica del mundo, y empoderar así a Europa con datos para mejorar las decisiones que se toman y la vida de todos sus ciudadanos”.

La nueva Agenda Digital española, **Plan España Digital 2025**, está orientada a impulsar la Transformación Digital de España como una de las palancas fundamentales para relanzar el crecimiento económico, la reducción de la desigualdad, el aumento de la productividad y el aprovechamiento de todas las oportunidades que brindan estas nuevas tecnologías. El Plan recoge un conjunto de medidas, reformas e inversiones, articuladas en diez ejes estratégicos, alineados a las políticas digitales marcadas por la Comisión Europea para el nuevo periodo. En el noveno de estos diez ejes se indica: **“Favorecer el tránsito hacia una economía del dato, garantizando la seguridad y privacidad y aprovechando las oportunidades que ofrece la Inteligencia Artificial”**.

**El potencial de los datos, tanto en Europa como en España, está todavía por explotar porque (1) falta un marco regulatorio adecuado que ofrezca certidumbre al uso que todos los agentes pueden hacer de estos datos; (2) faltan marcos de cooperación adecuados para compartir datos entre sectores y entre estos y las Administraciones Públicas; y (3) insuficiente transparencia en la regulación sobre los derechos de propiedad en este ámbito.**

Algunos aspectos se deben abordar a escala europea y otros son responsabilidad nacional, como los relacionados con la creación de un ecosistema del dato y con mecanismos de colaboración público-privados.

En materia de Economía de datos e Inteligencia Artificial, el Plan fija los siguientes objetivos específicos:

- Convertir a España en un referente en la transformación hacia una Economía del Dato. *“Esta transformación necesita asimismo de una gestión responsable, que favorezca el fuerte compromiso del sector público para continuar incrementando la apertura de los datos públicos, la creación de mecanismos de colaboración fuerte entre el sector público y el privado, el impulso público a la compartición de datos y al desarrollo de proyectos faro para usar los datos tanto públicos como privados para el bien común y la existencia de empresas con gran experiencia y voluntad para desarrollar experiencias de colaboración público-privadas en este ámbito, como instrumentos decisivos para alcanzar el objetivo propuesto.”*
- Impulsar la Inteligencia Artificial como motor de innovación y crecimiento económico social, inclusivo y sostenible.
- Desarrollar un marco ético y jurídico para la IA basado en valores compartidos.
- Preparar a España para las transformaciones socioeconómicas que origina la IA.
- Fortalecer la competitividad a través de las actividades de I+D en el conjunto de las Tecnologías Habilitadoras Digitales (THD).

Y entre las medidas se encuentran:

- **Estrategia Nacional de Inteligencia Artificial.**
- **Oficina Del Dato y Chief Data Officer (CDO).** El Gobierno creará en la Administración General del Estado la Oficina del Dato que se encargará de diseñar y proponer estrategias que permitan poner a disposición de las empresas y la ciudadanía los datos públicos de las Administraciones. Entre sus tareas se incluirán las siguientes: el establecimiento de las medidas necesarias para impulsar la compartición y uso de los datos públicos y privados, **la creación de entornos para compartir de manera segura los datos de la Administración con las empresas y de estas con la Administración, así como de las empresas entre sí, de forma alineada y coordinada con las instituciones existentes que elaboran estadísticas oficiales.** También se encargará del seguimiento de las tendencias europeas en las estrategias europeas de datos y de cloud; **así como invertir en la generación de “lagos de datos” sectoriales y la propuesta de mecanismos de gobernanza.**
- **Consejo Asesor de Inteligencia Artificial**
- **ESTRATEGIA CLOUD: Espacios Compartidos Europeos del Dato.** España desempeñará un rol activo para formar parte de los espacios compartidos de la **European Cloud Federation** y para potenciar un espacio ibérico, junto a Portugal, que impulse el desarrollo de tecnologías avanzadas de computación de datos: HPC, Quantum computing o Edge computing, entre otras, con el objetivo de convertirse en un hub de conectividad y, consecuentemente, un potencial punto de concentración de infraestructuras de datos. Es irrenunciable apostar por la disponibilidad de datos y la innovación digital a nivel europeo, y desde España se trabajará para desarrollar alianzas en este sentido, impulsando las inversiones privadas en Data Centers y situando al país como un hub de negocios alrededor del espacio cloud europeo.

También, en la nueva **Estrategia Nacional de Inteligencia Artificial**<sup>5</sup> de diciembre de 2020, en su eje 3, “desarrollar plataformas de datos e infraestructuras tecnológicas para dar soporte a la IA” propone:

*La regulación deberá reforzar los compromisos para la protección, la gestión integral, la apertura y el acceso a la información de la que dispone el sector público y para que, de manera coordinada con el sector privado, se facilite la creación de nuevos servicios de valor añadido basados en datos. En esta dirección, y en línea con lo propuesto en la nueva Estrategia Europea de Datos que fomenta la creación de data pools sectoriales, es necesario investigar y desarrollar un marco de acción para instrumentar modelos de compartición de datos para incentivar la innovación y experimentación a distintos niveles:*

- **G2B (Government to Business): compartir datos entre el Gobierno y las empresas.**
- **G2G (Government to Government): compartir datos entre Administraciones Públicas.**
- **G2C (Government to Citizen): compartir datos entre el Gobierno y la ciudadanía.**

<sup>5</sup> Estrategia Nacional de Inteligencia Artificial, diciembre 2020;  
[https://portal.mineco.gob.es/RecursosNoticia/mineco/prensa/noticias/2020/201202\\_np\\_eniav.pdf](https://portal.mineco.gob.es/RecursosNoticia/mineco/prensa/noticias/2020/201202_np_eniav.pdf)

Y se establece como medidas, dentro de la Línea de Actuación “Impulsar las plataformas de datos”:

- **Medida 13. Creación de espacios compartidos de datos sectoriales e industriales** y repositorios descentralizados y accesibles, que faciliten la creación de servicios de valor añadido basados en infraestructuras de datos y la construcción de aplicaciones de IA seguras e innovadoras en un entorno de confianza.
- **Medida 16. Puesta en marcha del Proyecto Datos por el Bien Social**, basados en el uso de datos públicos de las Administraciones y de datos generados por la Ciudadanía para generar un retorno colectivo positivo (ej. Ciencia ciudadana, contribución de evidencia para políticas públicas, etc.).

En su eje 5, “potenciar el uso de la IA en la administración pública y en las misiones estratégicas nacionales” se indica:

*Crear repositorios de datos públicos que permitan el acceso en condiciones óptimas de seguridad, legalidad, integridad confidencialidad y protección de la privacidad de los ciudadanos para desarrollar nuevas aplicaciones y oportunidades, tanto para el sector público (sistemas en el ámbito de la gestión sanitaria, la educación, la seguridad, la transición ecológica, la gestión urbana o la movilidad sostenible) como para el sector privado.*

Y se incorporan varias medidas relacionadas en la línea de actuación 5.1, “aprovechar los beneficios para la administración del uso de la IA”:

- **Medida 21. Incorporar la IA en la administración pública** para mejorar la eficiencia en la relación con los ciudadanos, industria, empresas, y la sociedad en general y eliminar cuellos de botella administrativos.
- **Medida 22.** Poner en marcha un **laboratorio de innovación** para desarrollar nuevos servicios y las aplicaciones de la IA en la Administración Pública (GobTechLab). Colaborar con los programas de IA en los distintos Ministerios y organismos públicos, para garantizar la puesta en común de los programas y la difusión de las soluciones.
- **Medida 23. Competencias IA en la AAPP. Fomentar el uso de IA en la Administración**, desde el acceso a la función pública, la formación de los/as empleados/as públicos/as, hasta la definición de puestos de trabajo.
- **Medida 24.** Programa “**IA para una gestión pública basada en datos**”. Utilizar la IA para tener un conocimiento real de la sociedad y tomar decisiones en la previsión macroeconómica y la articulación de políticas públicas basada en datos fiables y actualizados, así como utilizar la IA para mejorar las políticas de empleo y de capacitación de los trabajadores.

El “**Plan de Recuperación, Transformación y Resiliencia – Plan España Puede**”<sup>6</sup> es un proyecto de país que traza la hoja de ruta para la modernización de la economía española, la recuperación del crecimiento económico y la creación de empleo, para la reconstrucción económica sólida, inclusiva y resiliente tras la crisis de la COVID, y para responder a los retos de la próxima década.

<sup>6</sup> Plan España Puede, Plan de Recuperación, Transformación y Resiliencia, octubre 2020;

[https://www.lamoncloa.gob.es/presidente/actividades/Documents/2020/07102020\\_PlanRecuperacion.pdf](https://www.lamoncloa.gob.es/presidente/actividades/Documents/2020/07102020_PlanRecuperacion.pdf)

El Plan comporta un importante volumen de inversión pública y privada en los próximos años, financiándose con los fondos del Plan Next Generation EU, el Fondo de Recuperación europeo. Este Fondo permitirá a España obtener financiación por hasta 140.000 millones de los cuales alrededor de **72.000 millones se desembolsarán en forma de transferencias** y el resto a través de préstamos.

El Plan **contempla diez políticas palanca**, en aquellas políticas prioritarias y que contribuyen al crecimiento sostenible e inclusivo, potenciando la descarbonización de la economía y la digitalización. La **colaboración público-privada constituye uno de los principios directores indispensables**, que sirve para aumentar la capacidad de inversión de los proyectos tractores y movilizar empresas y agentes sociales, llegando al conjunto del tejido productivo.

El Plan se estructura en torno a **cuatro ejes transversales** que vertebrarán la transformación del conjunto de la economía y que el Gobierno ha situado desde el principio en el centro de su estrategia de política económica: **la transición ecológica, la transformación digital, la igualdad de género y la cohesión social y territorial**. En el **eje de digitalización se indica que la ciberseguridad, la economía del dato, la inteligencia artificial y otras tecnologías digitales habilitadoras son vectores estratégicos para impulsar la segunda oleada de digitalización en España**. Hace referencia a la **Agenda España Digital 2025** como la hoja de ruta para guiar la acción en este terreno.

El **Programa Digital Europe**<sup>7</sup> tiene como objetivo impulsar las inversiones de la UE, los Estados miembros y la industria en las áreas clave de inteligencia artificial, informática avanzada y manejo de datos, ciberseguridad y las habilidades digitales necesarias para implementarlas.

Busca conectar a las empresas, las administraciones públicas y los ciudadanos con las últimas tecnologías y recursos. **Contempla establecer y hacer accesibles Espacios de Datos e instalaciones de prueba y experimentación en toda Europa para la inteligencia artificial en las áreas de salud, medio ambiente / clima, movilidad, fabricación y energía.**

Y, específicamente, implementar plataformas digitales urbanas abiertas, interoperables y confiables adaptadas a las necesidades de las comunidades, que ofrezcan un fácil acceso estandarizado a nuevos conjuntos de datos y el despliegue a gran escala de servicios impulsados por IA en Smart Energy, Smart Mobility, gestión de residuos y recursos secundarios, industria y (re)fabricación, salud y gobierno electrónico.

Por último, a nivel europeo se está trabajando en la **Ley de Gobernanza de Datos** (Data Governance Act COM/2020/767) y en el Reglamento de Datos (Data Act). El objetivo de este instrumento es ampliar la disponibilidad de datos con miras a su utilización, mediante el aumento de la confianza en los intermediarios de datos y el refuerzo de los mecanismos para el intercambio de datos en el conjunto de la UE.

El instrumento aborda las situaciones que se exponen a continuación:

- La cesión de datos del sector público para su reutilización, en los casos en que esos datos estén sujetos a derechos de terceros<sup>8</sup>.

<sup>7</sup> Página del Digital Europe Programme; <https://ec.europa.eu/digital-single-market/en/europe-investing-digital-digital-europe-programme>

<sup>8</sup> Las expresiones «datos cuya utilización depende de derechos de terceros» o «datos sujetos a derechos de terceros» se refieren a datos que puedan estar sujetos a la legislación sobre protección de datos o sobre

- El intercambio de datos entre empresas a cambio de algún tipo de remuneración.
- La cesión de datos personales con ayuda de un «intermediario de datos personales», cuya labor consistirá en ayudar a los particulares a ejercer los derechos que les confiere el Reglamento General de Protección de Datos (RGPD).
- La cesión de datos con fines altruistas.

*ARTIFICIAL INTELLIGENCE – Exploiting the potential of data and putting AI to work. The Digital Europe programme aim is to develop capacity in AI in Europe, in line with the Communication on AI for Europe and the Coordinated Plan on AI. **To this end, the first two years of the programme will focus on developing an infrastructure which offers businesses and the public sector access to AI tools and components and data resources, as well as reference testing and experimentation facilities in some prioritised application sectors.** Particular attention will be paid to ensuring that fundamental rights and ethics requirements, such as privacy, diversity, non-discrimination, accessibility, societal and environmental wellbeing, are met when developing and deploying AI technology. Moreover, transparent implementation of AI should ensure that basic legal safeguards for citizens such as due process and Good administration are maintained.*

*Actions will focus on three main aspects:*

- **establishing EU-wide common, interoperable data spaces providing access to data for AI. Common European data spaces are a means of organising access and use of data from both the public and private sector in a marketconforming, lawful, ethical and trusted manner;**
- *developing world-class large-scale reference Testing and Experimentation Facilities (TEF) for AI hardware, software, components, systems and solutions, and underlying resources (data, computing, cloud) in a number of sectors;*
- *and scaling up the Common European AI platform, building on the existing AI-on-demand platform.*

**Data spaces include:**

- i. IT systems (digital industrial and personal data platforms);**
- ii. domain-specific data governance frameworks as articulations of the overall technical governance framework;**
- iii. standards, including semantic standards and interoperability protocols – both domain-specific and cross-cutting;**
- iv. competitive and seamless Access and use of cloud infrastructures based on the rollout of cross-border cloud federations.**

**Digital Europe Programme<sup>9</sup>**

propiedad intelectual, o que puedan contener secretos comerciales u otra información sensible a efectos comerciales.

<sup>9</sup> <https://ec.europa.eu/digital-single-market/en/europe-investing-digital-digital-europe-programme>

## 2 Espacios de Datos como elementos clave para la explotación de datos

*Un Espacio de Datos es una estructura que proporciona confianza y seguridad para la compartición voluntaria de datos entre diversos agentes de manera homogénea a través de mecanismos combinados de gobernanza, organizativos, legales y técnicos, que facilitan la interoperabilidad para acceder a los datos o transferirlos, y que posibilitan su reutilización eficiente y legítima en un contexto de soberanía y control para las partes sobre sus propios datos.*

Un **Espacio de Datos** se define como una infraestructura descentralizada para compartir e intercambiar datos confiables en ecosistemas de datos basados en principios comúnmente acordados. Los Espacios de Datos:

- Requieren componentes básicos, como plataformas de datos, que brinden soporte para compartir e intercambiar datos de manera efectiva, así como para la ingeniería y el despliegue de capacidades de procesamiento e intercambio de datos.
- Requieren componentes básicos como los mercados de datos, donde los proveedores de datos pueden ofrecer y los consumidores de datos pueden solicitar datos, así como aplicaciones de procesamiento de datos;
- Requieren soberanía de datos, es decir, la capacidad de cada parte interesada para tomar decisiones sobre cómo se estructuran, construyen y gestionan los procesos digitales, la infraestructura y los flujos de datos, sobre la base de un esquema de gobernanza apropiado que permita especificación de términos y condiciones.

Un Espacio de Datos se diseña teniendo en cuenta 4 principios:

- **Principio de diseño 1. La soberanía de los datos:** es la capacidad de una persona o entidad corporativa para la autodeterminación exclusiva con respecto a sus bienes de datos económicos. Este es un concepto innovador y transformación que subyace en los Espacios de Datos.
- **Principio de diseño 2. Nivel de juego con los datos:** implica que los nuevos entrantes no se enfrentan a barreras de entrada debido a situaciones monopolística, cuando existe un nivel de juego de datos, los participantes compiten por la calidad del servicio y no por la cantidad de datos que controlan. Que exista un nivel de juego de datos es una condición fundamental para crear una economía de intercambio justa de datos.
- **Principio de diseño 3. Infraestructura blanda descentralizada:** la infraestructura de intercambio de datos no es una infraestructura IT centralizada y monolítica. En cambio, es una colección de facto de implementaciones interoperables de Espacios de Datos que cumplen con un conjunto unificado de acuerdos en todas las disciplinas: funcional, técnica, operativa, legal y económica. Una 'infraestructura blanda' es simplemente invisible y está formada por acuerdos. El principio de soberanía de datos sigue los requisitos funcionales y no funcionales de interoperabilidad, portabilidad, capacidad de búsqueda, seguridad, privacidad y confiabilidad.
- **Principio de diseño 4. Gobernanza público-privada:** para el diseño, la creación y el mantenimiento del campo de juego de los datos, es esencial una gobernanza sólida. Todas las partes interesadas deben sentirse representadas y comprometidas. Estos incluyen usuarios (personas o empresas) o proveedores de servicios de datos, así como sus socios tecnológicos y profesionales.



Los Espacios de Datos son, por tanto, un elemento clave para compartir y posibilitar el acceso a los datos, pero deben estar complementados por infraestructuras y tecnologías de explotación de datos que permitan extraer el valor a los datos. Estas infraestructuras se pueden identificar como **Espacios o nodos de experimentación y explotación de datos (Espacios de Innovación con Datos) y conforman el conjunto de tecnologías en torno al ciclo completo del dato (acceso, captación, comunicación, protección, almacenamiento, procesamiento, análisis, visualización), permitiendo la creación de aplicaciones inteligentes que ayuden a la toma de decisiones por dominio o para múltiples dominios con las que poder extraer valor de los datos.**

Un **Espacios de experimentación y explotación de datos (Data Innovation Spaces)** se construye dentro de un marco de gobernanza y siguiendo una arquitectura para la plataforma de explotación. Además, provee una serie de guías para la prestación de servicios en cualquiera de las fases del ciclo del dato (captación, comunicación, protección, almacenamiento, procesamiento, análisis, visualización).

Así, los Espacios de Innovación con Datos se apoyan en cinco pilares fundamentales:

1. **Gobernanza:** se entiende como el conjunto de entidades que participan, sus roles, atribuciones y pautas de funcionamiento.
2. **Infraestructura y plataforma:** conjunto de recursos de computación adecuadamente estructurados y dimensionados, junto con un entorno (software) operativo de explotación (plataforma) cuyos objetivos principales son
  - Permitir el almacenamiento eficiente de grandes cantidades de datos para facilitar cómputos efectivos sobre los mismos
  - Permitir la especificación y lanzamiento fácil de trabajos de análisis sobre los datos disponibles a través de la plataforma
  - Implementar mecanismos de acceso seguro a los datos, así como la especificación de políticas flexibles de acceso y uso, acordes con la voluntad del dueño de los datos
  - El establecimiento de mecanismos y servicios que permiten el acceso federado a los recursos de otros *DSES* con el fin de aumentar la capacidad de cómputo o la cantidad y calidad de los datos disponibles.
3. **Datos:** información accedida a través del Espacio de Datos o directamente a través de conexiones con empresas como organismos del sector público. Estos datos constituyen una valiosa materia prima para desarrollar servicios digitales e innovadores. Los datos sean de naturaleza abierta o privada, de libre acceso o sujetos a condiciones específicas, se almacenarán en la infraestructura, con una durabilidad que dependerá de su naturaleza en origen y de la voluntad del dueño de los datos. Una vez almacenados, los datos podrán ser accedidos o consumidos bajo demanda, apoyándose en los diversos mecanismos provistos por la plataforma, como conectores, APIs o volúmenes seguros.

Es de vital importancia resaltar que cualquier acceso a cualquier dataset estará sometido a la política establecida por el dueño del dataset, delegando en los mecanismos de la plataforma el cumplimiento de las distintas políticas de gobernanza, soberanía, protección y compartición de datos que sean necesarias. Con este fin, se aprovecharán los distintos modelos de referencia internacionales en la materia. Cabe destacar que, a nivel europeo, la Directiva 2003/98/CE, relativa a la reutilización de la información del sector público<sup>10</sup>, ha creado un marco a escala de la UE para fomentar el uso transfronterizo de datos financiados con fondos públicos y contribuir al desarrollo de servicios y productos de datos paneuropeos.

<sup>10</sup> [Directiva 2003/98/CE del Parlamento Europeo y del Consejo, de 17 de noviembre de 2003, relativa a la reutilización de la información del sector público.](#)

4. **Algoritmos y servicios:** engloba Servicios, aplicaciones, técnicas y algoritmos que pueden ejecutarse sobre la Infraestructura y que permiten extraer valor de los datos de múltiples formas. Se asegura un compromiso de privacidad en la explotación de los datos respetuosos con una gobernanza dirigida por los derechos de los ciudadanos que puedan verse afectados por los mismos.
5. **Personas:** equipo que proporciona un amplio conocimiento en cada una de las disciplinas anteriores, capaces de investigar en ellas y asesorar a las empresas y la administración pública sobre su uso.

***Conjuntamente, los Espacios de Datos y los Espacios de Innovación con Datos, configuran los llamados Espacios de Compartición y Explotación de Datos (DSES), que permiten acelerar la digitalización de nuestros sectores productivos poniendo a disposición de empresas industriales y de servicios, de las administraciones públicas y centros de investigación, un entorno donde puedan encontrar una gran cantidad de datos de valor sobre los que diseñar, y donde puedan experimentar en el diseño y construcción de experimentos sobre dichos datos que ayuden a mejorar procesos industriales y servicios, así como a crear nuevas aplicaciones y servicios software que ayuden a la industria a acelerar su grado de digitalización.***

Esta propuesta persigue impulsar de forma decisiva los procesos de digitalización y evolución tecnológica a la industria y empresas españolas, permitiéndoles el acceso a recursos computacionales para la explotación de sus datos en solitario o en conjunto con los datos disponibles en los Espacios de Datos, a través de empresas que les ayuden en la puesta en marcha de procesos y servicios de explotación de dichos datos.

En este aspecto, la puesta en marcha de los DSES ayudará a la creación de ecosistemas que giren en torno al Dato, facilitando el desarrollo, prototipado y despliegue de técnicas de análisis y explotación de datos en cualquier dominio de la economía y la sociedad. En segundo lugar, permitirá poner en valor de un modo más efectivo muchos de los esfuerzos en materia de innovación e investigación que se realizan en el seno de centros tecnológicos y Universidades, potenciando la transferencia de éstos al tejido productivo usando los Espacios de Datos como conducto.

Con este enfoque de valor, un Espacio de compartición y explotación de Datos se apoya en las siguientes características:

1. ***Relación de confianza con diversos agentes productores de datos*** de alto valor para su almacenamiento y protección en el DSES, posibilitando así la innovación por parte de diferentes agentes: empresas, organismos de investigación y administraciones públicas.
2. Una ***infraestructura de almacenamiento y procesado de datos eficiente y escalable.***
3. Un ***sistema de gestión que garantice el cumplimiento de las políticas de uso y acceso*** a los datos estipuladas por sus propietarios.
4. Un ***sistema de compartición y explotación de los datos que permita el acceso a los mismos*** por parte de una diversidad de agentes, pero enfocado principalmente a las empresas y las administraciones públicas. De este modo estas últimas también podrán aumentar su productividad y la calidad del servicio que ofrecen a los ciudadanos.
5. Un ***sistema de planteamiento de experimentos ampliamente accesible*** con diversos grados de explotación.
6. Un ***sistema de “exportación” de los resultados/procedimientos*** que permita a empresas y administraciones portar procedimientos derivados de la explotación de dichos datos a sus propios sistemas de explotación.

### 3 Servicios habilitados por los Espacios de Compartición y Explotación Datos

Los Espacios de Compartición y Explotación de Datos (*DSES*) permiten **la explotación de datos de forma profesional, en el que las empresas, administraciones y organismos de investigación puedan disponer de un entorno tecnológico en que probar, desplegar y operar aplicaciones y servicios con robustez seguridad y confiabilidad.**

Este entorno de explotación de datos ofrece:

- Acceso a una plataforma gobernando infraestructuras de computación, y ofreciendo herramientas y tecnologías base para desarrollar, probar y desplegar aplicaciones basadas en la explotación de datos de fuentes diversas accesibles desde la propia plataforma.
- Almacenamiento seguro de datasets privativos para el desarrollo de aplicaciones y servicios de aplicación particular a una empresa o sector.
- Almacenamiento y acceso a datasets disponibles para su compartición con diversos grados de protección y con diferentes políticas de uso, para su inclusión en el desarrollo de aplicaciones y servicios de valor añadido, tanto para uso privativo de empresas, como para el desarrollo de servicios al ciudadano por parte de las administraciones públicas.
- Diversos mecanismos y políticas de preparación de datos: anonimización, curado, ...
- Puesta a disposición pública de Datasets, incluyendo mecanismos de acceso a datasets disponibles a través de otros *DSES* federados.
- Colaboración con proveedores de aplicaciones y servicios accesibles/integrables con el *DSES*.
- Canal de comercialización (Marketplace) para empresas tecnológicas.
- Otros servicios, como diccionarios, catálogos, *app stores* analíticos, etc.

Ofrecer facilidades para acceder a datasets de origen público ayuda a fomentar nuevos tipos de negocios basados en la explotación de dichos datos. Bien con su consumo periódico, bien a través de la obtención de nuevos descubrimientos consecuencia del análisis de los mismos. En conjunto, se plantea como una herramienta fundamental en el arsenal que se pone a disposición del sector productivo.

En paralelo un *DSES* también se puede configurar como un espacio público de experimentación, siendo una plataforma de servicio público fundamental para provocar que las empresas puedan salvar el enorme gap en digitalización (y, por consiguiente, en productividad/ competitividad) que tienen con respecto a sus competidores en otros países que ya se encuentran más avanzados en la explotación de datos propios y ajenos. Dentro de esta faceta de experimentación se ofrece, directa o a través de otras organizaciones de su entorno, el siguiente conjunto de servicios y funcionalidades:

- **Validación Tecnológica**
  - Acceso a infraestructuras, herramientas y tecnologías para desarrollar y probar aplicaciones basadas en el análisis de datos e Inteligencia Artificial
  - Acceso a datasets para el desarrollo de experimentos.
    - Intermediar entre los proveedores de datos y sus posibles consumidores.
  - Generación de datos sintéticos con los requerimientos de los experimentos.
  - Asesoramiento a empresas que buscan tecnologías innovadoras para incorporar a su cartera, mediante la selección y prueba de las mismas.
  - Desarrollo de estudios de viabilidad tecnológica para apoyar la transformación de ideas innovadoras en conceptos demostrables.

- Ayuda al diseño de prototipos para explorar ideas y tecnologías emergentes antes de entrar en producción considerando también las oportunidades potenciales que ofrece la producción en series pequeñas.
  - Apoyo en la realización de pruebas de concepto que demuestren la viabilidad de una idea o proyecto a través de su implementación reducida, temporal o provisional
  - Desarrollo de productos mínimos viables que puedan validarse con datos reales y/o en un entorno relevante.
  - Realización de trabajos de I+D específica: Aplicación de innovación tecnológica para desarrollar nuevos productos y servicios basados en datos o mejorar los existentes.
  - Certificación de producto basado en datos: Apoyo en la certificación de las pruebas de rendimiento y control de calidad.
- **Asesoramiento en manejo del Dato**
    - Diagnóstico de madurez de uso de datos: Evaluación de la preparación de la empresa para la adopción de las tecnologías basadas en datos para enfrentar los desafíos de los próximos años.
    - Selección de tecnologías para explotación de datos: Asesoramiento a empresas que buscan tecnologías innovadoras para incorporar a su cartera
    - Plan estratégico de explotación de datos: Apoyo en la definición de su visión y estrategia, tanto a nuevas empresas, como a PYMEs y grandes corporaciones que requieren nuevas ideas para seguir siendo relevantes y competitivas en el mercado
    - Aspectos legales y éticos de la explotación de datos y la Inteligencia Artificial: Asesoramiento especializado en aspectos legales y éticos de la IA y el manejo de datos.
    - Protección y privacidad de datos: Asesoramiento en técnicas y regulaciones de protección de datos y privacidad.
    - Compartición de datos: Técnicas y procedimientos para la compartición de datos de forma segura y fiable.
- **Formación.**
    - Formación técnica: Capacitación mediante formación en tecnologías basadas en datos.
    - Formación en Negocio: Formación para la capacitación y el desarrollo en habilidades empresariales y emprendimiento y el desarrollo de nuevos modelos de negocio e internacionalización.
    - Formación en aspectos legales: Formación en tratamiento, protección, privacidad y otros aspectos legales en torno a los datos.
- **Sensibilización y financiación**
    - Observatorio Tecnológico y de Mercado: Información actualizada sobre las tendencias en el mercado y evaluación del potencial del mercado de nuevos productos y servicios validando modelos de negocio.
    - Creación de comunidad. Creación de una comunidad en torno a las tecnologías de explotación de datos que conecte a todos los miembros del ecosistema de innovación de la Comunitat Valenciana
    - Demostración de productos propios o de terceros: Workshops de productos frente a potenciales clientes
    - Asesoramiento en búsqueda de financiación: Facilitación del acceso a diferentes fuentes de financiación (UE, nacional, regional y privada) con el objetivo de lograr una combinación efectiva de fondos.

## 4 Casos de uso de un Espacio de Compartición y Explotación de Datos

Se presentan a continuación, a modo ilustrativo una serie de escenarios de aplicabilidad para un *DSES*. Primero se plantean una serie de casos de uso simples, definidos de una manera general. A continuación, se enumera una serie de ejemplos basados en experiencia de negocio que comprenderían o estarían asociados a al menos uno de los casos de uso simples previamente definidos.

Estos casos de uso complementan y ejemplifican cuáles son y cómo se usarían algunos de los servicios y características de un *DSES*.

### 4.1 Casos de uso simples

A continuación, se plantea una lista inicial de casos de uso simples de *DSES*. Nótese que, en los siguientes casos, se utilizará empresa o institución de manera intercambiable.

Se identifican los siguientes casos de uso:

1. Empresa con necesidad de consumir un servicio de tipo software ofertado en el *DSES*. Algunos ejemplos o tipos de servicio podrían ser los siguientes:
  - *Radiatus*<sup>11</sup> (entrenamiento o ejecución de modelos)
  - Modelos concretos ofertados como servicio (p.ej., salud, segmentación de imágenes, etc.)
2. Empresa que necesita acceder a un dataset alojado en el Espacio de Datos. Esto requiere usar, al menos, los siguientes elementos:
  - Catálogo de datasets, que proporciona la suficiente información como para ayudar y facilitar las decisiones en cuanto a su uso o consumo por parte del posible usuario.
  - Limitaciones de visibilidad de los datos del dataset. Al decir que los datos no son visibles para el usuario nos referimos a cuando al usuario solo se le facilitan una serie de metadatos describiendo la estructura de dicho dataset. Esto le permite preparar su análisis, pero no puede hacer una inspección visual de los datos.
  - Herramienta o plataforma que permite la manipulación o trabajo con el dataset sin que éste salga del *DSES*, e.g., *Radiatus*.
3. Empresa que necesita recursos de computación adicionales para ser utilizados de manera independiente para ejecución de experimentos u otros usos aceptados.
4. Empresa que requiere federar recursos de computación con los suyos propios para ejecución de experimentos u otros usos aceptados.
5. Empresa que requiere federar un servicio en sus instalaciones con una instancia del mismo servicio, desplegada en el *DSES*.
6. Empresa que requiere realizar un aprendizaje federado consumiendo datasets del *DSES*.
7. Colaboración con otros *DSES* a nivel nacional o europeo. El objetivo de esta colaboración podría ser la creación de:
  - Conjuntos de recursos compartidos orientados a su consumo en modalidad IaaS, así como para el despliegue de servicios sobre los mismos.
  - Un catálogo de servicios conjunto para los *DSES* que podrán ser desplegados en cualquiera de ellos, utilizando recursos de un solo *DSES* o de varios simultáneamente.
8. Empresa que quiere depositar, compartir y/o monetizar (u obtener alguna compensación) un dataset a través del *DSES*. Para ello, el *DSES* deberá dar:

<sup>11</sup> <https://radiatus.iti.es/>

- Garantías de privacidad y de no fuga de datos.
  - Posibilidad de ejecución segura, p.ej., enclaves. Esto puede ser especialmente relevante para acoger datos de reducida visibilidad para el usuario consumidor.
  - Funcionalidad de caja fuerte.
9. Empresa que quiere depositar, compartir y/o monetizar (u obtener alguna compensación) modelos de inteligencia artificial a través del *DSES*.
  10. Empresa que quiere compartir y/o monetizar (u obtener alguna compensación) recursos de computación a través del *DSES*.
  11. Empresa que quiere compartir y/o monetizar (u obtener alguna compensación) un servicio a través del *DSES*.

## 4.2 Posibles ejemplos o situaciones de uso

En este apartado se definen posibles ejemplos extraídos de la experiencia en el mercado y en los que un *DSES* podría ser utilizado o relevante. También se incluyen algunos posibles ejemplos de SaaS que podrían ser ofertados en este marco.

### 1. Brokering & Safe Experimentation (asociado a los casos de uso 01, 02 y 08)

En este marco participarían una empresa tecnológica que tiene una idea de negocio (e.g., predicción de calidad de soldaduras), pero que no dispone de datos para poder generar o evolucionar modelos o algoritmos que le permitan avanzar su idea. Por otra parte, interviene una empresa que tiene los datos y que puede estar interesada en el producto de la primera empresa.

- *¿Cómo puede ayudar un DSES a resolver el problema?*

Un *DSES* ejerce de bróker de datos y de espacio de experimentación, ofreciendo almacenamiento donde el proveedor de datos puede depositar sus datos de manera segura y la empresa tecnológica puede disponer de potencia de cómputo y herramientas donde realizar sus análisis y modelado. Los datos podrían almacenarse como un nuevo dataset o existir sólo durante la vida del experimento o colaboración, siendo borrados después.

### 2. Sectorial Common Data Space (asociado a casos de uso 01, 02, 03, 06 o 07)

Se trata de la necesidad de crear un Espacio de Datos específico de un sector, cubriendo toda la cadena de valor (e.g. Manufacturing, Transporte, Salud, ...), para facilitar el análisis de datos y mejorar el proceso de toma de decisiones.

- *¿Cómo puede ayudar un DSES a resolver el problema?*

Un *DSES* puede ser desplegado para cubrir las necesidades específicas de un conjunto acotado de actores, como son los de una cadena de valor. El *DSES* sería, por tanto, desplegado para este sector y teniendo en cuenta las infraestructuras disponibles y la tipología de datos. El despliegue puede contemplar la federación de recursos de computación. Igualmente, los datos podrían ser complementados por otros hospedados en el *DSES*, ofreciendo la posibilidad de hacer un aprendizaje federado o de consumir datasets adicionales.

### 3. Federación de datasets distribuidos (asociado a casos de uso 01, 02, 06 y 08)

En este caso se contempla la existencia de distintos bancos de imágenes (por ejemplo, imágenes DICOM (pesadas)) distribuidos. Estas imágenes no están anonimizadas y no pueden ser usadas por científicos de datos, e.g., para mejorar modelos de predicción de cáncer en base a imágenes, por cuestiones de privacidad.

- *¿Cómo puede ayudar un DSES a resolver el problema?*

Un DSES podría ofrecer sus servicios orientados a la anonimización, evaluación de completitud u homogeneización de datos, entre otros. Igualmente, los datasets se hospedarían en el DSES de manera temporal o definitiva para facilitar esos análisis, ofreciendo funcionalidades de análisis seguro o garantías de que no se pueden producir fugas de datos.

### 4. Proof of concept with Open Data (asociado a casos de uso 01 y 02)

Se plantea el caso donde una empresa de análisis de datos puede necesitar utilizar datasets abiertos (p.ej., copernicus) o de acceso libre para validar una posible aplicación.

- *¿Cómo puede ayudar un DSES a resolver el problema?*

Un DSES podría ofrecer conectores/plugins estándar a datasets públicos que facilitaran su uso a terceros. El DSES lista aquellos open datasets para los que se ofrecen conectores, así como instrucciones para su uso. Igualmente, se ofrecen datasets propios de manera gratuita o remunerada.

### 5. DataThon (Asociado a los casos de uso 01 y 02)

Ofrecer un marco en el cual se puedan desarrollar eventos de tipo DataThon. En un DataThon, una o varias empresas proporcionan datasets y plantean una serie de retos a resolver. El público objetivo son investigadores, start-ups, u otros participantes de un ecosistema local que, en este marco, se ponen en contacto y a los que se invita a colaborar dentro del marco de dichos retos.

- *¿Cómo puede ayudar un DSES a resolver el problema?*

Un DSES puede ser el marco en el que se desarrollen estos DataThones. Esto implica facilitar la ingesta y acceso a los distintos datasets así como la creación de tenants y espacios de ejecución independientes para los distintos participantes o el uso de planificadores de recursos.

### 6. Product Certification (Asociado a los casos de uso 01, 02 y 11)

Se puede dar el caso de empresa con aplicaciones que requieren una validación de calidad o precisión (e.g., modelos de machine learning). Esta validación es compleja y puede requerir solicitar los servicios de un tercero además de disponer de datasets con los que realizar dicha validación.

- *¿Cómo puede ayudar un DSES a resolver el problema?*

Un DSES ofrece el servicio para certificar la robustez de un modelo/aplicación basada en datos, generando el entorno de pruebas necesario (infraestructura, herramientas y datos) para medir la calidad y validar el producto.

### 7. Data Incubation (Asociado a los casos de uso 01, 02 y 11)

De cara a la evaluación de modelos, de ejecución de pruebas de concepto y algunos ejercicios de investigación, puede ser necesario crear datasets sintéticos

- *¿Cómo puede ayudar un DSES a resolver el problema?*

Un servicio de creación de datasets sintéticos personalizable es un servicio incluido en el catálogo de servicios de un DSES. Igualmente, el entrenamiento y la validación se puede realizar con datasets hospedados en el propio DSES.

### 8. Data Broker / Data or Other Marketplaces (Asociado a casos de uso 08, 09, 10 y 11)

Presenta el caso en el que una empresa pueden querer ceder o compartir datos temporalmente para obtener un rendimiento económico a partir de su uso por parte de terceros. Esto es igualmente aplicable a algoritmos y modelos de Machine Learning / Inteligencia artificial, servicios o recursos de computación.

- *¿Cómo puede ayudar un DSES a resolver el problema?*

Un DSES puede servir de lugar para hospedar estos recursos, ofreciendo las herramientas de monetización y accountability necesarias para poder llevarlo a cabo.

### 9. Composed services (Asociado a casos de uso 01 y 11)

Una empresa necesita componer una cadena de servicios para realizar una tarea específica. Para ello, puede utilizar uno o más servicios o modelos como módulos independientes que conforman dicha cadena de servicio.

- *¿Cómo puede ayudar un DSES a resolver el problema?*

Un DSES puede ser de utilidad en este caso ofreciendo, primero, parte o todos los servicios necesarios para conformar la cadena de servicio.



## 5 Retos legales, políticos y regulatorios; de mercado y técnicos de los Espacios de Datos

Existen tres tipos de retos para poner en marcha este tipo de iniciativas:

1. Los retos legales, políticos y regulatorios
2. Los retos de mercado
3. Los retos técnicos

### 5.1 Los retos legales, políticos y regulatorios

Desde la perspectiva legal, política y regulatoria se percibe un conjunto de retos para potenciar el despliegue de los ecosistemas y Espacios de Datos:

- Regulaciones con relación al uso y reutilización de determinados tipos de datos: personales, sensibles, en posesión de las entidades públicas, mercado...
- Regulaciones con relación a determinados agentes y roles en los ecosistemas y Espacios de Datos.
- Desarrollo de los derechos de las personas, empresas y organizaciones de todo tipo con relación a sus datos o a los datos que poseen.
- Falta de estándares en la materia y las certificaciones asociadas.

En los últimos años se ha ido desarrollando una serie de regulaciones con relación a determinados tipos de datos. Así, el Reglamento General de Protección de Datos (2016/679) y, para el caso de España, la ley de 6 de diciembre de 2018 de Protección de Datos Personales y garantía de los derechos digitales, que lo complementa, regula la protección de las personas físicas en lo que respecta al tratamiento de sus datos personales y a la libre circulación de estos datos. Entró en vigor el 25 de mayo de 2016 y fue de aplicación el 25 de mayo de 2018. Por otro lado, la Directiva europea de Datos Abiertos (2019/1024) de 20 de junio de 2019 regula el tratamiento de los datos abiertos y la reutilización de la información del sector público. Actualmente (enero de 2021), se está ultimando la *Data Governance Act* para facilitar el intercambio y generar confianza en el intercambio de los datos sujetos a determinados derechos que están disponibles en las administraciones públicas y, por supuesto, sin modificar sustancialmente los derechos sobre los mismos.

La Unión Europea tiene previsto desarrollar otras políticas para fortalecer mecanismos de intercambio de datos en Europa, tal y como establece en “Una Estrategia Europea de Datos” de febrero de 2020. Además de las políticas generales o transversales que se generen bajo este paraguas, existen otras políticas sectoriales desplegadas o previstas con relación a industrias concretas, como la del automóvil, los sistemas de pago, los contadores de energía, de salud, etc.

Estas regulaciones y políticas determinan en qué condiciones y cómo se pueden compartir determinados tipos de datos y abren también el ámbito a generar y regular nuevos roles, como representantes o intermediarios de datos, que pueden jugar un papel determinante en el desarrollo de los Espacios de Datos y los DSES. También se regulan nuevos modos de prestación de datos, como la prestación altruista de datos no personales para fines sociales, científicos, etc. Por último, otro reto en esta categoría tiene relación con la certificación de determinados sistemas o agentes en los ecosistemas de datos para lo que será preciso determinar las normas y requisitos asociados.

Desde el punto de vista político, el mayor reto al que nos enfrentamos es el del convencimiento de que se pueden establecer políticas de buen uso de aquellos datos que impacten de lleno en aspectos de privacidad. El otro aspecto que será fundamental es el establecimiento de mecanismos de compensación adecuados para los propietarios de los datos por el uso y explotación de los mismos. Entendemos que estos mecanismos afectarán por igual a datos de procedencia pública como privada.

Finalmente, será necesario convencer de que los mecanismos tecnológicos provistos por el *DSES* son suficientes para asegurar la correcta gestión de las políticas relacionadas con la soberanía del dato incluyendo la definición de aquellas condiciones y términos que constituyen las políticas que el dueño de los datos crea oportunas y controlar el uso de los datos de acuerdo con las mismas.

## 5.2 Los retos de mercado/financieros

Aunque el concepto de *DSES* no hace referencia exclusivamente a estructuras económicas pues puede aplicarse a servicios vinculados con las administraciones públicas y entidades con fines sociales, existen desde la perspectiva de mercado determinados problemas y retos que se precisan superar para potenciar la competitividad e innovación basada en datos:

- Gobernanza de datos.
- Desequilibrios en el poder del mercado de datos.
- Desconocimiento e insuficiente cantidad de datos disponibles para innovación basada en datos. Dificultad para conocer a priori el valor y calidad de los datos existentes en otras organizaciones.
- Falta de confianza e incentivos para favorecer el acceso o intercambio de datos en contextos B2B.
- Insuficiente desarrollo de nuevos actores claves en ecosistemas de datos, como intermediarios de datos, entidades de certificación, etc.
- Competencias y cultura organizativa insuficientemente preparadas para la economía del dato.
- Costes de despliegue y operación de este tipo de infraestructuras.

Todos ellos representan barreras intangibles y, en algunos casos, tangibles desde la perspectiva de mercado para el desarrollo de una economía de datos. Las claves están en la aportación de valor, en la confianza, en la seguridad y en los costes de despliegue y acceso a datos.

Para la aportación de valor entre las partes será necesario el establecimiento de mecanismos de compensación adecuados para los propietarios de los datos por el acceso y uso de los mismos.

Con relación al aspecto tangible, el coste, un *DSES*, en lo físico, va a consistir en una infraestructura computacional que va a requerir espacio, energía, comunicaciones, soportes computacionales y recursos organizativos. Todo ello tiene un coste, tanto de capital inicial como de gastos de operación que hay que abordar a lo largo de su vida útil. Estos gastos van desde la contratación de personal, hasta el reemplazo de los recursos existentes, pasando por la innovación en la construcción y mejora de la plataforma de software y servicios que permite la explotación efectiva de los recursos computacionales subyacentes.

Superados los retos de confianza y seguridad, es preciso reducir los costes de transacción para incorporar al tejido empresarial en este tipo de infraestructuras, aspecto especialmente relevante para la pequeña y mediana empresa.

La resolución de los retos técnicos necesitará de inversiones para su resolución e integración en la plataforma de explotación de datos. Sin embargo, estas inversiones pueden realizarse de forma gradual.

Cabe resaltar la enorme importancia que representa que un *DSES* sea una infraestructura de utilidad pública, por lo que su planteamiento deberá ser merecedor de inversiones públicas adecuadas que garanticen, al menos, su sostenibilidad en el tiempo.

Finalmente, aunque planteado como infraestructura de utilidad pública, podrá pedirse una contribución, en forma de cuota por pertenencia, a su mantenimiento por parte de sus usuarios objetivo: las empresas y el sector público. Este último también podrá intervenir financiando total o parcialmente la participación de las empresas.

A modo ilustrativo, y basado en algunas experiencias previas como se verá en el apartado de casos de éxito (ITI DataHub), se pueden estimar los costes de inversión en la infraestructura que cabe esperar para un *DSES*.

La aspiración de un *DSES* es constituirse en el depósito real o virtual, concentrado o distribuido, por defecto de los datos de valor que se generen en una cadena de valor o en un entorno geográfico (por ejemplo, una región). Ello implica que el volumen de datos a almacenar va a crecer sin parar a lo largo del tiempo (los datos históricos también son valiosos y, por tanto, no deben descartarse).

Por otra parte, la carga computacional que debe soportar un *DSES*, según los objetivos que se planean, provendrá del conjunto de agentes productivos y de investigación de la cadena de valor, conjunto que forzosamente se verá acotado por limitaciones evidentes y que, en todo caso, presentará un crecimiento suficientemente limitado como para que las nuevas necesidades de cómputo sean satisfechas por las mejoras inherentes en los mantenimientos periódicos que deben realizarse sobre el equipamiento de cómputo y comunicaciones (ley de Moore en la sustitución de equipos).

Por ello, el factor más relevante a considerar en cuanto a analizar la evolución de costes es la dependencia respecto a la cantidad de datos a almacenar, cantidad que como hemos argumentado antes, no va a parar de crecer.

Por lo tanto, aunque presentamos figuras estimativas de costes de las unidades de cómputo, la atención debe dirigirse a la dependencia de los costes en el volumen de datos.

Se presenta a continuación una gráfica (*Figura 2.*) de evolución de costes en base a la evolución de un *DSES*, almacenamiento y en computación.

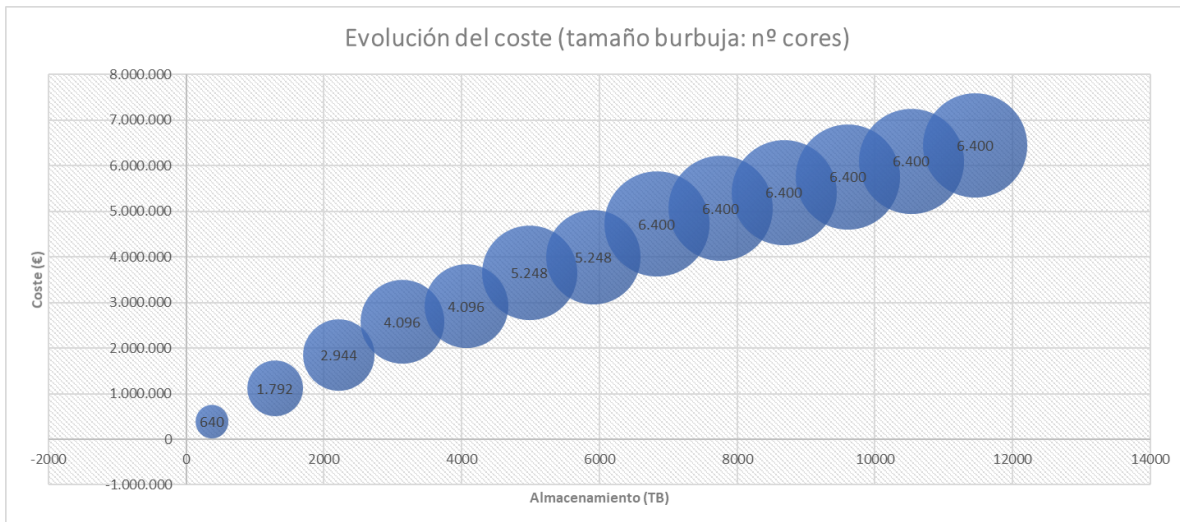


Figura 2. Evolución del coste de un DSES

### 5.3 Los retos técnicos

Conseguir el nivel de funcionalidad e interoperabilidad necesario para alcanzar los objetivos planteados para un DSES precisa de la puesta a punto de una serie de infraestructuras que faciliten los requisitos en términos de gobernanza, legales y operativos del ecosistema de datos. Los retos técnicos a los que se enfrentan estas infraestructuras basadas en datos son:

- Problemas técnicos para la interoperabilidad abierta, confiable y segura, que impacta también en los costes de acceso o transacción.
- Infraestructuras y tecnologías de datos.
- Ciberseguridad.

Es necesaria la creación e integración de diversas tecnologías para agilizar el proceso de acceso, uso y reutilización seguro y confiable, con el aseguramiento de las políticas sobre los datos de quienes ostenten los derechos. Para poder poner fácilmente a disposición de nuestro público objetivo dichas tecnologías, es preciso contar con herramientas capaces de gestionar los recursos computacionales que conforman la infraestructura del DSES.

Además, es necesario incorporar el know-how y herramientas adicionales que ataquen aspectos concretos del acceso y uso de datos de diversa índole (visión artificial, genómica, etc.) y para ello se precisa de personal capacitado para crear dichas herramientas y también para prestar servicios de consultoría técnica en este ámbito para las empresas y Administraciones Públicas.

Finalmente, será necesario asegurar que los mecanismos tecnológicos provistos por el DSES ofrecen la correcta gestión de las políticas relacionadas con la soberanía del dato, incluyendo la definición de aquellas condiciones y términos que constituyen las políticas que el dueño o el poseedor de los datos crea oportunas y controlar el acceso y uso de los datos de acuerdo con las mismas, considerando todo el ciclo de vida del ecosistema. Por ejemplo, un poseedor de datos sujetos a determinados derechos podrá retirar quizá privilegios a los usuarios de esos datos en determinadas circunstancias, asegurando el sistema que el usuario de esos datos no podrá acceder o usar más esos activos de datos -por ejemplo, si se hubieran caducado.

## 6 Modelos y casos de éxito de espacios de Datos

### 6.1 Modelos de referencia

#### 6.1.1 GAIA-X Una infraestructura de datos federada para Europa.

GAIA-X<sup>12</sup> surge con el fin de convertirse en la próxima generación de infraestructura de datos para la Unión Europea. Su concepto se basa en crear un **sistema federado y seguro que cumpla con los estándares de soberanía digital** y permita dar soporte a la innovación empresarial. GAIA-X conecta infraestructuras centralizadas y descentralizadas para convertirlas en un sistema homogéneo y fácil de usar. De esta forma, se convierte en la cuna de un ecosistema digital abierto donde los datos y servicios son recopilados y compartidos en un entorno de confianza. Con el objetivo de crear una infraestructura de datos eficiente, competitiva, segura y fiable, los **principios** que rigen GAIA-X son la *apertura-transparencia, interoperabilidad, federación, autenticidad y confianza*.

GAIA-X está diseñado para habilitar **ecosistemas federados**, con especificaciones y estándares comunes, políticas armonizadas y una gobernanza que permita mantener el equilibrio entre los requerimientos del consumidor y del proveedor.

#### ARQUITECTURA GAIA-X

La arquitectura de GAIA-X define los conceptos técnicos y funcionales necesarios para conseguir una correcta federación e interoperabilidad del Ecosistema GAIA-X. La **seguridad está en el centro de su diseño**, junto con elementos como *la privacidad, la descentralización y la facilidad de uso*, y es que para este modelo la protección de los Participantes, Datos y Sistemas que forman parte de un ecosistema GAIA-X es vital.

El Ecosistema GAIA-X está compuesto por el **Ecosistema de Infraestructura** y el **Ecosistema de Datos**, ambos conectados a través de **Servicios federados**, y soportado en una base sólida de **políticas, normas y estándares** que rigen su funcionamiento. La **Figura 3** ofrece una descripción general de alto nivel de su arquitectura<sup>13</sup>.

- **Ecosistemas de Infraestructura:** Enfocado en brindar/consumir Servicios de infraestructura necesarios para almacenar, transferir y procesar datos. En GAIA-X estos Servicios están representados principalmente por el Activo denominado Nodo. Con el enfoque de sistema federado GAIA-X proporciona Servicios mediante múltiples Proveedores y Nodos del ecosistema. Las partes interesadas en este ecosistema son:
  - ✓ Proveedores de servicios en la nube, que incluye todo tipo de proveedor de infraestructura en la nube sean pequeños, especializados o de gran escala.
  - ✓ Computación de alto rendimiento, este grupo cubre Proveedores de recursos informáticos de alto rendimiento, como universidades y laboratorios industriales.
  - ✓ Clouds sectoriales, el grupo está compuesto por Proveedores de servicios en la nube que ofrecen servicios a sectores específicos.
  - ✓ Edge Clouds: Son parte integral del Ecosistema de infraestructura.
  - ✓ Proveedores de interconexión y red

<sup>12</sup> <https://www.data-infrastructure.eu/GAIA/Navigation/EN/Home/home.html>

<sup>13</sup>GAIA-X: Technical Architecture Release. June, 2020; [https://www.data-infrastructure.eu/GAIA/Redaktion/EN/Publications/gaia-x-technical-architecture.pdf?\\_blob=publicationFile&v=5](https://www.data-infrastructure.eu/GAIA/Redaktion/EN/Publications/gaia-x-technical-architecture.pdf?_blob=publicationFile&v=5)

- **Ecosistemas de Datos:** En este ecosistema, los datos son tratados como la principal materia prima para la innovación. Las partes interesadas en este ecosistema son:
  - ✓ Proveedores de datos, quien crea datos disponibles para ser intercambiado.
  - ✓ Propietario de los datos, se refiere al autor original o propietario legal de un activo de datos.
  - ✓ Consumidor de datos, consumidor que recibe datos de un Proveedor de datos.
  - ✓ Proveedores de servicios inteligentes avanzados, aquellos que proporcionan Servicios de orden superior dentro de GAIA-X (ej: servicios basados en Artificial, aplicaciones e IoT).
  - ✓ Proveedores de servicios de soberanía de datos

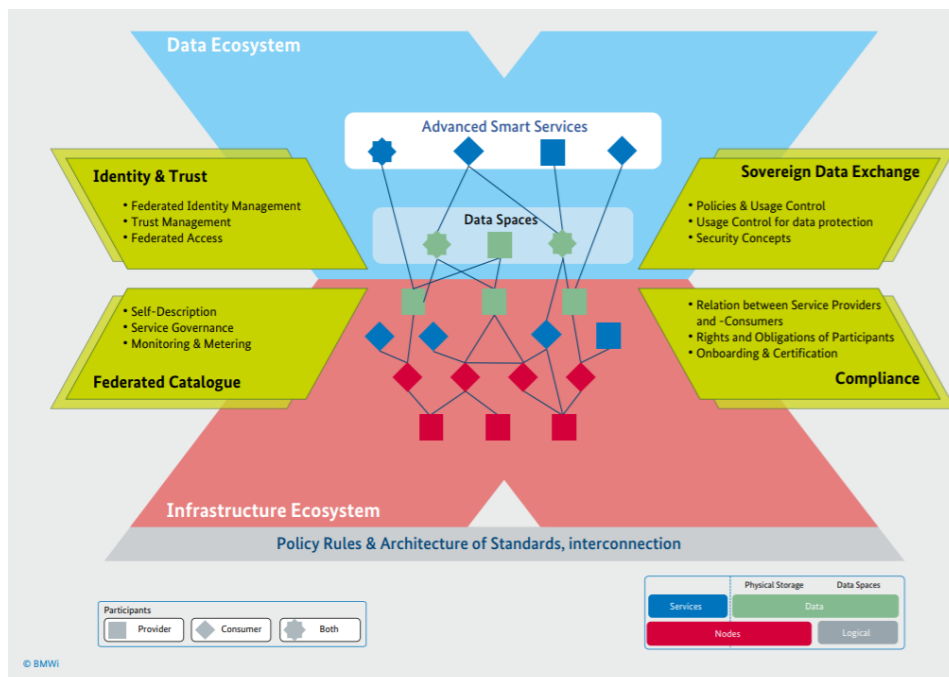


Figura 3. Visión general de alto nivel de la Arquitectura GAIA-X Fuente: GAIA-X Project

## GOBERNANZA

La confianza dentro de los participantes del Ecosistema GAIA-X es un aspecto fundamental, por ello este elemento es abordado desde dos perspectivas diferentes: la **confianza en la identidad** y la **confianza en las capacidades** de un Participante o Activo. Para lograrlo, cobra importancia el concepto de **Identidad**, como el elemento clave para poder acceder al Ecosistema. De esta forma, para acceder al ecosistema se realiza una verificación criptográfica de la Identidad usando un protocolo de gestión de identidad definido por GAIA-X. En términos generales, la identidad en GAIA-X se compone de un identificador único y una lista de atributos del mismo.

La **Figura 4** muestra el diseño el Modelo de Identidad Federada de GAIA-X y que aplica a Proveedores, Consumidores, Proveedores de Identidad y Visitantes.

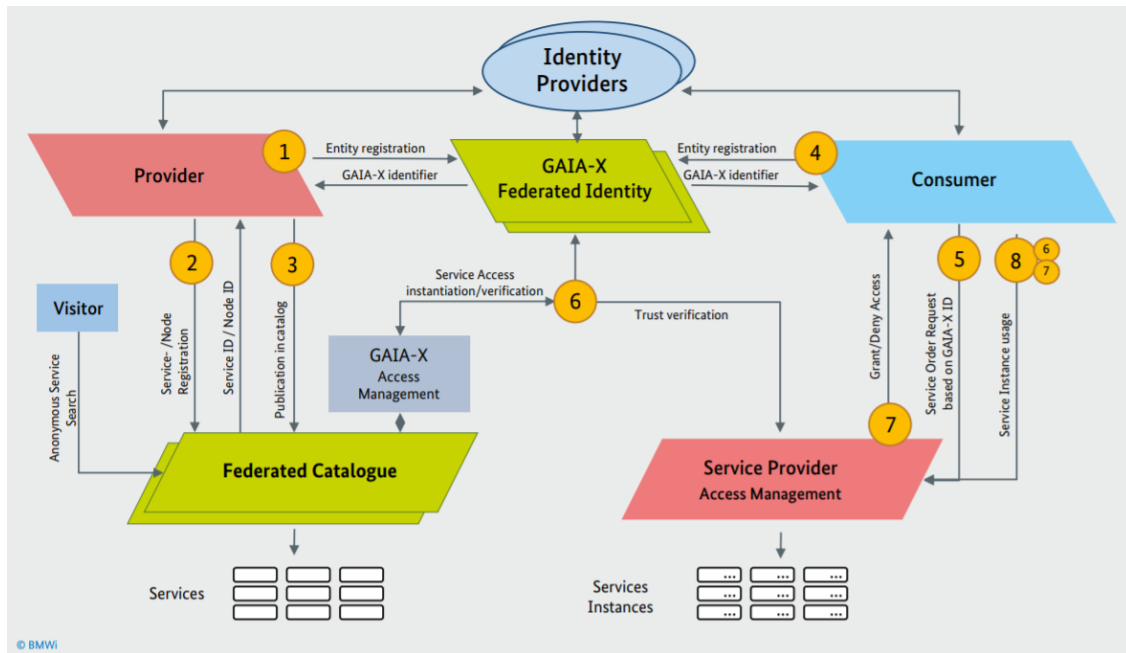


Figura 4. Modelo de Identidad GAIA-X. Fuente: GAIA-X Project

#### En el Anexo Caso GAIA-X

se detalla más información acerca de los elementos centrales de la arquitectura GAIA-X, el proceso de incorporación de Proveedores, Consumidores, Servicios y Nodos, así como la visión de GAIA-X desde el punto de vista de la seguridad de la información, la protección de datos, los mecanismos de control de acceso y las políticas de uso.

#### 6.1.2 Iniciativa International Data Spaces IDS, IDSA.

La Iniciativa **International Data Spaces (IDS)** promovida por **IDSA (International Data Spaces Association)** tiene como objetivo principal estandarizar el intercambio y la compartición de datos entre sus participantes, garantizando siempre la soberanía de sus datos. De manera que, su principal característica radica especialmente, en que los *Proveedores de datos (empresas)* siempre podrán mantener el control de los mismos, saber quién está usándolos, durante cuánto tiempo, para qué aplicación, en qué términos y bajo qué condiciones.

IDSA brinda el soporte necesario para el desarrollo de una arquitectura abierta, confiable, federada y de referencia para los *International Data Spaces (IDS)*. Una arquitectura en la que se garantice la soberanía de datos mediante la definición de unas condiciones básicas de gobernanza.

#### ARQUITECTURA IDS-RAM

El Modelo de arquitectura de referencia (IDS-RAM)<sup>14</sup> desarrollado por IDSA proporciona varios elementos, roles e interacciones estandarizados para una infraestructura en la que se realice un intercambio de datos soberanos. Este modelo utiliza una **estructura de cinco capas**: La capa de negocios, la capa funcional, la capa de procesos, la capa de información y la capa de sistemas. Abordándolas desde **tres perspectivas**: seguridad, certificación y gobernanza. La estructura general del modelo es ilustrada en la *Figura 5*.

<sup>14</sup> IDS Project. Reference Architecture Model versión 3. April, 2019; <https://www.internationaldataspaces.org/wp-content/uploads/2019/03/IDS-Reference-Architecture-Model-3.0.pdf>

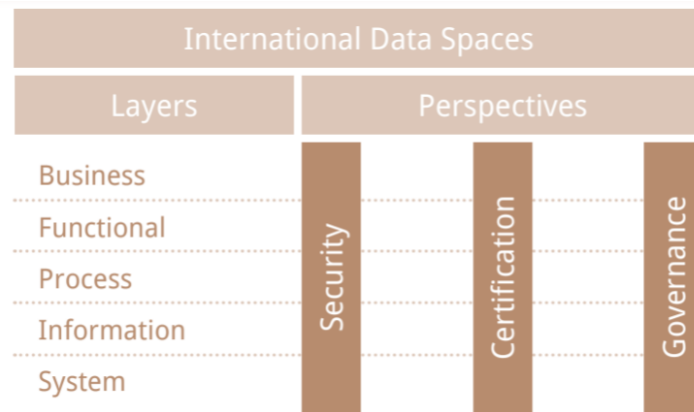


Figura 5. Arquitectura de referencia IDS. Fuente: IDS

### 1. CAPA DE NEGOCIOS.

Esta capa define y categoriza los diferentes roles que los participantes pueden asumir, además especifica los patrones básicos de interacción entre ellos y por tanto sirve de input para generar los requerimientos a tener en cuenta en la capa funcional. Los 4 roles definidos en la arquitectura son:

- **Categoría 1: Participante esencial.** Son participantes necesarios para realizar un intercambio de datos en el IDS y puede por tanto puede asumirlo cualquier organización que posea, quiera ofrecer, quiera consumir o usar datos (*Data Owners, Data Providers, Data Consumers, Data Users* y *App Provider*).
- **Categoría 2: Intermediario.** Participantes que actúan como entidades de confianza y de intermediación. Dentro de esta categoría se encuentran, por ejemplo: *Proveedores de servicio de intermediación* (broker), *Cámaras de compensación* (Clearing House), *Proveedores de identidad*, *Proveedores de Apps* o *Proveedores de Vocabulario*.
- **Categoría 3: Proveedor de Servicios/Software.** Esta categoría comprende las compañías de IT que proveen servicios y/o software a los participantes del IDS. Dentro de los participantes que puede asumir este rol se encuentran los *Proveedores de Servicios* y los *Proveedores de Software*.
- **Categoría 4: Órgano de gobierno:** Este tipo de participantes no están involucrados en las actividades diarias del IDS, como por ejemplo los órganos de gobiernos: El Órgano de certificación y la Asociación IDS.

La [Figura 6](#). proporciona una visión global de los **roles y las interacciones** que pueden darse entre ellos. Además de las descritas en la figura, es importante mencionar que tanto el Proveedor de software como el Proveedor de Identidad están conectados a todos los demás roles.

Un aspecto fundamental para participar **es la certificación**, la gran mayoría de los roles asumidos necesitan certificarse tanto a nivel de capacidades organizacionales como de sus componentes técnicos. Esta certificación permite generar un mayor grado de confianza en todo el ecosistema. El esquema de Certificación aplicado se muestra en la [Figura 7](#).



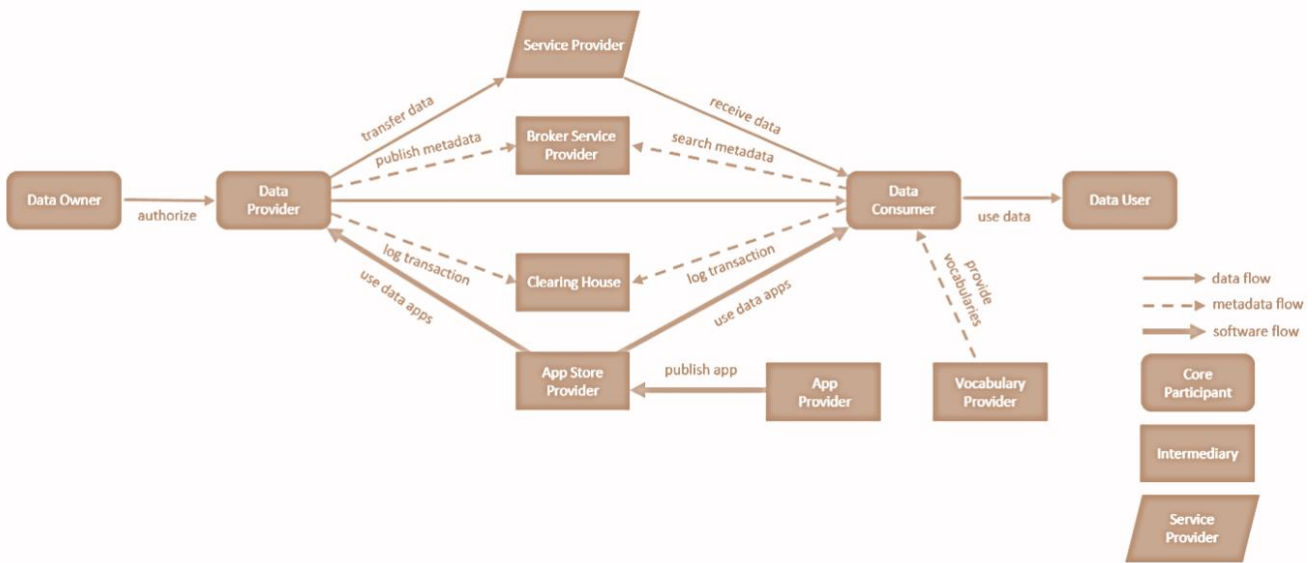


Figura 6. Roles e interacciones. Fuente: IDS

Por su parte, el **IDS Connector** se posiciona como uno de los componentes centrales de la arquitectura, toda vez que son los responsables del intercambio de datos. Es importante resaltar que la transferencia de los datos se realiza entre los Conectores del Proveedor de datos y los del Consumidor de datos. En este proceso, las **identidades digitales** cobran gran importancia, ya que son necesarias para compartir e intercambiar datos en un *International Data Space (IDS)*. Por tanto, es necesario que los participantes cuenten con una identidad digital que puede ser del tipo estático (basada en la certificación de los participantes y componentes) o dinámico (basado en el seguimiento activo de los participantes y componentes).

Por último, esta capa aborda **los contratos de uso**, los cuales comprenden un conjunto de Políticas de uso que describen ciertos permisos y obligaciones. Cada contrato de uso entre los participantes consta de una parte técnica, centrada en la descripción de las interfaces técnicas y una no técnica dedicada a los aspectos relacionados con el intercambio de datos previsto.

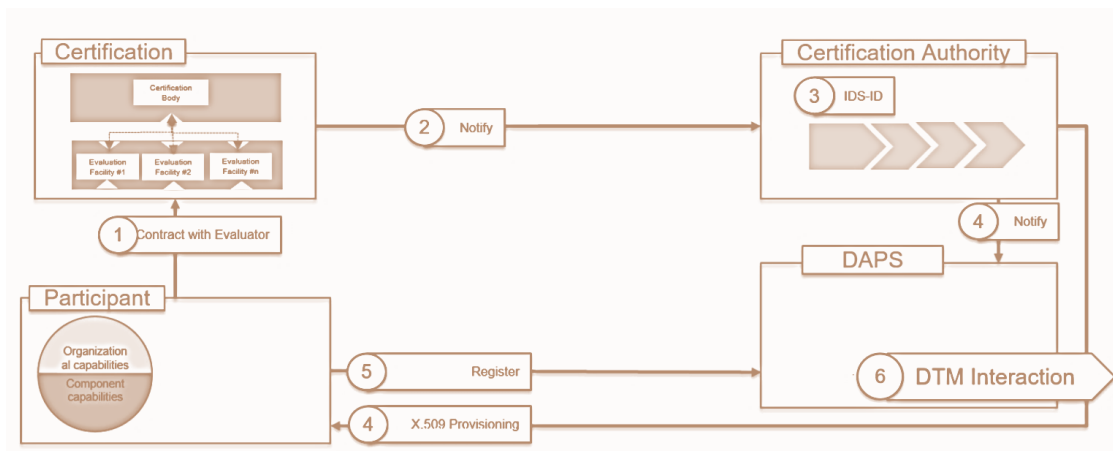


Figura 7. Proceso de Certificación. Fuente: IDS

## 2. CAPA FUNCIONAL.

Esta capa define los requisitos funcionales del *DSES*, la arquitectura de referencia IDS-RAM toma estos requisitos funcionales junto con algunos que, aunque no son funcionales sí son vitales para el correcto funcionamiento del modelo y los agrupa en 6 categorías:

- **Confianza:** Un requisito no funcional pero vital para el modelo, comprende los roles y la gestión de identidad y certificación de usuarios.
- **Seguridad y soberanía de datos:** Un requisito no funcional, pero una de las características fundamentales del IDS. Este grupo aborda 4 aspectos: la autenticación y autorización, las políticas de uso, la comunicación fiable y la certificación técnica.
- **Ecosistema de datos:** Los datos son la base del modelo y por tanto su funcionalidad en cuanto a ser capaz de describir, encontrar e interpretar correctamente estos datos es un aspecto clave. Este grupo comprende aspectos como: la descripción de las fuentes de datos, la intermediación (brokering) y el vocabulario necesario para crear y estructurar metadatos.
- **Estandarización e interoperabilidad:** Este grupo es el más importante de los 6 ya que aborda los aspectos fundamentales para lograr con éxito el intercambio de datos y asegura la operabilidad del IDS.
- **Aplicaciones de valor agregado:** Bien sea antes o después del intercambio de datos, estos pueden necesitar ser transformados o procesados, por esa razón, IDS ofrece ciertas aplicaciones (Data Apps). Estas aplicaciones tienen un ciclo de vida (implementación, disponibilidad, instalación y soporte) que impone requisitos funcionales dentro del modelo.
- **Mercado de datos:** Teniendo en cuenta que los datos que se intercambian en los IDS pueden llegar a tener un valor monetario, el modelo integra aspectos como: facturación, restricciones de uso, gobernanza, entre otros.

### PERSPECTIVA DE GOBERNANZA

La perspectiva de gobernanza de este modelo define los requisitos a cumplir por el Ecosistema empresarial para lograr una interoperabilidad segura y confiable. En ese sentido, la arquitectura de referencia (IDS-RAM) debe verse como un marco funcional que proporciona mecanismos que las organizaciones participantes pueden personalizar de acuerdo con sus requisitos individuales.

El *Modelo de Gobernanza de Datos* define un marco de derechos y procesos de toma de decisiones con respecto a la definición, creación, procesamiento y uso de datos. Por un lado, las actividades principales se orientan a establecer la directriz general del sistema de toma de decisiones y por otro, la gestión de datos se orienta a las actividades de creación, procesamiento y uso de datos. En el contexto de un *International Data Spaces (IDS)*, la gobernanza de datos comprende también derechos de uso de datos compartidos e intercambiados dentro del ecosistema.

En el *Anexo Caso International Data Spaces (IDS)* se detalla más información acerca del proceso de incorporación a IDS, la adquisición de identidad de los Usuarios, el proceso de publicación de Servicios y se especifica el Modelo de Información y lenguaje común que facilita la compatibilidad y la interoperabilidad entre los participantes y los componentes del IDS.

### 6.1.3 EOSC (European Open Science Cloud)

La EOSC (European Open Science Cloud) impulsada desde 2015 por la Comisión Europea, surge con el objetivo de promover el acceso y la reutilización de datos y activos (publicaciones, software, etc) derivados de investigaciones financiadas con fondos públicos. La EOSC es un **entorno federado y virtual en la nube creado con el fin de alojar y procesar datos de investigación** y así apoyar la ciencia y la innovación abierta de la UE. Esta iniciativa reúne diferentes infraestructuras de datos institucionales, nacionales y europeos para desarrollar un ecosistema de ciencia abierta e inclusiva en Europa.

La EOSC se convierte en un único punto de acceso gratuito en el que diferentes bases de datos son interoperables. Basado en los principios de **apertura, comportamiento ético e integridad de la investigación**, la EOSC se guía por políticas aplicables tanto a Usuarios, Proveedores como a los propios Recursos.

La **Estrategia Europea de Datos** reconoce actualmente a la EOSC como el núcleo de un *DSES* de ciencia, investigación e innovación que se articulará con los 9 Espacios de Datos sectoriales previstos en esta estrategia. En ese sentido, se prevé que a partir de 2024 esta iniciativa se empiece a articular y conectar con stakeholders del sector público y el sector privado y para 2025 se estima la implementación total de las operaciones EOSC al servicio de los investigadores de la UE. Así mismo, en el marco del nuevo **Programa Horizon Europe 2021-2027**, la EOSC ejercerá su rol como European Partnership, y es que Horizon Europe proporciona un marco ideal para la colaboración y puesta en común de recursos a nivel europeo, nacional, regional e institucional.

#### ARQUITECTURA DE LA EOSC

Para la EOSC los **Datos** y las **Infraestructuras de datos** forman lo que puede considerarse como una relación "Yin-Yang", en donde uno no es posible sin el otro. El **principio fundamental** por el que se rige la EOSC es la investigación como núcleo central del modelo, por tanto, un fuerte compromiso con las comunidades de investigación es necesaria para comprender sus requisitos y asegurarse de que la forma en que EOSC opera y se ofrecen los servicios es de utilidad y valor para los investigadores. La EOSC involucra una visión en la que los científicos deben adoptar un nuevo enfoque en donde el conocimiento debe ser compartido en todas las etapas del ciclo de vida de la investigación, en contraste con una visión más antigua en la que los resultados solo eran compartidos a través de publicaciones una vez el trabajo había finalizado.

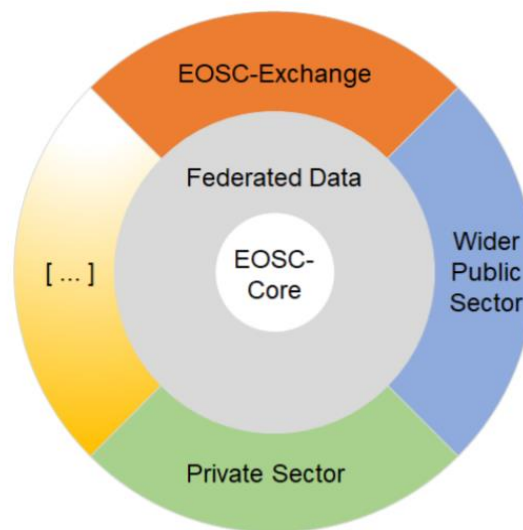
Los principios en los que se basa la gestión de datos en la EOSC son la **apertura**, los **principios FAIR** y la **federación de Infraestructuras**. Surge pues, con la visión de ser una **federación de infraestructuras de datos de investigación** (existentes y previstas) las cuales a través de una capa transversal permiten la conexión y funcionamiento como **una sola Infraestructura europea**.

De acuerdo con la Agenda estratégica de innovación e investigación de la EOSC <sup>15</sup>, esta será principalmente será una federación de servicios y repositorios de datos temáticos existentes, interconectados mediante un marco de intercambio de datos definido. Sobre la base de las infraestructuras de datos de investigación existentes, la arquitectura de EOSC se desarrollará en fases o iteraciones. Cada fase o iteración agregará más funcionalidades y servicios que darán soporte a un mayor número de usuarios y podrá por tanto satisfacer una gama más amplia de casos de uso.

<sup>15</sup> <https://www.eosc.eu/sites/default/files/EOSC-SRIA-V09.pdf>

La primer fase o iteración se enfoca en conseguir lo que se denomina como el **mínimo viable de la EOSC (MVE)** con el que se buscará la federación de las infraestructuras de datos de investigación y el acceso abierto a los datos y servicios. Ese mínimo viable incluye el EOSC-Core y el EOSC-exchange que trabajan con los datasets que se federarán en la EOSC, relacionándose entre sí tal como se muestra en la *Figura 8*.

- El **EOSC-Core**, ofrece a través de una plataforma online los medios y los mecanismos necesarios para nombrar, localizar, descubrir, compartir, acceder y reutilizar los datos y servicios; así como un marco común para administrar la identidad y el acceso de los usuarios.
- El **EOSC-Exchange**, compuesto por un conjunto de servicios que almacenan, explotan los datos y fomentan su reutilización. Ejemplos de servicios incluidos en *EOSC-Exchange* son aquellos que almacenan, preservan o transportan datos de investigación.



*Figura 8. Relación entre los elementos de la EOSC.*

Respecto a los **Servicios y Datos federados**, si bien es cierto que existen cientos de componentes y datos en los estados miembros y países asociados que potencialmente podrían ser federados, la EOSC actualmente ha federado una cantidad relativamente pequeña de Servicios accesible a través de su Marketplace. Sin embargo, se espera que crezca sustancialmente.

En este sentido, los **estándares comunes** son imprescindibles para cumplir con los principios FAIR y por ello es necesario utilizar estándares comunes para datos, vocabularios comunes, metadatos y procedimientos en cuanto a cómo, cuándo y dónde los datos son compartidos. Para permitir el uso interdisciplinario dentro de EOSC, las estructuras de **gobernanza** deben primero facilitar la aceptación de estándares construidos sobre una base de calidad de los datos y metadatos.

Para posteriormente, centrarse en el **marco de interoperabilidad** (EOSC IF), brindando un acceso a los servicios que permitan a los investigadores llevar a cabo sus actividades de investigación de manera más eficaz mediante un intercambio más rápido y fluido de publicaciones, datos, software y otros resultados de la investigación. Por tanto, estos servicios deben ser fácil de usar y debe respaldar todas las fases del ciclo de vida de la investigación. Para lograr la interoperabilidad de los servicios y la compatibilidad de los datos federados en el marco de la EOSC, *el marco de interoperabilidad* incluye especificaciones técnicas, políticas y procedimientos para la gestión de datos siguiendo los principios FAIR y de seguridad. Su estructura está basada en el Marco Europeo de Interoperabilidad y se compone de 4 capas: La legal, organizacional, semántica y técnica, como se muestra en la *Figura 9*. El desarrollo

de casos de uso y la implementación del Marco de Interoperabilidad EOSC están previstas para la próxima fase de la EOSC.

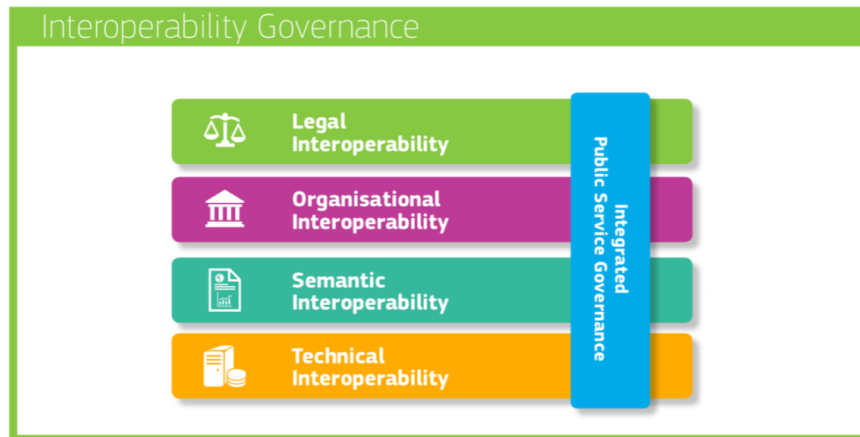


Figura 9. Marco de Interoperabilidad

#### 6.1.4 Otras referencias

- Data Sharing Coalition<sup>16</sup>
- iSHARE<sup>17</sup>
- Engagement VOLontaire pour une Logistique Urbaine Efficente, EVOLUE<sup>18</sup> (Urban logistics data sharing)
- Plataforma Opal – Open Algorithms<sup>19</sup>
- Data Market Austria<sup>20</sup>
- Data Trading Alliance<sup>21</sup>
- FIWARE<sup>22</sup>
- EBRINS<sup>23</sup>

## 6.2 Casos de éxito

### 6.2.1 Caso 1. ITI DATAHUB basado en NEODIS

**Reto y solución.** ITI tiene un enfoque multisectorial y en torno a las tecnologías del ciclo del dato. El reto que se plantea es justamente el descrito a lo largo del presente documento, es decir, proporcionar un conjunto de tecnologías que permitan extraer valor de los datos, fomentar la innovación económica y social mediante la explotación de los mismos y contribuir a generar un modelo de sociedad empoderada por los datos.

<sup>16</sup> <https://datasharingcoalition.eu/>

<sup>17</sup> <https://www.ishareworks.org/en>

<sup>18</sup> <https://www.francesupplychain.org/evolue>

<sup>19</sup> <https://www.opalproject.org/home-es>

<sup>20</sup> <https://datamarket.at/en/>

<sup>21</sup> <https://data-trading.org/en/alliance-outline/>

<sup>22</sup> <https://www.fiware.org/>

<sup>23</sup> <https://ebrains.eu/>

**ITI DataHub**, es un *DSES*, que ha recibido el label GOLD de BDVA como iSpace<sup>24</sup> desde 2017, y ha sido puesto a disposición de la alianza *Inndromeda*<sup>25</sup> *eDIH*, y que implementa el modelo NEODIS, cuyo objetivo es acelerar la innovación basada en Big Data, Inteligencia Artificial y High Performance Computing a través de un centro demostrador y de experimentación que aúna esfuerzos en estas áreas en la Comunitat Valenciana. ITI ha trasladado a este Datahub parte de sus desarrollos experimentales en materia de plataforma de soporte, que ya venía aplicando en su propio centro de datos.

NEODIS es el modelo diseñado por ITI de Espacio Público de Gestión y Explotación de Datos, así como de Experimentación y Certificación, que proporciona las guías y componentes tecnológicos (*building blocks*) para la creación de un *DSES* ajustable a distintas configuraciones de infraestructura y tipología de datos, de forma que puede desplegarse para distintos tipos de objetivos, desde un Espacio de Explotación de Datos corporativo, un Espacio de Explotación de Datos para un sector o cadena de valor o un Espacio Público de Explotación de Datos para una región o país.

NEODIS propone una plataforma tecnológica alrededor de la cual se genera un **liderazgo firme** propiciando un **consenso** que facilita las adhesiones de los agentes generadores de datos y del resto de agentes intervinientes, por medio de la **confianza**, apoyada en:

- Casos de uso que clarifican la utilidad
- Neutralidad y capacidad de albergar contribuciones de múltiples procedencias
- Capacidad de gestión
- Formación
- Solvencia técnica y capacidad de ejecución con calidad y excelencia
- Capacidad de innovación y evolución tecnológica

**Beneficios:** ITI DataHub proporciona los siguientes beneficios:

1. Una **infraestructura de almacenamiento y procesamiento de datos eficiente y escalable**.
2. **Relación de confianza con diversos agentes productores de datos** de alto valor para su almacenamiento y protección en el DataHub, posibilitando así la innovación por parte de diferentes agentes: empresas, organismos de investigación y administraciones públicas.
3. Un **sistema de gestión que garantiza el cumplimiento de las políticas de uso y acceso** a los datos estipuladas por sus propietarios.
4. Un **sistema de compartición y explotación de los datos que permite el acceso a los mismos** por parte de una diversidad de agentes, pero enfocado principalmente a las empresas y las administraciones públicas. De este modo estas últimas también pueden aumentar su productividad y la calidad del servicio que ofrecen a los ciudadanos.
5. Un **sistema de planteamiento de experimentos ampliamente accesible** con diversos grados de explotación.
6. Un **sistema de “exportación” de los resultados/procedimientos** que permite a empresas y administraciones portar procedimientos derivados de la explotación de dichos datos a sus propios sistemas de explotación.

<sup>24</sup> <https://www.bdva.eu/I-Spaces>

<sup>25</sup> <https://www.inndromeda.es/>

### Componentes y tecnologías:

El ITI DataHub está compuesto por un clúster de computación con dos entornos diferenciados. Primero, un entorno de infraestructura compartida basada en OpenStack orientado a soluciones de Big Data e IA. Segundo, un entorno de infraestructura de alta computación HPC para soluciones que requieran recursos dedicados y uso de GPUs. Estos dos entornos se encuentran distribuidos en varios clústeres. Respecto a las comunicaciones de red se dispone de varios tipos de switches de interconexión que van desde 1 GB a 100 GB.

#### *Software y servicios*

Actualmente se ofrecen cuatro servicios distintos:

- Baremetal, que es una versión que proporciona únicamente las máquinas para computación y es el usuario el que tiene que instalar todo el sistema.
- Una versión más avanzada con OpenStack (IaaS) más un conjunto de herramientas para Big Data Analytics (Jupyter y Spark) que facilitan mucho el despliegue de los experimentos.
- Radiatus, que ofrece un servicio que facilita la configuración y despliegue de servicios robustos, eficientes, tolerantes a fallos y elásticos para soluciones de Big Data Analytics.
- Una versión que permite disponer de recursos mediante GPUs para llevar a cabo sus experimentos con elevadas necesidades de computación.

#### *Modelo de gobernanza*

ITI DataHub ha desplegado un sistema de gobernanza de las infraestructuras y los datos del DSES y que se basa en la existencia de un comité compuesto por 5 instituciones que representan el ámbito tecnológico, científico, empresarial y de la Administración Pública, que ha establecido los criterios para albergar experimentos.

### 6.2.2 Caso 2. HARMONY Alliance & HARMONY Plus.

**Reto y solución.** En conjunto, las neoplasias hematológicas ocupan el tercer lugar, tras el cáncer de pulmón y el cáncer colorrectal, en términos de mortalidad ajustada por edad en Europa. En las últimas décadas, ha mejorado sustancialmente el diagnóstico y tratamiento de los cánceres hematológicos, sin embargo, muchos siguen siendo incurables.

*Harmony Alliance*<sup>26</sup> es un proyecto IMI (de colaboración público-privada entre EFPIA y la Comisión Europea) cuyo objetivo final radica en mejorar los tratamientos de distintos tipos de cáncer hematológico, que afectan al sistema circulatorio y linfático. A partir del éxito obtenido por Harmony Alliance, surge *Harmony Plus*<sup>27</sup> que permite desbloquear y difundir conocimientos sobre otras 3 neoplasias hematológicas (cánceres hematológicos). HARMONY PLUS permite el uso de los últimos aprendizajes y desarrollos en el campo del análisis de big data y la inteligencia artificial (IA) para la atención sanitaria. 39 socios y 8 asociados de 10 países hacen parte del ecosistema Harmony Plus.

La *Alianza Harmony* utiliza Big data para, en última instancia, mejorar el tratamiento de estas neoplasias. La plataforma de Big Data establecida por HARMONY es una herramienta única y lo suficientemente madura para avanzar en la comprensión de muchas neoplasias hematológicas. Los análisis están abordando muchas necesidades no satisfechas, como la definición de mejores

<sup>26</sup> <https://www.harmony-alliance.eu/>

<sup>27</sup> <https://harmonyplus.com/>

estrategias de tratamiento para pacientes individuales, la predicción de acontecimientos adversos en subpoblaciones específicas y la definición de conjuntos estándar de medidas de resultado relevantes. Los servicios integrados de HARMONY PLUS llevarán la plataforma de Big Data al siguiente nivel al ampliando el acceso a las herramientas analíticas existentes y emergentes.

Harmony Alliance pretende convertirse en colaborador de distintos proyectos internacionales en el contexto de la pandemia, como herramienta para examinar las características de los coronavirus. HARMONY, con su Plataforma de Datos COVID-19, desempeña un importante papel como centro de datos, facilitando la entrega de conjuntos de datos armonizados, conocimientos técnicos y experiencia de sus socios para proyectos lanzados o co-creados por o con instituciones sanitarias europeas

**Beneficios:** Con el objetivo de estudiar un gran nº de pacientes, se ha creado la Plataforma Harmony Big Data, que reúne datos clínicos armonizados de aproximadamente 45.000 pacientes europeos. Esto permite caracterizar molecularmente las neoplasias hematológicas, comprender su fisiopatología e identificar nuevas dianas farmacológicas, predecir la evolución de la enfermedad y la respuesta a los fármacos en subgrupos de pacientes.

HARMONY y HARMONY PLUS trabajan intensamente y de forma coordinada para proporcionar herramientas de análisis de datos en formatos que puedan ser adoptados por los hospitales y utilizados por los médicos para mejorar los resultados de los pacientes de HM.

#### Componentes y tecnologías:

- Análisis basados en tecnologías como Big Data y Data Science y nuevas técnicas de Inteligencia Artificial (IA) para mejorar la capacidad de análisis y aumentar la precisión de los resultados.
- La Plataforma Harmony es un repositorio central donde se recogen datos anónimos de forma segura siguiendo todos los requisitos legales y éticos, se armonizan y luego se analizan.
- Al adaptar la plataforma y las herramientas para que las utilicen los hospitales y los médicos con el fin de mejorar los resultados de los pacientes, HARMONY PLUS pretende llevar los resultados al lugar donde se encuentra el paciente.

#### 6.2.3 Caso 3. Madrid Green Data Space

**Reto y solución:** Madrid Green Data Space<sup>28</sup> es un Espacio de Datos (en fase de prototipo y validación) diseñado y desplegado para dar un soporte digital (en lo referente a la capa de datos) a algunos de los experimentos que se están diseñando en el portfolio asociado al demostrador profundo (Deep Demonstrator) de la ciudad de Madrid, en su hoja de ruta hacia la descarbonización de la ciudad, auspiciado por el programa de EIT Climate-KIC de Deep Demonstrations for Healthy, Clean Cities (<https://www.climate-kic.org/programmes/deep-demonstrations/healthy-clean-cities/>). El proyecto ha sido financiado por EIT Digital en el contexto de una actividad Cross-KIC, conjuntamente con EIT Climate-KIC.

Este Espacio de Datos gestiona un catálogo de conjuntos de datos de diversas instituciones (públicas y privadas), incluyendo distintos niveles de administraciones públicas (locales, regionales, nacionales e internacionales). Estos conjuntos de datos han sido seleccionados en función de los experimentos que han sido definidos hasta el momento en este catálogo, y siguiendo un protocolo que ha sido definido dentro del modelo de gobernanza de datos para el Deep Demonstrator.

Asimismo, además del catálogo de datos (cuyos metadatos están basados en DCAT-AP) se ofrecen conectores especializados a distintas fuentes de datos (bases de datos, ficheros en formato shapefile,

<sup>28</sup> <https://mgds.oeg.fi.upm.es/>



datos procedentes de la Agencia Espacial Europea, etc.) y se han generado nuevos conjuntos de datos con índices para la ciudad de Madrid que combinan datos de vegetación, parques y jardines, etc., con datos socio-demográficos. Estos datos también están disponibles en el catálogo y disponibles para su visualización en la Web del Dataspace.

**Beneficios:** Gracias a este Espacio de Datos, los experimentos del Deep Demonstrator asociado a Madrid tienen un punto de acceso unificado a todo el catálogo de conjuntos de datos del que están haciendo uso. Asimismo, el modelo de gobernanza de datos define un conjunto amplio de aspectos que son necesarios para llevar a cabo de manera efectiva la gestión del Espacio de Datos en el conjunto de entidades que participan en el mantenimiento de dicho Espacio de Datos.

#### Componentes y tecnologías:

- Se ha partido de la implementación de la arquitectura de referencia propuesta por IDSA, dado que la implementación de referencia de GAIA-X no estaba aún disponible en el comienzo del proyecto.
- Debido al grado de madurez bajo en el que se encuentra esta implementación, ha sido necesario realizar modificaciones sobre dicha implementación para corregir errores de despliegue y para eliminar dependencias con otros sistemas.
- Asimismo, se ha desarrollado un nuevo broker y se ha tenido que actualizar el componente de autorización.

#### 6.2.4 Caso 4. Innovalia: Espacio de Datos Europeo IDSA para la Conexión del Gemelo Digital a lo Largo del Ciclo de Vida del Producto

**Reto y solución.** El aumento de la complejidad de los productos modernos y la aparición de conceptos como la Ingeniería de Sistemas Basada en Modelos (MBSE) hacen del desarrollo de estructuras federadas de compartición y análisis distribuido de datos, un gran reto.

A este reto se le ha conseguido dar respuesta a través del consorcio de más de 50 empresas, unidas en el desarrollo del primer prototipo de espacio europeo de datos industriales, que conecta infraestructuras privadas, infraestructuras gestionadas por operadores y espacios de datos iSpaces, explotando las capacidades en el Edge, cloud y HPC para la gestión y análisis *Big Bata* industrial.

De este espacio de datos industriales 4.0, “embriónico” según reconoció la EC en 2021, se han seleccionado, tanto el **testbed del Mars Rover** galardonado en el IoT solution World Congress 2021, como el piloto **“Trusted digital twins data connection across the product life cycle”** seleccionado por el Radar de Innovación de la EC29. El Grupo Innovalia y la industria vasca juegan un papel protagonista en el desarrollo de la tecnología de intercambio de datos utilizada, siendo un avance prometedor para la industria 4.0.

Los procesos desarrollados en el **proyecto lighthouse Boost 4.0**, se pueden consultar en el sitio web de la Digital Factory Alliance<sup>30</sup> (DFA). Igualmente, sobre el demostrador se puede encontrar más información, tanto en los sitios de la propia Asociación IDSA, como la Fundación FIWARE. Así como en

<sup>29</sup> <https://www.innoradar.eu/>

<sup>30</sup> <https://digitalfactoryalliance.eu/>

el libro “Technologies and Applications for Big Data Value”<sup>31</sup>, capítulo 23; cuya publicación está estimada en el segundo cuatrimestre de 2021.

El rápido crecimiento del Internet de las Cosas ha hecho que el verdadero reto, para el *Big Data* en las fábricas 4.0, esté en la velocidad, la transparencia y la confianza, con las que se puede establecer y acceder a redes de datos interoperables, altamente heterogéneas y multidominio. Así como en la sincronización, en tiempo real. O lo que es lo mismo, establecer “hilos digitales” sin fisuras.

### *Demostrador Mars Rover*

El demostrador del *Mars Rover* está basado en la fabricación de una única pieza, y demuestra la viabilidad de la implementación del gemelo digital y del “hilo digital”, mediante la colaboración empresarial<sup>32</sup> a lo largo del ciclo de vida del producto. Consiguiendo así la explotación combinada de la Ingeniería Basada en Modelos (MBE que soporta los estándares abiertos PLM STEP estandarizados<sup>33</sup> y la semántica del Marco de Información de Calidad (QIF)<sup>34</sup>), los “hilos digitales” y las tecnologías de gemelos digitales basadas en conectores de confianza.

### **Beneficios**

- Diseño mejorado y más robusto basado en el rendimiento del producto.
- Excelencia de fabricación en todo el ciclo de vida del producto.
- Previsión inteligente de defectos y planificación de tareas de mantenimiento predictivo.

### **Componentes y tecnologías:**

Se han interconectado y abierto los datos industriales en las máquinas y plataformas de fabricación digital, sin perder el control sobre el uso de los mismos gracias a una implementación de la arquitectura de referencia IDSA. Esto se logra, a través de la incorporación de *gateways* de información confiables, certificados y estandarizados según la norma DIN SPEC 27070, conocidos como conectores de datos en el ecosistema IDS, con un enfoque para aprovechar la continuidad de los datos.

### **Componentes Tecnológicos**

- Integración de infraestructuras y espacios de datos privados, gestionados y públicos.
- Interacción y computación de datos optimizada en el Edge, nube y HPC
- Compartición de datos basados en implementaciones abiertas (OSS) de conectores IDSA.
- Modelos de datos abiertos.
- Gestión de información a lo largo de ciclo de vida de producto
- Plataformas y modelos avanzados de análisis visual y prescriptivos.
- Gestión de servicios de análisis big data basados en microservicios.
- Trazabilidad y gobernanza basados en tecnológica de distributed ledger
- Modelos de análisis de información 3D.
- Plataforma de interoperabilidad semántica (IoF, Industry Commons, QIF, ProSTEP).
- Servicios de ingeniería basada en modelos.
- Plataforma de validación y verificación de interoperabilidad de componentes del espacio de datos. Integration Test Camp (ITC)<sup>35</sup>.

<sup>31</sup> <https://www.bdva.eu/node/1616>

<sup>32</sup> FIWARE, +GF+, EPFL, Innovalia Metrology, Pro-STEP

<sup>33</sup> International Organization for Standardization. ISO 10303 Industrial Automation Systems and Integration – Product Data Representation and Exchange. Geneva: ISO.

<sup>34</sup> International Organization for Standardization (2020). ISO 23952:2020. Automation systems and integration — Quality information framework (QIF) — An integrated model for manufacturing quality information. Geneva: ISO.

<sup>35</sup> <https://www.sqs.es/q-idsa-itc/> <https://internationaldataspaces.org/adopt/ids-lab/>

### TRIMEK - fabricación cero defectos (ZDM) con Metrología Masiva 4.0

Las tecnologías mostradas en el demostrador del *Mars Rover*, se han escalado a un proceso de fabricación cero defectos en el sector de automoción. La fabricación sin defectos (ZDM) impulsada por la metrología masiva tiene como objetivo mejorar el rendimiento y la eficiencia de los procesos de control de calidad esenciales en las líneas de fabricación.

**Los pilares clave** de este piloto son: (1) la implementación de un proceso en línea de metrología de alta definición, para capturar grandes volúmenes de datos de nubes de puntos en 3D; (2) la integración y análisis de datos heterogéneos, tanto de calidad, como de producto, procedentes de diferentes fuentes, incorporando al flujo de metrología datos procedentes de las etapas de diseño del producto; y, por último, (3) el desarrollo de un proceso avanzado de colaboración y analítica visual centrado en el ser humano y basado en el mapeo avanzado de colores.

En la actualidad la industria demanda soluciones de metrología más holísticas, flexibles y rápidas. El ensayo de la empresa española Trimek (Innovalia Metrology) aprovecha la ingeniería basada en modelos, para conseguir flujos de trabajo masivos de metrología 4.0, que implementen: un flujo de *Big Data*, para el procesamiento, y mapas avanzados de color texturizados de alta velocidad, para el análisis visual de alta fidelidad de nubes de puntos 3D masivas (de 10 a 100 millones de puntos) en menos de 30 segundos. El ensayo de Trimek ZDM Massive Metrology 4.0 considera dos procesos empresariales principales para su implementación: **Metrología de alta densidad** y **Metrología masiva virtual**.

#### Beneficios:

Estas capacidades han permitido a la industria del automóvil trabajar con fluidez con archivos CAD 10 veces más grandes y controlar más de 600 características geométricas. La velocidad de procesamiento se ha multiplicado por 5 y el tiempo necesario para programar el escaneo de cada pieza se ha reducido un 20%. En conjunto, esto se traduce en una reducción de costes del 10%.

#### Resumen capacidades del Sistema:

- **Uso de datos:** 64 bits, CAD para 400 Mbs y tamaño nube de puntos > 100 millones
- **Formato del archivo CAD:** Multi CAD
- **Algoritmos alineación Pieza/CAD:** Multi-alineado
- **Color de mapeado:** Multi-núcleo y texturizada
- **Visualización:** Adaptativa

## 7 Recomendaciones para el despliegue de DSES en España

### 7.1 Actores en los ecosistemas y DSES

De acuerdo con la BDVA<sup>36</sup>, el éxito en el desarrollo de un espacio de intercambio de datos que genere valor económico en la sociedad, radica no solo en su estrategia sino también en el esfuerzo conjunto de las partes interesadas. En el Position Paper *“Towards a European-Governed Data Sharing Space”*<sup>37</sup>, la BDVA, tomando como base el modelo de la triple hélice, resume los actores estratégicos (Industria, Academia y Gobierno) y los roles que éstos pueden asumir, como se muestra en la *Figura 10.*, así como las principales herramientas y mecanismos que se pueden utilizar para el desarrollo de un Espacio de Datos común.

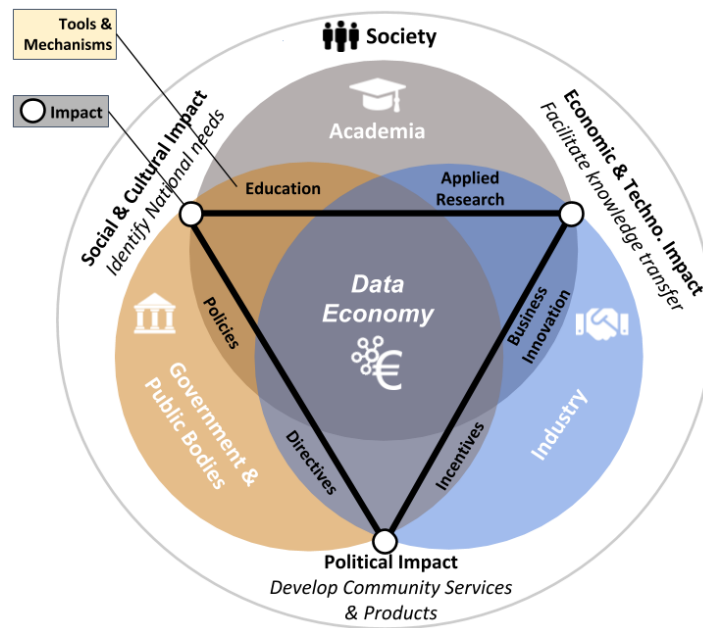


Figura 10. Actores, Roles y Herramientas. Fuente: BDVA

#### 7.1.1 Actores principales

En un Espacio de Datos podemos diferenciar una serie de actores en función de su relación con el acceso y uso de los datos del espacio en cuestión:

- Los propietarios de datos o actores que ostentan los derechos legítimos sobre ellos.
- Los poseedores de datos, que disponen de datos propios o de terceros sobre los que pueden otorgar legítimamente determinados derechos de acceso y uso a otros actores del espacio.
- Los consumidores de datos, a los que los poseedores de datos pueden otorgar derechos de acceso, uso o control.

<sup>36</sup> <https://www.bdva.eu/>

<sup>37</sup> BDVA. Towards a European-Governed Data Sharing Space. November 2020.

[https://www.bdva.eu/sites/default/files/Bdva%20DataSharingSpaces%20PositionPaper%20V2\\_2020\\_Final.pdf](https://www.bdva.eu/sites/default/files/Bdva%20DataSharingSpaces%20PositionPaper%20V2_2020_Final.pdf)

- Otros agentes que podrían proporcionar determinados servicios en el ecosistema de datos, como representación de otros actores, intermediación, provisión de servicios (registro, trazabilidad, información, procesamiento, etc.)

Dependiendo de qué organización esté desempeñando un determinado rol dentro del ecosistema del Espacio de Datos, pueden contemplarse algunas consideraciones. Por ejemplo, cuando las **Administraciones Públicas** desempeñan el rol de propietarias o poseedoras de datos, podrían, a través de este tipo de infraestructuras, hacer accesibles y reutilizables - a otras administraciones, a las empresas y a la sociedad - los datos generados con inversiones públicas sobre los que ostenten derechos, siempre que no sean personales ni estén sujetos a derechos de propiedad intelectual o industrial de terceros.

Igualmente, las personas físicas podrían poner a disposición de otros actores del Espacio de Datos sus datos no personales a través de diversos mecanismos, como la provisión altruista.

Para el caso de las empresas, podrían beneficiarse de un potencial intercambio de valor (económico o de otro tipo) a través del mercado abierto, seguro y confiable de datos que estas infraestructuras proporcionan para el acceso, uso, interoperabilidad y control de los datos.

Además, alrededor de estos roles principales pueden desplegarse otros agentes con roles de intermediación, representación, prestación de servicios asociados a los datos, etc.

### 7.1.2 Agentes del Conocimiento y otros agentes

El desarrollo de la visión de los **Espacios de Datos** como “estructuras que proporcionan confianza y seguridad para la compartición voluntaria de datos entre diversos agentes de manera homogénea a través de mecanismos combinados de gobernanza, organizativos, legales, de negocio y técnicos, que facilitan la interoperabilidad para acceder a los datos o transferirlos, y que posibilitan su reutilización eficiente y legítima en un contexto de soberanía y control para las partes sobre sus propios datos” requiere todavía superar numerosos retos. No se podrá lograr esta visión sin la cooperación activa de los responsables de las políticas relacionadas, los agentes del conocimiento, y otras infraestructuras de promoción y apoyo que contribuyan a los retos legales, técnicos y de mercado, respectivamente.

En España se están desplegando varias políticas relacionadas con la materia a través de determinadas estrategias y planes asociados. La estrategia España Digital 2025 y la Estrategia Nacional de Inteligencia Artificial son posiblemente las referencias más relevantes, como se ha señalado en apartados anteriores de este documento. A través de estas estrategias y de planes vinculados -como el Plan de Digitalización de las Administraciones Públicas 2021-2025, se está favoreciendo su desarrollo. El rol de los agentes públicos encargados de las políticas públicas en la materia y en articular los instrumentos de apoyo asociados (talento, mercado, infraestructuras de apoyo, financiación, etc.) son vitales para el éxito de su adopción en nuestro país.

Los agentes de conocimiento, como **universidades, OPIs y Centros Tecnológicos**, son también vitales para desarrollar investigaciones para los retos tecnológicos de los **DSES**. Algunos de estos desafíos han sido señalados también en la sección correspondiente de este documento con relación a la interoperabilidad abierta de datos, el uso de catálogos inteligentes, generación automática de metadatos estandarizados para los datos, conectores e interfaces, mecanismos de trazabilidad y de forzado de políticas de acceso y uso, la operación sin transferencia de datos y transferencia de algoritmos (*federated learning*), etc.

Por último, se requiere cerrar la brecha entre la complejidad de este tipo de aproximaciones y la capacidad actual de las organizaciones, sean empresas u organizaciones públicas. En este sentido,

existen determinadas infraestructuras que pueden contribuir a cerrar este gap. Las Plataformas Tecnológicas son desde luego algunas de ellas. Las asociaciones sectoriales tienen también un rol en esta función. Otros actores son los **Digital Innovation Hubs**, DIH, que sirven de ventanilla única de entrada para atender las necesidades en innovación y transformación digital de las empresas, especialmente pymes, a las que atienden.

## 7.2 Del Piloto a la Explotación.

### 7.2.1 Pasos para la creación de un Espacio de Compartición y Explotación de Datos (DSES)

La *Figura 11* muestra una vista simplificada de lo que podría ser el diseño de arquitectura inicial para un DSES, con lo que son las piezas fundamentales para llevar a cabo su despliegue. Como se puede observar, que sea inicial no implica que sea simple. A continuación, se realiza un pequeño recorrido por los distintos elementos que consideramos para esta primera aproximación, explicando su rol y el porqué de incluirlos.

Empezando desde fuera hacia dentro, lo primero que debe de encontrar un usuario o una empresa que quiera o necesite trabajar con un DSES es una **interfaz de acceso**, por ejemplo, un frontend web. La misión primera del frontend es la de informar a los usuarios, sea cual sea su naturaleza, de qué es y qué les ofrece el DSES. Además, deberá de ofrecer una serie de funcionalidades adicionales, como la de registrarse para poder hacer uso del servicio, gestionar la cuenta de usuario u ofrecer aspectos tan importantes como los términos y condiciones de uso. Sin embargo, más allá de las herramientas de gestión o informativas, el frontend contiene un **elemento fundamental en infraestructuras de este tipo: el catálogo**. El catálogo tiene tres misiones principales. Primero, informar de los servicios ofertados por el DSES de una manera clara y amigable, ofreciendo un conjunto de información básica sobre dichos servicios. Segundo, facilitar que un usuario registrado pueda solicitar y consumir estos servicios. Por último, en tercer lugar, ofrecer los medios para que usuarios registrados puedan solicitar la inclusión de servicios propios en el catálogo del DSES.

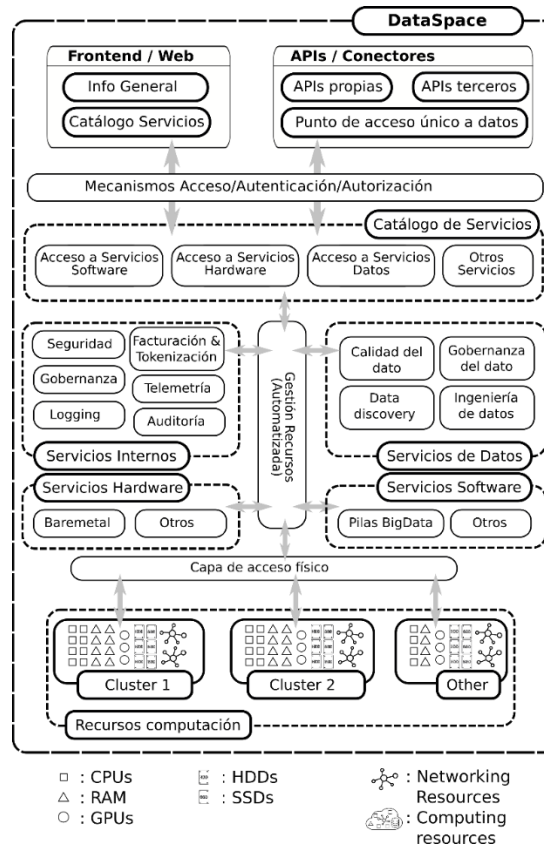


Figura 11. Posible modelo de referencia para un DSES

Ahora bien, **¿qué tipo de servicios podemos encontrar en el catálogo de un DSES?** A priori, contemplamos cuatro tipos de familias de servicios principales:

- **Servicios software:** Dentro de esta categoría se incluye, al menos, dos tipos de servicios:
  - Primero, aquellos servicios que han de ser desplegados en recursos de computación, máquinas virtuales o contenedores. Ejemplos de estos servicios podrían ser Spark, Jupyter, Hive u otros.
  - En segundo lugar, se ofrecen servicios que no requieren, necesariamente, ser ejecutados en recursos de computación del DataSpace. Posibles ejemplos serían algoritmos o modelos de inteligencia artificial o incluso contenedores que se ofrezcan en modalidad de descarga.
- **Servicios hardware:** Estos servicios engloban la solicitud de acceso a recursos de computación “baremetal” o a recursos HPC.
- **Servicios de datos:** Dentro de los servicios de datos distinguimos dos servicios distintos.
  - Primero, el **acceso y consumo de datasets** ofertados por el DSES, cumpliendo las condiciones y términos de uso estipulados por el dueño de los datos.
  - En segundo lugar, la posibilidad de **almacenamiento de datos propios** del usuario en recursos de computación propios. En este caso, el usuario podría decidir si quiere almacenarlos para uso propio o si se pueden abrir a otros usuarios. En cualquier caso, se tendrían que observar las normativas correspondientes a privacidad de datos, e.g., la GDPR.
- **Otros servicios:** En esta categoría se incluyen servicios no relacionados con el consumo de recursos de computación, como podrían ser soporte, formación o consultoría. En general, hablamos de recursos ofertados por personal o empresas asociadas al DSES.

Como se ha comentado, cada servicio ha de proveer un conjunto de información que se considere suficiente como para poder usarlo y desplegarlo, si fuera necesario. Desde nuestro punto de vista, esta información ha de incluir, al menos, una descripción del mismo, condiciones o términos de uso del servicio/dataset, restricciones (si las hubiere), un enlace que permita solicitar dicho servicio, costes asociados (si los hubiere) e instrucciones de uso. Además, en el caso de los servicios software y hardware se debería ofrecer información relativa a la disponibilidad actual, o a futuro, de los recursos de computación en el *DSES*.

Hay que tener en cuenta que todo acceso o interacción con el *DSES* no va a ser manual o basada en la interacción con el usuario. El *DSES* deberá ofrecer un conjunto de APIs y conectores que faciliten:

- a. El uso, despliegue y/o consumo de servicios o datasets de manera automatizada;
- b. su interacción con otras infraestructuras o iniciativas de interés a nivel europeo con las que se considere valiosa una interacción.

En el primer caso, y centrándonos en el caso de ingesta o consumo de datos, podría ser especialmente interesante definir un punto único de acceso a datos, lo que podríamos ver como un caso especial de API propia. Esto supondría una interfaz común o única para cualesquiera fueran los repositorios o tecnologías de almacenamiento de datos que tuviera, internamente, el *DSES*. Este aspecto se amplía en la Sección 7.2.1.1.

En el segundo caso, estaríamos hablando, por ejemplo, de ofrecer APIs o conectores de terceros que permitan obtener servicios de estas iniciativas, como sería el caso de GAIA-X o la red europea de DIHs. Ejemplos de estos conectores podrían ser herramientas para la sincronización de catálogos de servicios o conectores para el consumo de datasets, como los definidos por IDSA en su arquitectura de referencia. Esto permitiría ofrecer no sólo un catálogo de servicios ampliado incluyendo servicios ofrecidos por otras instituciones, sino incluso a infraestructura de computación con tipos específicos de hardware que no estuviera disponibles localmente (e.g., que otro centro solicitará usar GPUs locales).

Por supuesto, el acceso a servicios, datos o recursos de computación necesita de la existencia de una capa de seguridad intermedia. Por tanto, por debajo de este frontend y de estas APIs/conectores, es necesario que existan una **serie de mecanismos de acceso, autenticación y autorización**. Estos mecanismos permitirán a cada usuario registrado autenticarse y tener acceso a los servicios ofertados por el *DSES* o a un subconjunto de los mismos. El conjunto de servicios accesible por un usuario estará condicionado por el tipo de usuario. Distintos tipos de usuarios podrán tener distinto alcance y funcionalidad disponible. Esto puede ser controlado con distintos niveles de granularidad mediante herramientas de autorización basadas en roles. Aunque no entremos en detalle ahora, huelga decir que esta capa de seguridad ha de incluir una serie de herramientas que permitan prevenir o detectar problemas de seguridad comunes como intrusiones o denegaciones de servicio, entre otros.

Una vez que accedemos a los servicios, hay que ponerlos en marcha. Para ello necesitamos un módulo que nos permita hacer una **gestión de los recursos de computación**, bien de manera manual o bien automática. Este módulo ha de ser capaz de negociar o asignar a cada usuario o servicio solicitado los recursos de computación necesarios para su ejecución. Por otra parte, harán falta los servicios en sí, concretamente los servicios software, hardware y de datos más una serie de servicios internos auxiliares que permitan efectuar la gestión de recursos de manera adecuada o faciliten otros aspectos como la facturación o la auditoría. A continuación, profundizamos en los servicios de datos, por considerarlos los más relevantes en un *DSES*, conjuntamente con los internos, por ser imprescindibles para su correcto funcionamiento. Los servicios de software, e.g., pilas BigData o herramientas para visualización de datos, y hardware, e.g., préstamo de contenedores o máquinas virtuales, no los



comentaremos por entender que pueden ser muy variados y no ser el objetivo principal de este documento.

**Los servicios de datos** pueden ser, también, muy variados. Sin querer entrar en lo amplio que podría ser este abanico y dejando deliberadamente de lado posibles servicios de almacenamiento de datos o de prestación de datasets, querríamos destacar los siguientes:

- **Calidad del dato:** Definir mecanismos que nos permitan cuantificar, de manera cualitativa, la calidad de un dataset y de los datos que comprende de manera objetiva. Algunos criterios para utilizar podrían ser distribución típica de los datos y detección de outliers, existencia de campos con valores nulos, etc.
- **Data Discovery:** Servicios que permitan la detección automática de características de un dataset cuando este se incorpora al *DSES*. Por características nos referimos a generar un conjunto de metadatos descriptivos del dataset, por ejemplo, número de columnas, tipo de datos de cada columna, tamaño, número de registros, completitud y calidad del dataset, etc.
- **Gobernanza del dato:** El gobierno del dato es un aspecto crítico que ha de ser considerado, clave tanto para el proveedor como para el consumidor de datos. Aspectos como la soberanía, la trazabilidad, el linaje y proveniencia del dato dan garantías tanto del uso que se va a hacer como de el origen de los datos. La gobernanza no se limita a esto, por supuesto, aspectos como vocabularios, estructuras de metadatos para los datasets o definición de requisitos de calidad de datos para aquellos datasets que puedan o quieran ser incorporados han de ser también considerados.
- **Servicios de ingeniería de datos:** Disponer de una serie de herramientas que permitan mejorar la manera en que se usan los datasets disponibles en el *DSES*, optimizando consultas, estructuras de datos u otros aspectos relacionados. Estos servicios de ingeniería de datos permitirán que los usuarios finales no precisen disponer de los conocimientos tan especializados en el campo de ingeniería de datos.

La **sección 7.2.1.1** amplía información en cuanto a procesos relacionados con estos servicios de datos.

Pasamos ahora a los **servicios internos**. Estos servicios pueden ser también de una naturaleza variada, pero consideramos que son necesarios, al menos, los siguientes:

- **Seguridad:** Aquí ya no hablamos sólo de seguridad perimetral contra intrusiones o denegaciones de servicio. Un *DSES* ha de garantizar la seguridad e integridad tanto de los datos almacenados en su infraestructura como de los que los usuarios puedan cargar para la ejecución de servicios o experimentos. Esto aplica, especialmente a aspectos de privacidad y data sovereignty, también trabajados dentro de la gobernanza del dato, donde es especialmente importante garantizar que se cumple con las condiciones estipuladas por los dueños de los datos y servicios. Esto entronca directamente con la gobernanza.
- **Gobernanza:** Se ha de permitir la definición de políticas que condicionen cómo se ha de efectuar el acceso a recursos, cómo han de ser asignarse prioridades en la ejecución de experimentos o cuáles han de ser las políticas de permisos. De igual manera, debe de dotar de mecanismos que permitan implementar de manera adecuada distintas políticas de uso de servicios y datasets, entre otros.
- **Logging y auditoría:** Este servicio se encarga de almacenar todos aquellos eventos que se consideren relevantes, e.g., interacción con usuario, eventos de servicio o errores o mensajes de hardware. Igualmente, se encargará de almacenar y ofrecer esta información de manera que sea fácilmente consultable. Esta información puede ser utilizada, eventualmente, en

labores de auditoría o posibles resoluciones de conflictos derivadas de un mal uso de servicios o datos por parte de algún usuario.

- **Telemetría:** Proporcionará medios para estimar, medir o consultar la ocupación de los distintos recursos de computación, así como otros parámetros que se consideren de interés o que puedan afectar a la calidad del servicio ofertado. Esto es especialmente relevante de cara a ofrecer la disponibilidad actual de recursos de computación en un *DSES* en el catálogo de servicios. Igualmente, se medirán los recursos consumidos por cada usuario, así como otras variables que puedan tener un impacto en la tarificación (si la hubiera) de los servicios que tiene en ejecución o de los datasets que está consumiendo.
- **Facturación y tokenización:** Mediante la facturación se traducirán valores de telemetría u otros aspectos relevantes de un servicio en ejecución en dinero o tokens a ser abonados por los distintos usuarios. El caso de una gestión basada en pagos de dinero no necesita mayor explicación. Sin embargo, la posibilidad de usar un servicio de tokenización abre nuevas puertas en la gestión o control de acceso a recursos, servicios o datasets en el *DSES*. Una facturación basada en tokenización podría implicar, por ejemplo, que los usuarios necesitan abonar un montante de tokens para poder solicitar o consumir servicios. Aunque es posible que la adquisición de tokens se hiciera vía mecanismos de compra, también es posible tener mecanismos de generación de tokens. Esto podría utilizarse, por ejemplo, para garantizar que todos los usuarios tienen una capacidad mínima de consumir recursos cada mes, actuando también como mecanismos de control de acceso a recursos. En cualquier caso, este tipo de políticas son un ejemplo más de mecanismo a definir y configurar en el módulo de gobernanza.

Una vez que somos capaces de determinar el conjunto de recursos que han de ser asignados a una petición de servicio efectuada por un usuario, hemos de dar acceso efectivo a esos recursos. Aquí entrarían sistemas de gestión de infraestructura como servicio (IaaS) cuyos principales exponentes, a día de hoy, podrían ser Openstack o VMWare. Estas herramientas no sólo permiten asignar conjuntos de recursos a un usuario o empresa, sino también crear aquellas estructuras virtuales de red que fueran necesarias o proporcionar el aislamiento necesario entre servicios.

Por último, tenemos los recursos de computación disponibles o bajo control directo del *DSES*. Estos recursos englobarían tanto recursos de computación estándar (CPUs, RAM, HDDs, SSDs) como recursos de computación avanzada u HPC (high performance computing) como GPUs. Estos recursos son los que, al final, sustentan toda la infraestructura descrita, donde se ejecutan los distintos servicios que oferta el *DSES* y donde se residen las bases de datos que almacenan los distintos datasets.

#### 7.2.1.1 Servicios de Datos

Uno de los aspectos más importante en una plataforma de datos son los conectores que permitirán acceder a los mismos. La aparición de nuevas tecnologías de almacenamiento y procesamiento de datos ha permitido a los ingenieros de datos poder elegir entre un gran abanico de posibilidades, permitiendo así seleccionar la tecnología que mejor se adapta a cada caso de uso.

Sin embargo, esto también ha ocasionado que la complejidad de cara al usuario sea mucho mayor. La coexistencia de múltiples tecnologías en un *DSES* tiene como efectos colaterales:

- Complejidad para el usuario final:
  - Debido a que le obliga a conocer diferentes APIs y/o librerías para poder hacer uso de cada tecnología. Esto también ocasiona que la curva de aprendizaje sea muy grande.
  - Complejidad para encontrar las fuentes de datos, ya que estas se encuentran en diferentes almacenes de datos.

- Seguridad. Disponer múltiples almacenes de datos, obliga a los administradores de las infraestructuras a exponer puntos de comunicación con cada uno de ellos que, por razones obvias, tienen que estar securizados.

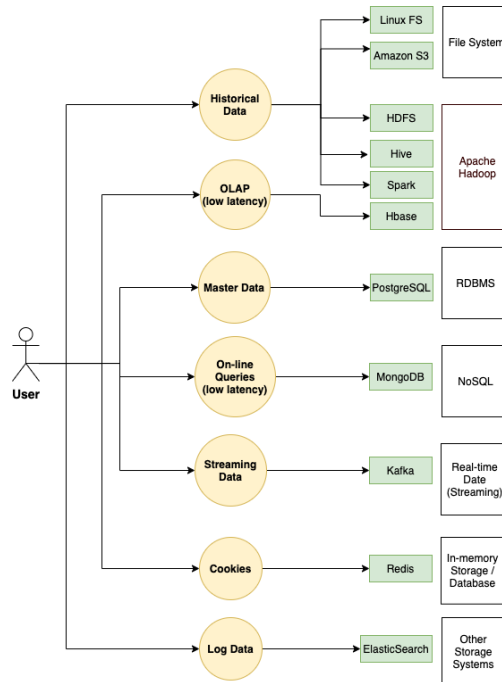


Figura 12. Complejidad en el acceso a datos.

Por el contrario, disponer de un único acceso a los datos simplifica la arquitectura de la plataforma, mejora enormemente la experiencia de usuario y facilita el rápido acceso a los datos.

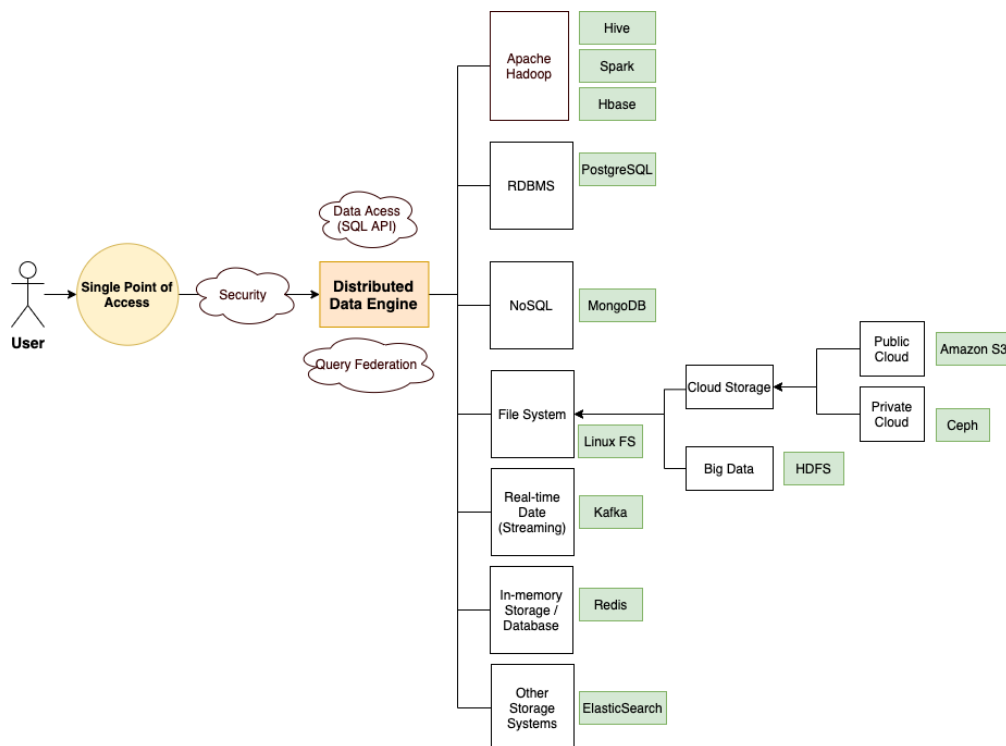


Figura 13. Simplicidad en el acceso a datos

El único punto de acceso, por otro lado, le permite conectarse a todos estos sistemas como fuente de datos. Expone los datos para consultas con SQL estándar del Instituto Nacional de Estándares Estadounidenses (ANSI) y todas las herramientas que utilizan SQL.

Una vez el dentro de la plataforma, los servicios de datos ofrecen funcionalidades de alto valor añadido y que ayudan a los usuarios menos versados en gestión de datos a introducirse en la cultura del dato, en el cambio estratégico de convertirse en una compañía Data Centric, siguiendo las mejores prácticas en ingeniería de datos y de manera segura.

En la *Figura 11* se mostró un ejemplo de los distintos tipos de servicio que puede ofrecer un DSES. En ella se mencionan los servicios de datos y, en particular, se incide en servicios relacionado con la calidad del dato, la gobernanza del dato, data discovery e ingeniería de datos. Para la construcción de estos servicios es necesario:

- La definición del Marco Global de **Arquitectura de Datos** donde se especifican:
  - Capas físicas y lógicas.
  - Políticas y estándares de almacenamiento y gestión de los datos.
  - Políticas y estándares de acceso a los datos.
- Definir e implantar los procesos de ingestión y almacenamiento de los datos.
- Definir e implantar los procesos de **Calidad del Datos**.
- Definir e implantar los procesos de normalización y transformación de los datos.
- Definir e implantar los procesos de **Data Provenance** (control del origen de los datos) y **Data Lineage** (trazabilidad del uso de los datos).
- Definir e implantar los procesos de securización y protección de los datos considerando todas las normativas relativas a privacidad.
- Definir e implantar los procesos de publicación de los datos para que sea accesibles a los usuarios a través del **Catálogo de Datos**.
- Definir e implantar el servicio de “auto descubrimiento” de datos (**Data Discovery**), que automatiza el proceso de alta y mantenimiento del catálogo de datos.

## 7.2.2 DSES sectoriales

### 7.2.2.1 DSES sector Manufactura

La industria manufacturera está inmersa en un proceso de digitalización que está transformando la forma en que se producen los productos, en particular continúa avanzando hacia la fabricación digital y conectada, impulsada por la cuarta revolución industrial. La I4.0 está marcando una aceleración rápida hacia las factorías digitales, conectadas y autónomas que permitirán producir de forma más eficiente y ágil nuevos productos. La mayor disponibilidad de datos impulsada por la sensorización, los sistemas ciber-físicos y las plataformas IoT, los cuales, combinados con la inteligencia artificial y la robótica, abordan el reto de conseguir que se desarrollen procesos flexibles e inteligentes, que habiliten la toma de decisiones en tiempo de ciclo, promuevan la hiperautomatización y potencien la gestión predictiva del mantenimiento, la calidad y el uso de recursos.

#### **Principales aplicaciones y beneficios para el sector.**

Los datos de planta deben retroalimentar los procesos a lo largo de toda la cadena de valor y contribuir a la virtualización de las plantas de producción. Es fundamental avanzar en herramientas predictivas que permitan maximizar el uso y vida útil de equipos y sistemas, asegurar la calidad del producto y gestionar de forma óptima y predictiva los recursos. Asimismo, el acceso a los datos es fundamental para garantizar la competencia y crear nuevas oportunidades de negocio, principalmente para las PyMES industriales. Los Espacios de Datos industriales permitirán avanzar hacia un modelo de

soberanía tecnológica basada en herramientas para compartir datos en su origen y desplegar una nueva generación de la IA Industrial que permita que los tejidos industriales sean capaces de dar respuesta ágil a escenarios cambiantes e inestables.

La **cadena de valor en el intercambio de datos** del sector involucra:

- Empresas Industriales: data owners.
- Proveedores: data providers.
- Prestadores de servicios: integradores y desarrolladores de aplicaciones.
- Consumidores/clientes: usuarios de datos y beneficiarios.
- Agentes tecnológicos y de conocimiento: consumidores de datos, desarrolladores de aplicaciones.
- Comercializadores: usuarios de datos, data providers.
- Transportistas, distribuidores, logística: data providers.
- Administraciones: crear condiciones favorables para el flujo del dato no personal.

**Herramientas o plataformas de datos de referencia** en el sector:

- Eurecat<sup>38</sup>: BigDataCOE (BDVA iSpace), centro de excelencia para desarrollar y proporcionar herramientas, conjuntos de datos y capacidades de análisis de datos donde las empresas comparten datos, prueban y validan modelos analíticos. Cuenta con un ámbito prioritario industrial con disponibilidad de plataformas de datos de diferentes procesos (por ejemplo: inyección, extrusión, fresado, estampado, soldadura, entre otros) y maquinarias.
- Open Manufacturing<sup>39</sup> es un entorno basado en la nube donde las grandes empresas industriales y de la cadena de suministro pueden compartir componentes y datos de código abierto para desarrollar soluciones innovadoras.
- <https://www.internationaldataspaces.org/>International Data Space Association<sup>40</sup>: Con el establecimiento de la Asociación Internacional de Espacios de Datos, las empresas y los centros de investigación participan activamente en el diseño de una arquitectura confiable para la economía de datos (incluida la industria). IDSA tiene como objetivo garantizar la soberanía de los datos mediante una arquitectura abierta.

**Datos Abiertos existentes:**

- The open industrial data project<sup>41</sup>
- Prognostics Center of Excellence Datasets<sup>42</sup>:
- data.europa.eu<sup>43</sup>
- También se hace uso de buscadores, catálogos y agregadores específicos, por ejemplo, el motor de búsqueda de Google, Kaggle<sup>44</sup>: y también Uearthed<sup>45</sup>:

Las **principales legislaciones o regulaciones específicas** para el sector en cuanto a tratamiento de datos son: La Estrategia europea de datos propone establecer los Espacios de Datos específicos de un sector o dominio, mediante acuerdos concretos para el intercambio de datos tanto a nivel local, como

<sup>38</sup> <https://eurecat.org/es/>

<sup>39</sup> <https://open-manufacturing.org/>

<sup>40</sup> <https://www.internationaldataspaces.org/>

<sup>41</sup> <https://openindustrialdata.com/>

<sup>42</sup> <https://ti.arc.nasa.gov/tech/dash/groups/pcoe/prognostic-data-repository/>

<sup>43</sup> [www.data.europa.eu](http://www.data.europa.eu)

<sup>44</sup> <https://www.kaggle.com/>

<sup>45</sup> <https://unearthed.solutions/>

regional y europeo. Un Espacio de Datos común estará compuesto por un entorno de TI seguro para el procesamiento de datos por parte de un número abierto de organizaciones, y un conjunto de reglas de naturaleza legislativa, administrativa y contractual que determinan los derechos de acceso y procesamiento de los datos. Los datos estarán disponibles de forma voluntaria y se podrán reutilizar según la decisión del titular de los datos. A la fecha de elaboración del presente documento, los instrumentos legislativos están en definición. Los borradores proponen un marco general que abarca medidas horizontales relevantes para todos los Espacios de Datos europeos. Asimismo, el marco legislativo dejará espacio para reglas, mecanismos de gobernanza y estándares específicos del sector cuando sea relevante. El objetivo de la iniciativa legislativa no es crear Espacios de Datos europeos comunes por ley, sino mejorar su desarrollo reforzando la confianza en el intercambio de datos y en los intermediarios de datos.

- Proposal for a REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL on European data governance (Data Governance Act) - SWD(2020) 295 final.
- COM(2020) 767 final, 2020/0340 (COD)

Estándares de interoperabilidad:

- Reference Architecture Model Industry 4.0 (RAMI 4.0)
- OPC Unified Architecture
- OPC UA
- Reference Architecture Model for the Industrial Data Space
- MQTT, CoAP, oneM2M, SAREF...

Dentro de **las barreras y retos** a los que se enfrentan este tipo de espacios en el sector son:

- Concreción de un marco regulador correcto en materia de gobernanza de datos, del acceso a ellos y de su reutilización entre empresas, administraciones y todos los actores de cada cadena de valor.
- Dotar de incentivos para el intercambio de datos y adoptar normas prácticas, justas y claras para el acceso a los datos y el uso de estos.
- Aumentar la disponibilidad de los datos industriales abiertos y armonizados para fomentar su interoperabilidad a nivel sectorial y transectorial.
- Fomentar inversiones en nuevas tecnologías e infraestructuras digitales.
- Reducir las barreras al uso de tecnologías de IA y su convergencia con otras tecnologías para promover su aplicación innovadora.
- Formar a la próxima generación de ingenieros, técnicos, investigadores en una cultura de datos y su explotación
- Impulsar educación en ciencia y tecnología práctica, con énfasis en la ciencia de datos aplicada y la informática industrial, para garantizar que los trabajadores, en toda la cadena operacional, sean capaces de aprovechar al máximo las oportunidades de la revolución digital impulsada por los datos

#### 7.2.2.2 DSES sector Salud

La necesaria transformación digital del sistema sanitario se ha puesto de manifiesto durante la reciente crisis de la COVID-19, habiéndose dado un gran impulso a soluciones de e-salud. En este contexto, el Ministerio de Sanidad en España trabaja en el desarrollo de una Estrategia de Salud Digital del SNS en la que se contará con la colaboración de las Comunidades Autónomas y con la participación de otros actores públicos y privados implicados.

Respecto a las tecnologías médicas y sanitarias -salud digital- el sector agrupa aquellos productos y servicios para el cuidado de la salud, que ayudan en el diagnóstico, vigilancia o tratamiento de una enfermedad o condición similar. La tecnología sanitaria mejora la calidad de vida de aquellos usuarios (pacientes) que las utilizan. Gracias a la participación de la ciencia y la tecnología en el sector los pacientes disfrutan de diagnósticos más precoces, tratamientos menos invasivos, reducción de períodos de tratamiento y estancias en los hospitales. Dentro de la tecnología sanitaria se encuentra la salud digital, que hace referencia a todos aquellos productos o servicio fruto de la aplicación de las TICs en el campo de la salud. El sector de la salud digital es muy amplio y extenso, incluyendo aplicaciones que van desde la teleconsulta y el telediagnóstico, a la robotización y el uso de la inteligencia artificial, pasando por la inclusión del Internet of Things en los dispositivos médicos. En el contexto de la salud digital los datos tienen un valor capital. Un adecuado Espacio de Datos de salud puede traer numerosos beneficios para el sector en su conjunto. Por ello, desde la Comisión Europea se apremia a la creación de un Espacio de Datos de salud europeo, anónimos y adecuadamente cifrados.

Así mismo, un intercambio insuficiente de los datos clínicos afecta negativamente a los servicios prestados a la ciudadanía. Existen diferencias entre los países y regiones europeas en la infraestructura IT, esencial para el intercambio de la información y datos clínicos. En muchas ocasiones supone una gran dificultad para los pacientes para acceder a sus propios datos en salud. Como ejemplo, la Historia Clínica Electrónica (HCE) no es una realidad en la UE. Diversidad y disparidad europea en estándares e interoperabilidad digital entre los sistemas sanitarios. Acceso muy limitado a datos para investigación e innovación, diseño de políticas y propuestas de regulaciones y normativas. Debe hacerse siguiendo las normativas europeas y estatales de protección de datos, alineando la normativa de los estados miembros de la UE en el acceso e intercambio de datos en salud. La elaboración de políticas y regulaciones se vería altamente beneficiada del acceso a datos de calidad, precisos y representativos.

El uso de herramientas digitales ha aumentado notablemente como consecuencia de la pandemia COVID-19, pero ese impacto aún no se ha traducido a los servicios sanitarios. Un número cada día mayor de herramientas para la sanidad incluyen la Inteligencia Artificial, cuyos beneficios se van describiendo cada día; se hace necesaria una regulación al respecto y una carta de derechos y deberes. Por tanto, la solución radica en la elaboración de un marco legal para el Espacio Europeo de Datos en Salud, que:

- Asegure acceso, compartición y uso óptimo de los datos en salud para propósitos sanitarios, I+D+i, diseño de políticas y regulaciones. A alcanzar a través de un marco legal y de gobernanza adecuado para el acceso e intercambio de los datos, reducción de las barreras tecnológicas para el manejo de datos (infraestructura IT, interoperabilidad, calidad del dato, estándares) y garantizar el acceso y control de los pacientes sobre sus propios datos.
- Promoción de un mercado único de salud digital, de productos y servicios
- Promoción del desarrollo e implementación de productos y soluciones de salud digital robustos, eficientes y de confianza.

***Principales aplicaciones y beneficios para el sector:***

- *Beneficios Económicos:*
  - Beneficios para los pacientes y ciudadanos, profesionales sanitarios, legisladores, investigadores e innovadores de un marco común y a gran escala sobre legislación, gobernanza, calidad del dato e interoperabilidad.
  - Mejora de la sostenibilidad, reducción del gasto y aumento de la eficiencia de los sistemas sanitarios, gracias a la transformación digital.

- Facilitar el desarrollo de innovaciones en el área de la salud digital, que avancen hacia sistemas de salud más eficientes y sostenibles, favorezcan la medicina personalizada, mejoren la monitorización, y la seguridad e integridad de los pacientes.
- Acercamiento hacia un mercado único de datos en salud, reduciendo las barreras entre países de la UE y facilitando el acceso a PYMES.
- **Impactos Sociales y Derechos Fundamentales:**
  - Nuevas herramientas de soporte para estrategias colectivas y tratamientos o dispositivos médicos nuevos, que mejorarán la calidad de vida de los pacientes.
  - Mejora del acceso a nuevas tecnologías y facilita la implementación de herramientas, en especial la IA.
  - Reducción de pruebas de diagnóstico innecesarias y mejora del control en los datos clínicos, junto con una mejora de su seguimiento.
  - Garantías del derecho a la privacidad y protección de datos de los individuos. Ampliación a la movilidad europea de los ciudadanos.
- **Impactos Medioambientales:**
  - La interoperabilidad, uso de datos ya consensuados y la interoperabilidad de estos, promoverá la reducción de pruebas y recursos innecesarios. Todo ello, con una reducción de la huella medioambiental de los sistemas sanitarios.
- **Impacto en las Administraciones:**
  - Reducción la carga administrativa para profesionales de I+D+i, profesionales sanitarios y autoridades, con una reducción de la descentralización.
  - Mejora de la transmisión de los datos, y mejora en los recursos para el diseño de políticas.

La **cadena de valor en el intercambio de datos** involucra:

- Datos personales codificados pertenecientes a pacientes, para su utilización en investigaciones futuras relacionadas sin necesidad de contar con el consentimiento de los participantes, pero sin menoscabo de las necesarias garantías y derechos de los pacientes.

**Herramientas o plataformas de datos de referencia** en el sector:

- Harmony y Harmony Plus (*ver en la sección 6.2.2*) dos proyectos IMI (de colaboración público-privada entre EFPIA y la Comisión Europea), son claros ejemplos de cómo es posible realizar investigación biomédica a partir de datos de investigaciones anteriores, cumpliendo con todas las necesarias garantías descritas en la legislación vigente, relativas a protección de datos y autonomía de los pacientes.

Las **principales legislaciones o regulaciones específicas** para el sector en cuanto a tratamiento de datos son:

- GDPR y LOPDyGDD.
- Ley 14/2007, de 3 de julio, de investigación biomédica
- Ley 41/2002, de 14 de noviembre, básica reguladora de la autonomía del paciente y de derechos y obligaciones en materia de información y documentación clínica.
- Real Decreto 1716/2011, de 18 de noviembre, por el que se establecen los requisitos básicos de autorización y funcionamiento de los biobancos con fines de investigación biomédica y del tratamiento de las muestras biológicas de origen humano, y se regula el funcionamiento y organización del Registro Nacional de Biobancos para investigación biomédica.



- 'Healthdata29: Guía legal y repositorio para fomentar la compartición de datos de salud', que ha recibido el pasado 28/01/21 el premio de la AEPD a la Proactividad y Buenas Prácticas.
- Código Tipo en el sector farmacéutico y su actualización mediante el nuevo Código de Conducta de protección de datos en el ámbito de la investigación clínica y la farmacovigilancia (actualmente en evaluación por la AEPD)

Si bien los datos personales en salud son especialmente sensibles, y se maneja un alto grado de incertidumbre, es posible realizar investigaciones en este ámbito. Para ello, y con la intención de aportar mayor seguridad a la hora de llevar a cabo dichas investigaciones, Farmaindustria ha trabajado activamente en el nuevo Código de Conducta de protección de datos en el ámbito de la investigación clínica y la farmacovigilancia, que ofrece máximas garantías para los pacientes, resto de intervinientes, y que se encuentra en evaluación por la Agencia Española de Protección de Datos.

### 7.2.2.3 DSES sector Pesquero y Acuícola

La Unión Europea (UE) es el segundo mercado de productos de la pesca y la acuicultura del mundo después de China. En este ecosistema, España ocupa el primer puesto en variedad de pescado en la UE y en el mundo, con más de 1000 especies pesqueras de interés comercial. Además, España es el primer productor industrial de la UE en productos de la pesca con aproximadamente el 20% de la producción, tanto en volumen como en valor.

En relación con la acuicultura, España es hoy uno de los principales países productores de acuicultura en Europa, alcanzando en 2018 unos ingresos en primera venta de 472,3 millones de euros, destacando como referencia a nivel europeo y mundial por la cantidad y calidad de los moluscos producidos, sin menoscabo de otras especies, como el rodaballo, o el besugo, cuya producción europea se produce únicamente en Galicia.

El sector transformador y comercializador de productos del mar y de la acuicultura de la Unión Europea (UE) tiene una gran importancia y representatividad socioeconómica dentro del conjunto del sector pesquero. Concretamente, con una producción de 359.081 toneladas y con un valor superior a los 1.750 M€, España es una referencia mundial en la producción de conservas de pescado y marisco. Respecto a la comercialización de productos de la pesca y la acuicultura, España es el segundo consumidor de productos pesqueros de la UE y el quinto a nivel mundial, con un consumo en hogares en el año 2019 de 22,53Kg por persona y año, con un gasto per cápita de 195,06 €/kg (Informe del Consumo de la Alimentación en España 2019. MAPA) 46 .

#### **Principales aplicaciones y beneficios para el sector.**

Mediante el Big Data y la Inteligencia Artificial se puede alcanzar una pesca más eficiente y sostenible. La monitorización electrónica a bordo permite obtener unos datos que pueden emplearse para mejorar la gestión pesquera, ayudando, por un lado, a la sostenibilidad de la flota al optimizar los recursos pesqueros al máximo, y por otro, al cumplimiento de las regulaciones de pesca. Además, con los datos obtenidos bajo la monitorización electrónica se puede aumentar la cobertura de la vigilancia científica, lo que facilitaría la obtención de más datos para una ordenación pesquera sostenible.

Mediante el análisis de indicadores medioambientales se pueden redefinir objetivos científicos y de gestión, y hacer predicciones masivas de cambio de distribución de especies bajo escenarios de cambio climático y presión pesquera. En cuanto a la acuicultura, se puede lograr una optimización de la producción en instalaciones acuícolas, por ejemplo, con sistemas de conteo y clasificación automática de organismos basado en análisis de imagen e inteligencia artificial. Además, dotando a los productos

<sup>46</sup>[https://www.mapa.gob.es/eu/alimentacion/temas/consumo-tendencias/informe2019\\_v2\\_tcm35-540250.pdf](https://www.mapa.gob.es/eu/alimentacion/temas/consumo-tendencias/informe2019_v2_tcm35-540250.pdf)

pesqueros de información, a través de los datos (por ejemplo, con un código qr), se puede mejorar la confianza del consumidor y convertir la compra en una experiencia de consumo.

Toda la cadena de valor de la pesca y la acuicultura puede generar y aportar datos. A su vez se beneficiarían de estos datos todas las entidades del sector, y en última instancia el propio consumidor.

**Herramientas o plataformas de datos de referencia** en el sector:

- Portal de datos abiertos<sup>47</sup> de la Unión Europea
- “Data Collection<sup>48</sup> Framework” de la Unión Europea
- Programa Nacional de Datos Básicos del sector pesquero español<sup>49</sup> (PNDB)

**Datos abiertos existentes:**

Para preservar los recursos pesqueros y su explotación sostenible se establece un marco comunitario (en inglés, Data Collection Framework), para la recopilación, gestión y uso de los datos del sector pesquero y el apoyo al asesoramiento científico, en relación con la Política Pesquera Común. Los estados miembros deben recoger datos biológicos, técnicos, ambientales y socioeconómicos del sector pesquero en el marco de un programa nacional plurianual elaborado en consonancia con el programa comunitario. La información que se obtiene de estos programas son los datos básicos necesarios para el asesoramiento científico en la gestión sostenible de los recursos pesqueros.

Las **principales legislaciones o regulaciones específicas** para el sector en cuanto a tratamiento de datos son: Reglamento (UE) 2017/1004 del Parlamento Europeo y del Consejo, de 17 de mayo de 2017, relativo al establecimiento de un marco de la Unión para la recopilación, gestión y uso de los datos del sector pesquero y el apoyo al asesoramiento científico en relación con la política pesquera común y por el que se deroga el Reglamento (CE) n.o 199/2008 del Consejo

Dentro de **las barreras y retos** a los que se enfrenta el sector se encuentra por ejemplo la necesidad de conseguir generar más datos, pues actualmente la generación de datos es muy escasa. Frente al océano que tenemos delante, solo estamos viendo el 6%. Además, faltan bases de datos estructuradas y con unos datos estandarizados.

Estamos ante un sector muy heterogéneo, donde en algunas empresas falta digitalización, antes de poder empezar con la generación de datos. Otra barrera que poco a poco se va superando es la falta de confianza del sector, y el miedo a perder la ventaja competitiva que podría suponer el compartir los datos. Esto lleva a que actualmente se generen pocos datos agregados.

#### 7.2.2.4 DSES Sanidad Animal

La sanidad animal se considera un factor clave, tanto en el mantenimiento del bienestar de los animales de compañía, como para el desarrollo de la ganadería, y es de vital trascendencia tanto para la economía nacional como para la salud pública, así como para el mantenimiento y conservación de la diversidad de especies animales. Para la salud pública, por la posible transmisión de enfermedades de los animales al hombre, y por los efectos nocivos que para éste puede provocar la utilización de determinados productos con el fin de aumentar la productividad animal.

<sup>47</sup> <https://data.europa.eu/euodp/es/home>

<sup>48</sup> <https://datacollection.jrc.ec.europa.eu/index.html>

<sup>49</sup> <https://www.mapa.gob.es/es/pesca/temas/proteccion-recursos-pesqueros/programa-nacional-datos-basicos/>

### **Principales aplicaciones y beneficios para el sector:**

Se está trabajando activamente en la digitalización de las granjas para tener un control más exhaustivo de los animales. Asimismo, se están implantando tecnologías en mascotas, con el objetivo de obtener datos más precisos de su estado de salud, movimiento, etc.

La **cadena de valor en el intercambio de datos** involucra:

- Desarrolladores de aplicaciones
- consumidores de datos
- usuarios de datos
- beneficiarios.

### **Herramientas o plataformas de datos de referencia en el sector:**

Vetinnova es un portal web, desarrollado por la Fundación Vet+i, Plataforma Tecnológica Española de Sanidad Animal orientado a facilitar y fomentar el networking y la colaboración entre todos los agentes públicos y privados del ámbito de la I+D+i en sanidad animal en nuestro país: las empresas de la industria de sanidad animal, los Organismos Públicos de Investigación, los grupos de investigación de Universidades, centros científicos y tecnológicos, otros centros privados de investigación, etc.

Esta herramienta, pionera en el ámbito de la sanidad animal en España, constituye una importante base de datos en materia de I+D+i que contribuye a potenciar la visibilidad del conocimiento científico y tecnológico disponible en este campo y fomentar su transferencia al tejido productivo para que éste desarrolle nuevos productos y servicios innovadores. Además, Vetinnova va fomenta la identificación de nuevas oportunidades de cooperación e internacionalización del sistema ciencia-tecnología-empresa dado que el portal web está traducido al inglés. Resulta crucial que la cooperación entre la comunidad generadora de conocimiento y las empresas se produzca en ambos sentidos y, por ello, Vetinnova también brinda la oportunidad para que las empresas puedan dar a conocer sus áreas de actividad y líneas de interés para establecer colaboraciones con otras entidades.

### **Datos abiertos existentes:**

- Listados de Proyectos
- Líneas de Investigación
- Servicios
- Patentes, etc.
- Grupos de Investigación
- empresas del sector.

No se han determinado una **legislación específica para el sector** a este respecto, pero se aplican las directrices de la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales.

Una de las **principales barreras** en el sector es la baja capacidad de tecnificación de las granjas y clínicas veterinarias.

#### **7.2.2.5 DSES sector Envase y embalaje**

En lo que concierne a la **cadena de valor del envase y el embalaje** conviene destacar que una mayor facilidad en el acceso a los datos se plantea como premisa fundamental para abordar – entre otros – los diferentes retos medio ambientales planteados en la iniciativa del Pacto Verde europeo<sup>50</sup>, que

<sup>50</sup>[https://ec.europa.eu/info/strategy/priorities-2019-2024/european-green-deal\\_es](https://ec.europa.eu/info/strategy/priorities-2019-2024/european-green-deal_es)

pretende convertir a Europa en el primer continente neutralmente climático en el año 2050 y que además contribuirá a conformar una sociedad más próspera y sostenible.

**Principales aplicaciones y beneficios para el sector.**

La implementación de un Espacio de Datos común que permita alojar diferentes aplicaciones orientadas a la circularidad contribuirá sustancialmente a la creación de cadenas de valor circulares a lo largo de la cadena de suministro, especialmente en el **sector del packaging** que ha sido considerado como objetivo primordial dentro del *Plan de Acción de Economía Circular*. En este sentido se contempla el desarrollo de los “pasaportes de producto” digitales, que incluirán amplia información sobre su trazabilidad a lo largo de todo su ciclo de vida incluyendo también su gestión final.

Las **principales legislaciones o regulaciones específicas** para el sector hasta ahora abordadas son:

- The European Data Strategy<sup>51</sup>
- European Green Deal<sup>52</sup>
- Estrategia Española de Economía Circular<sup>53</sup>
- Sustainable Products Initiative<sup>54</sup>

**7.2.2.6 DSES sector Turismo**

Según el Barómetro OMT (Organización Mundial del Turismo) sobre el turismo mundial, en 2019 el tráfico de turistas internacionales alcanzó la cifra de 1.460 millones, lo que supone un incremento del 4% con respecto al año anterior. El hito permitió a España batir su propio récord y seguir siendo el segundo país del planeta que más visitantes extranjeros recibe (83,7 millones de turistas), superado sólo por Francia, y manteniendo esa misma posición en cuanto a ingresos por turismo (92.337 millones de euros) por detrás, únicamente, de Estados Unidos. También en 2019 el turismo representó el 12,4% del PIB español y supuso 2,72 millones de puestos de trabajo, convirtiéndose así en el principal motor económico del país.

Sin embargo, el año 2020 supuso un año catastrófico para el turismo a causa del Covid-19. Según el INE, durante el año 2020 visitaron España 19 millones de turistas, lo que indica un descenso del 77,3% respecto al año anterior. Esto supuso un gasto total realizado por los turistas no residentes en España de 19.740 millones de euros, es decir un descenso del 78,5% respecto al de 2019.

De cara a 2021, expertos de la OMT no creen que se produzca una reactivación turística notable hasta el segundo semestre del año. Más a largo plazo, los escenarios de la OMT indican que el turismo internacional podría tardar entre dos años y medio y cuatro en volver a los niveles de 2019.

En este contexto, la pandemia ha acelerado la necesidad de hacer llegar la transformación digital tanto de empresas turísticas como de los destinos de nuestro país. La digitalización será el pilar bajo el cual se pueda mejorar la experiencia del turista en el destino, impulsar la competitividad de España a nivel mundial y progresar en el ámbito de la sostenibilidad. Un país como España, con una alta dependencia del turismo, como ha plasmado el efecto de la pandemia producida por la COVID19, necesita conocer

<sup>51</sup>[https://ec.europa.eu/commission/presscorner/detail/en/fs\\_20\\_283](https://ec.europa.eu/commission/presscorner/detail/en/fs_20_283)

<sup>52</sup>[https://ec.europa.eu/info/strategy/priorities-2019-2024/european-green-deal/actions-being-taken-eu\\_en](https://ec.europa.eu/info/strategy/priorities-2019-2024/european-green-deal/actions-being-taken-eu_en)

<sup>53</sup>[https://www.miteco.gob.es/es/calidad-y-evaluacion-ambiental/temas/economia-circular/espanacircular2030\\_def1\\_tcm30-509532.PDF](https://www.miteco.gob.es/es/calidad-y-evaluacion-ambiental/temas/economia-circular/espanacircular2030_def1_tcm30-509532.PDF)

<sup>54</sup><https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12567-Sustainable-products-initiative>

al detalle lo que sucede en tiempo real en sus destinos, pero lo que es más importante, conocer al turista.

Las decisiones turísticas de calado, tanto por parte de las empresas del sector como por parte de los destinos, no puede basarse en percepciones, sino que precisa de datos reales, para tomar las decisiones adecuadas en cada momento, pero también, para poder realizar predicciones que permitirán a las empresas adelantarse a cualquier situación futura, y nuevas tendencias.

Los datos no son ajenos al sector turístico, pero al contrario de lo que pueda ocurrir en otros sectores, la gran mayoría de los mismos se gestionan por parte de empresas privadas, muchas de ellas, fuera del territorio español. Ello ha llevado a una concienciación, cada vez más importante, sobre la necesidad de gestionar todos los datos que se producen en un territorio

La industria turística genera una cantidad ingente de datos, que gracias a tecnologías BigData y al cada vez más frecuente uso de la Inteligencia Artificial, ofrece grandes oportunidades al sector turístico. Datos turísticos que, combinados con otros, como los meteorológicos, de movilidad, etc, ofrecen incontables beneficios y nuevos modelos de negocio para la industria turística.

#### ***Principales aplicaciones y beneficios para el sector.***

Gran número de herramientas tecnológicas para el sector turístico, en especial para el hotelero y la gestión de destinos, tienen como base de desarrollo la IA y el Big Data. Estas tecnologías permiten optimizar los recursos, pero, sobre todo, tratar al turista de una forma muy personalizada y adaptada a sus necesidades.

De esta forma, el principal beneficio para el sector es la personalización. Esta es, sin duda alguna, es la clave para la mejora de la competitividad de las empresas y de los destinos turísticos. La personalización permitirá al sector el poder ofrecer la experiencia deseada por el turista y lograr su plena satisfacción a la vez que su fidelización. Esto se basa en un profundo análisis de todos los datos que genera un turista, antes de su visita, durante su visita y ya de regreso (el famoso, antes, durante y después). Conocer en detalle los gustos, preferencias, inquietudes, etc. de un turista, nos va a permitir poder ofrecerle una experiencia a medida, e incluso adelantarnos a sus deseos. A su vez, el destino podrá gestionarse de una manera más eficiente, poniendo a disposición sus servicios en función de toda esa información.

Gracias a la gestión de los datos, empresas y destinos pueden saber con detalle qué busca un turista y ofrecérselo. El big data y la IA son pilares básicos para la transformación de un territorio en destino turístico inteligente (DTI).

Sin embargo, las ventajas de la Inteligencia artificial y el big data aplicadas al sector turístico son numerosas y se pueden resumir en:

- Mayor eficiencia de los procesos y tareas
- Mayor conocimiento del cliente y por tanto permite ofrecer servicios más adecuados
- Reducción de los gastos e Incremento de la rentabilidad
- Permite descubrir nuevas tendencias y poder preverlas
- Mayor eficiencia en la gestión
- Accesibilidad a la información
- Ahorro de tiempo
- Mejorar atención al turista
- Mejora de la reputación de la empresa y el destino
- Conocer al detalle qué sucede a nivel de destino y empresa turística

- Identificar nuevos modelos de negocio basados en datos.

A nivel de destino, el poseer datos en tiempo real, supone el poder gestionar de manera más eficiente todos los recursos del destino, desde el transporte, hasta la gestión de residuos, ya que un destino turístico necesita conocer al detalle los flujos de turistas que se mueven por su territorio para poder ofrecer durante todo el año, la misma calidad de servicios a sus ciudadanos y turistas. Para ello, necesita conocer el número de turistas, sus preferencias, cómo se mueven, qué consumen, qué actividades realizan, y en función de ello, determinar sus necesidades en cuanto a gestión de agua, de efectivos policiales, de medios de transporte, etc. Y a su vez, poder ofrecer todos los atractivos del destino adaptados a las necesidades de los turistas, adaptados a las tendencias, como, por ejemplo, en estos momentos, con rutas digitalizadas que eviten masificaciones para poder ofrecer experiencias seguras, informando de rutas alternativas que eviten aglomeraciones y poder realizar recomendaciones en función de las preferencias de los turistas.

Todo ello, supone ir más allá de la simple recogida y gestión del dato, ya que se trata de dotar de inteligencia turística a todo el destino.

#### ***Cadena de valor en el intercambio de datos:***

La complejidad del sector turístico se basa en la heterogeneidad de empresas de diverso tamaño que forman parte de su cadena de valor. Esta complejidad, se traslada al mundo de la gestión de los datos de la propia industria turística, donde las heterogeneidades de las fuentes de datos, se mezcla con la necesaria colaboración público-privada y la dificultad añadida de tratar una gran cantidad de datos personales que deben ser debidamente protegidos. Los principales actores de esta cadena son:

- Proveedores de tecnología (parte de la oferta)
- Empresas turísticas, por la parte de la demanda (agencias de viajes online y tradicionales, transporte de pasajeros, hoteles y alojamientos turísticos, restauración, actividades turísticas.)
- Administración pública (a todos los niveles) y entes gestores del destino.
- Universidades y entidades de formación.
- Centros de I+D+i
- Federaciones, asociaciones, clústers y otras agrupaciones.
- Ciudadanos y turistas.

#### ***Herramientas o plataformas de datos de referencia en el sector:***

- DATAESTUR<sup>55</sup>, una iniciativa puesta en marcha por la Secretaría de Estado de Turismo a través de SEGITTUR<sup>56</sup>, consiste en un site único de datos del turismo nacional en el que se puede acceder a las cifras procedentes de distintas fuentes de información. Los objetivos de DATAESTUR son:
  - Generar un mejor conocimiento público de las principales magnitudes del turismo;
  - Identificar las fuentes públicas y privadas que pueden ser de utilidad para conocer el sector o tomar decisiones sobre el mismo;
  - Diseñar un sistema de producción y acceso a los datos ligado a la gestión de los problemas públicos de los destinos.

<sup>55</sup><https://www.dataestur.es/>

<sup>56</sup><https://www.segittur.es/transformacion-digital/proyectos-transformacion-digital/sistema-de-inteligencia-turistica-2/>

- DATAESTUR ofrece datos públicos de turismo procedentes de fuentes tradicionales como son el Instituto Nacional de Estadística, el Banco de España, AENA, Puertos del Estado o Turespaña, así como de datos privados procedentes de entidades como Telefónica. Los datos estarán estructurados en cinco categorías:
  - Datos generales (llegadas de turistas internacionales)
  - Economía (gasto turístico, aportación al PIB, empleo)
  - Transporte (pasajeros aéreos, tráfico de pasajeros por puertos, trenes y carreteras)
  - Alojamiento (ocupación hotelera, precio de alojamientos)
  - Conocimiento (escucha activa, SICTUR y revistas científicas de turismo).

DATESTUR también cuenta con área de conocimiento donde se publican informes de escucha activa con información sobre lo que se dice en redes sociales de España como destino turístico, así como intereses de los turistas a la hora de viajar a España. Esta iniciativa forma parte de un proyecto más complejo en el que el análisis de los datos constituye la base para mejorar el conocimiento del turista, a la vez que contribuirá a incrementar la competitividad de España como destino turístico.

#### **Datos abiertos existentes:**

El open data relacionado con el turismo, suele identificarse directamente con los datos de transporte público, pero las diferentes consejerías o institutos de turismo de las Comunidades autónomas, disponen de diferentes bases de datos abiertos sobre la situación del turismo en su región. En resumen:

- Datos ofrecidos por los diferentes entes gestores de la Comunidad Autónoma y sociedades de promoción de destinos turísticos.
- Datos recopilados por centros tecnológicos turísticos e iniciativas público-privadas.
- Información de la actividad turística de España.
  - ✓ FRONTUR: datos de movimientos turísticos en fronteras<sup>57</sup>
  - ✓ EGATUR<sup>58</sup>: Encuesta de gasto turístico
  - ✓ Encuesta de turismo de residentes
  - ✓ Alojamiento (encuesta de ocupación e índice de precios): Hoteles, Campings, Apartamentos turísticos: Alojamientos de turismo rural; Albergues
  - ✓ Cuenta satélite del turismo de España
  - ✓ Transporte de pasajeros

Las **principales legislaciones o regulaciones específicas** para el sector son:

- RGPD de la UE
- Ley Orgánica de Protección de Datos y Garantía de Derechos Digitales (LOPDGDD)
- Directiva europea PSD2 (Payment Services Directive)

Uno de los **principales retos** para el sector es la compartición de datos por parte de todas las empresas que forman parte de la cadena de valor de la industria turística. Sería necesario establecer algún sistema de incentivos, que facilitara las posibilidades de compartir datos.

En cuanto a las **barreras a las que se enfrenta el sector**, al igual que sucede en otros sectores, están ligadas a la falta de interoperabilidad y a los silos que aún existen entre los diferentes sectores y entre

<sup>57</sup> <https://www.dataestur.es/general/frontur/>

<sup>58</sup> <https://www.dataestur.es/general/egatur/>

la administración pública y sector privado. La falta de talento digital en la industria turística es también una de las barreras a las que se enfrentan las empresas tecnológicas del sector turístico. Además, podríamos resumir los retos y barreras en las siguientes áreas:

- Mayor colaboración público - privada.
- Implantación de la cultura del dato en empresas y organizaciones (data-driven organizations)
- Falta de digitalización de las empresas turísticas, sobre todo de Pymes y micropymes.
- Apoyo del sector público al tejido empresarial turístico, sobre todo PYMES, en materia de nuevas tecnologías (IA, big data, automatización, IoT...)
- Poca inversión en I+D+i por parte de las empresas turísticas.
- Formación específica en nuevas tecnologías relacionadas con el sector turístico.

#### 7.2.2.7 DSES sector Agua

Según el informe del año 2020 de la UNESCO sobre el Agua y el cambio Climático<sup>59</sup>, el uso global de agua se ha multiplicado por seis en los últimos 100 años y sigue aumentando a un ritmo constante de un 1% cada año debido al crecimiento demográfico, al desarrollo económico y al cambio de los patrones de consumo; por otro lado, el cambio climático genera un entorno en el que la escasez de este recurso puede llegar a ser crónica.

En este marco, las previsiones del Grupo Intergubernamental de Expertos sobre el Cambio Climático exponen una serie de riesgos para los países del sur de Europa relacionados con la ausencia de precipitaciones y los distintos usos del agua, como son la Energía Hidroeléctrica, el Regadío, el agua para consumo humano y los servicios ecosistémicos.

En el caso de España, este contexto marca la definición de una serie de prioridades de investigación para solventar los retos expuestos y que se recogen en la actual “Estrategia Tecnológica del Agua”<sup>60</sup> definida por la Plataforma Tecnológica Española del Agua (PTEA). En ese documento, el sector del agua apuesta por un conjunto de líneas de trabajo como la obtención e sistemas de tratamiento y uso del agua más eficientes; el fomento de la economía circular, incluyendo los diferentes usos del agua en función sus distintas calidades; el conocimiento en mayor profundidad del ciclo del agua en relación con sus usos y la interdependencia entre sus usos (específicamente el nexo agua – energía – alimentación); la gestión integrada del agua; la interconexión entre la explotación sostenible del agua y la biodiversidad y servicios ecosistémicos asociados o la mitigación y remediación del cambio climático, entre otros. En todas ellas, y con carácter transversal, se recogen además las Tecnologías de la Información como una herramienta clave para complementar las nuevas soluciones de tratamiento, depuración y circularidad del agua y la gestión del ciclo del agua en su concepto más amplio.

Estos son los retos a los que se enfrenta el sector del agua en España, que tiene un peso en la economía española del 0,65% de la economía española, genera 38.000 empleos directos, y representa el 0,3% de los insumos que requieren otros sectores productivos de la economía<sup>61</sup>

#### **Principales aplicaciones y beneficios para el sector.**

Los Espacios de Datos se encuadran dentro de las líneas de trabajo “Agua 4.0 Tecnologías Habilitadoras Digitales” recogidas en la Estrategia Tecnológica del Agua de la PTEA; en ella, se definen la

<sup>59</sup> Informe Mundial de las Naciones Unidas sobre el Desarrollo de los Recursos Hídricos 2020: agua y cambio climático. UNESCO. 2020

<sup>60</sup> Estrategia tecnológica del Agua para el período 2020 – 2022.

<sup>61</sup> [www.iagua.com](http://www.iagua.com)



monitorización de procesos, IoT, las redes 5G, big data, tratamiento de datos masivos, block Chain, o supercomputación, como herramientas de un alto valor añadido.

Los Espacios de Datos en el sector del agua, afectan a toda su cadena de valor. Por un lado, a las Autoridades Públicas dependientes de la Administración, como Organismos de Cuenca, Confederaciones o autoridades de planificación hidrológica; pero, por otro, a las entidades basadas en conocimiento, como Universidades, OPIs y Centros Tecnológicos que focalizan sus esfuerzos en el desarrollo de nuevas soluciones que den respuesta en los ámbitos descritos anteriormente; y, por último a las empresas que prestan servicios a las administraciones o que explotan el agua a u nivel municipal o supra municipal.

Contar con Espacios de Datos en el que confluya, por ejemplo, información sobre el clima, las especies animales y vegetales de una determinada zona geográfica, las pautas de consumo, la cantidad de agua regenerada, el agua desalada (cuando se está en una zona costera), las características del terreno, la afluencia turística permitiría beneficios desde varios puntos de vista. Por un lado, para incrementar la sostenibilidad de los recursos, reduciendo su extracción y mejorando la eficiencia en todos los procesos; por otro, para prever la demanda y ajustar los parámetros operacionales y reduciendo su impacto ambiental; por último, y sin ser estos los únicos beneficios, facilitando la toma de decisiones a medio y largo plazo, tanto desde el punto de vista de la explotación del agua, como desde el punto de vista ambiental, social y económico.

La **cadena de valor en el intercambio de datos** involucra a:

- Ministerios de Medioambiente y Sanidad
- Demarcaciones Hidrográficas
- Comunidades Autónomas
- Diputaciones Provinciales
- Ayuntamientos
- Gestores de Agua Públicos, Privados y Mixtos
- Otros agentes: ingenierías, laboratorios, asociaciones y ciudadanos

**Herramientas o plataformas de datos de referencia en el sector:**

- SANePLAN: acceso restringido contempla la importación masiva de datos de infraestructuras de distribución de agua potable y saneamiento y su monitorización. Cuenta con posibilidad de publicación selectiva de información.

**Datos abiertos existentes:**

- MAPAMA: Red de Seguimiento del Estado de Información Hidrológica. Contiene a) Capas generales de cuencas, subcuentas hidrográficas o demarcaciones y b) información específica sobre vertidos, calidad de aguas subterráneas, calidad de agua superficial, evaluación de recursos
- AMET (OPEN DATA) dispone de información meteorológica y predicción por municipio
- SINAC, con información sobre calidad de aguas de abastecimiento
- EIEL, Distintas Diputaciones Provinciales

Las **principales legislaciones o regulaciones específicas** para el sector son:

- **Directiva 2000/60/CE** del Parlamento Europeo y del Consejo, de 23 de octubre de 2000, por la que se establece un marco comunitario de actuación en el ámbito de la política de aguas

- **Directiva 2006/44/CE**, de 6 de septiembre de 2006, relativa a la calidad de las aguas continentales que requieren protección o mejora para ser aptas para la vida de los peces
- **Directiva 2006/7/CE**, del Parlamento Europeo y del Consejo, de 15 de febrero de 2006, relativa a la gestión de la calidad de las aguas de baño
- **Directiva 91/271/CEE** del Consejo, de 21 de mayo de 1991, sobre el tratamiento de las aguas residuales urbanas
- » **Directiva 2006/118**, de 12 de diciembre de 2006, relativa a la protección de las aguas subterráneas contra la contaminación y el deterioro
- **Directiva 2007/60/CE**, de 23 de octubre de 2007, relativa a la evaluación y gestión de los riesgos de inundación

El aspecto legal tiene dos vertientes: una la vinculada al modelo económico relacionado con los Espacios de Datos y los servicios a desarrollar sobre ellos y el otro, que algún sector puede hacer más complejo el reto, si cabe. En el primer caso, entre los factores a tener en cuenta están:

- Origen de los datos (público o privado)
- Financiación de los datos a lo largo del ciclo de vida del Espacio de los Datos: recogida, puesta a disposición y actualización
- Explotación económica, precio y valor del dato

Estos tres ejes marcan una serie de conflictos relacionados con una hipotética explotación económica alrededor de los datos:

- ¿cómo deben valorarse estos servicios? ¿a precio de coste? ¿a precio de venta?
- ¿cómo deben revertirse esos ingresos entre los proveedores?
- ¿qué colisiones existen en función de una financiación previa en la generación de los datos?
- ¿cómo debe ser el modelo de negocio, en relación a los proveedores de los datos, en función de la naturaleza pública o privada de los mismos?

Complementariamente, hay otros aspectos legales, relacionados, por ejemplo, con la vertiente de datos personales o de empresas, de infraestructuras críticas, pero de solución legal quizás más avanzada por las oportunidades que ofrece la propia tecnología.

En el segundo caso, centrandolo el foco en el sector del agua, hay un problema de competencias asociado, pues la gestión del recurso Agua en todo su ciclo de vida afecta a múltiples administraciones: desde la Cuenca Hidrográfica a los Ayuntamientos y a las Autonomías, con intervención de gestores privados o público privados en la explotación del mismo. En este contexto, un Espacio de Datos en el sector de agua, puede necesitar, previamente, una definición legal, marco o acuerdo entre administraciones desde el punto de vista del objetivo del Espacio de Datos y de las competencias de los distintos actores; adicionalmente, se hace también necesario, un marco legal que balacee el potencial de estos Espacios de Datos con los intereses de los distintos actores, especialmente los privados.

Una de las **principales barreras** en el sector gira entorno a lo Competencial debido a los distintos actores que intervienen en el ciclo del agua: públicos y privados. Entre los públicos, además, locales, autonómicos y estatales.

### 7.2.2.8 DSES sector Transporte

La tendencia hacia la digitalización, creciente, del sector transporte, junto con la multiplicidad de operadores y servicios de movilidad (Mobility as a Service, MaaS) abre un amplio Espacio de Datos en el que participan operadores (tanto públicos como privados) y usuarios.

Los operadores de MaaS necesitan datos que reflejen las pautas de movilidad de los usuarios lo más fielmente posible. Sin embargo, hasta ahora los gobiernos, organizaciones y empresas han tenido problemas para decidir qué tipos de datos son apropiados para compartir con el público y qué riesgos, reales y percibidos, podrían estar asociados con la apertura del acceso a los datos. Hay algunos problemas muy reales en juego, relacionados con los costos, la privacidad y la seguridad.

#### **Principales aplicaciones y beneficios para el sector.**

- Integración tecnológica lo suficientemente amplia como para poder ofrecer a los usuarios un servicio completo de movilidad con múltiples opciones en una sola plataforma digital desde la que pueda elegir cómo se quiere mover. Esto, junto con la posibilidad de integrar datos de usuario, etc., nos daría la posibilidad de hacer una gestión de pago y acceso única, sin tener que pagar tramo a tramo, validando identidades en cada operador, sino de una forma continua, sin costuras (*seamless transportation*).
- Neutralidad y requerimientos (condiciones de la neutralidad) para lograr una mayor transparencia en la gestión de las plataformas, especialmente en lo que se refiere a quién comercializa los servicios de movilidad ofrecidos en la plataforma y qué uso se hace de los datos generados.
- Regulación del acceso a los datos públicos y creación de un entorno que incluya a todos los medios de Movilidad Multimodal (promotores de la neutralidad).
- Cuestiones clave que afectan a la operatividad de las plataformas como la “identidad soberana” -que el usuario tenga el control sobre sus propios datos y con quién compartirlos-; el pago único, la integración tarifaria; así como la existencia de intermediarios para la gestión de pagos y distribución de los beneficios (“revenue sharing”).
- Cuestiones clave que afectan a los datos: obstáculos a la colaboración para la creación del usuario único “soberano”; el papel de la administración pública para garantizar la consistencia de la regulación del uso de los datos en las plataformas; y el efecto de los incentivos para que los operadores y los usuarios compartan sus datos.

La neutralidad implica un compromiso por parte de los operadores de plataformas multimodales dirigidas a garantizar la transparencia y sostenibilidad en el servicio, su accesibilidad y seguridad, la colaboración equitativa entre los actores (o no discriminación), el uso compartido de los datos (*data sharing*) de cara a mejorar la eficacia de todo el sistema, todo ello buscando el beneficio primero del usuario (*user centric approach*). Para impulsar plataformas digitales de servicios de movilidad “neutras”, con una gestión transparente de los datos, debemos apoyarnos en una serie de herramientas o propuestas concretas:

#### *Herramientas relacionadas con el “control y uso de datos”:*

- Desarrollar el concepto de la identidad soberana<sup>62</sup>.

<sup>62</sup> La identidad soberana es una forma de identidad digital en la que el usuario tiene pleno control de sus datos. Además de permitirle manejar quienes pueden acceder a ellos y en qué términos. Es uno de los conceptos que más renombre ha comenzado a tomar dentro de la tecnología blockchain (ver también: <https://publications.iadb.org/>)

- Definir la responsabilidad de los operadores en función del tipo de incidencia (brecha de seguridad; avería de vehículos).
- Incorporación de entidades conciliadoras para gestionar el cobro y la distribución de ingresos.
- Pago único e integración tarifaria para consolidar la integración de servicios.
- Creación de modelos de acuerdo que regulen la colaboración entre operadores y de estos con los usuarios.
- Fijación de estándares para la plataforma en cuestiones clave como las tarifas de viaje y el equilibrio entre sostenibilidad/interés de cada modo.

*Herramientas para garantizar la transparencia de la información ofrecida al usuario:*

- Planificación de rutas priorizando las opciones de viajes más sostenibles (menos emisiones), seguras, accesibles y razonables en coste.
- Alternativas que garanticen inclusividad de colectivos específicos.
- Datos de emisiones directas de cada viaje, externalidades e impacto medioambiental.
- Información sobre la accesibilidad de cada modo/viaje.
- Información en tiempo real sobre el estado de los vehículos e incidencias en las rutas de viaje, y que esa información se comparta con el Transporte Público.
- Compartición de los datos de movilidad generados con los operadores públicos y administraciones, para que estos puedan planificar sus matrices de viajes Origen/Destino.

*Herramientas orientadas a superar conflictos de intereses de los distintos operadores/plataformas:*

- Establecimiento de un marco común para integrar los datos de todos los operadores, públicos y privados.
- Creación de un sistema que garantice la coexistencia de varias plataformas de forma que la oferta de cada una de ellas sea completa y equilibrada.

*Los principales actores **en la cadena de valor** del sector son:*

- Operadores
- Planificadores
- Gestores urbanos

**Datos abiertos existentes:**

Los servicios de movilidad generan enormes cantidades de datos útiles tanto para los operadores (perfiles de movilidad, históricos de datos) como para los planificadores y gestores urbanos (incidencias en las rutas de viaje; emisiones; ubicación, accesibilidad de los distintos modos).

Pero los datos de acceso abierto conllevan también grandes retos como la privacidad y la seguridad, ambas cuestiones clave y muy sensibles. A medida que las ciudades, los operadores de movilidad y las administraciones hacen uso de los datos, deben tener cuidado de reconocer sus límites: las empresas privadas sirven a los clientes y las ciudades deben atender al conjunto de la ciudadanía, independientemente de su capacidad para pagar un determinado servicio.

Sin duda, compartir estos datos y hacerlos públicos genera reticencias a pesar de los beneficios potenciales tanto para las ciudades como para las empresas de movilidad. De ahí que los distintos actores demanden unas “reglas del juego” que ayuden a garantizar la neutralidad y transparencia de las plataformas:

- Reglas del juego en torno a la tipología de datos generados:
  - Definir los distintos tipos de datos (del operador/usuario).

- Qué niveles de privacidad se establecen.
- Cómo la plataforma devuelve los datos.
- Acuerdo sobre incentivos y su tipología:
  - Por compartir datos de usuario “completos”.
  - Por ceder sus perfiles de movilidad (sus históricos) con el fin de ajustar las opciones a sus necesidades.
  - Posibilidad de ofrecer otros servicios complementarios de ocio y compras.

Dentro de los **datos abiertos existentes** podemos mencionar:

- EUROPEAN DATA PORTAL<sup>63</sup>., proporciona un punto de acceso único a información del sector transporte de la UE.
- PORTAL DE DATOS ABIERTOS DE ESRI ESPAÑA<sup>64</sup>., pone a disposición datos relacionados con diferentes sectores dentro de los que se encuentra el de transporte. En este portal se puede consultar información relacionada con incidencias de tráfico en España, elementos de la red de autobuses, red de ferrocarriles, entre otras.
- MITMA<sup>65</sup> Open Data Movilidad, ofrece datos de movilidad en España durante el periodo de pandemia por la COVID-19 a nivel nacional.

#### **legislación específica para el sector**

- ✓ Ministerio de Transportes, Movilidad y Agenda Urbana: Estrategia<sup>66</sup> de Movilidad Segura, Sostenible y Conectada 2030.
- ✓ Anteproyecto<sup>67</sup> de Ley de Movilidad Sostenible y Financiación del Transporte (finalizada consulta pública).

**Barreras y retos:** Las plataformas digitales integradas permiten a los usuarios elegir las opciones más sostenibles y convenientes para ellos (más rápida; más económica) y crear viajes predecibles con facilidad. Pero este tipo de integración depende de datos abiertos compartidos con **importantes barreras** que salvar: la neutralidad; la interoperabilidad y las cuestiones relacionadas con la propiedad, uso y privacidad de los datos. Es por ello que requiere altos niveles de coordinación y la voluntad tanto del sector público como de los proveedores de servicios de movilidad privados para que funcione.

<sup>63</sup> <https://www.europeandataportal.eu/en>

<sup>64</sup> <https://opendata.esri.es/>

<sup>65</sup> <https://www.mitma.gob.es/ministerio/covid-19/evolucion-movilidad-big-data/opendata-movilidad>

<sup>66</sup> <https://esmovilidad.mitma.es/>

<sup>67</sup> <https://www.mitma.gob.es/el-ministerio/buscador-participacion-publica/consulta-publica-previa-la-ley-de-movilidad>

### 7.3 Recomendaciones finales

El desarrollo de una Economía de Datos en Europa requiere, entre otras iniciativas, **el despliegue progresivo de data pools o Espacios de Datos sectoriales** como los que se han introducido anteriormente. Las diferentes políticas europeas y españolas, como el Plan España Digital 2025 o la Estrategia Nacional de Inteligencia Artificial, incluyen medidas para la creación de estos espacios que faciliten además la creación de servicios de valor añadido basados en infraestructuras de datos y la construcción de aplicaciones de IA seguras e innovadoras.

Su puesta en marcha representa todavía un enorme desafío por las implicaciones que este tipo de estructuras cooperativas supone en términos de gobernanza, regulatorios, de mercado y de viabilidad técnica. Además, aunque el contexto dentro de la Estrategia europea para Datos es favorable, muchas de las piezas necesarias, sean regulatorias, de mercado o técnicas, están todavía en construcción. Se trata, por tanto, de promover iniciativas que minimicen las incertidumbres, generen confianza y que permitan progresar inteligentemente maximizando los esfuerzos. Para ello, se propone considerar las siguientes recomendaciones:

- **Creación de un comité asesor de apoyo y asistencia a las administraciones** encargadas de las políticas vinculadas al dato que sirva también como interfaz con las políticas europeas en la materia, en representación de los agentes públicos y empresariales. Esta estructura debe contribuir a facilitar la conversión de estas estrategias y políticas en iniciativas y proyectos concretos sectoriales basados en los valores europeos, desde íntegramente públicos hasta íntegramente privados (por ejemplo, en una determinada red de valor empresarial), incluyendo iniciativas público-privadas. Serviría de “espejo” nacional a la iniciativa “Data Innovation Board” propuesta desde Europa.
- **Desarrollo de iniciativas a partir de la experiencia, solvencia y buenas prácticas** existentes en el ecosistema español de Ciencia-Tecnología-Innovación en plataformas y arquitecturas de datos que estén alineadas con estrategias e iniciativas europeas, y con las regulaciones que se vayan implementando al respecto.
- **Incentivación de implementaciones** en aquellos sectores en los que se den condiciones favorables (por interés, referencias, impacto potencial de los datos) y que sirvan de referencia para otros ámbitos posteriormente. Se considera necesaria la incentivación para vencer tanto la incertidumbre inicial como el “efecto plataforma” de este tipo de iniciativas donde el valor es exponencial con el número de nodos / datos en cada nodo.
- **Promoción de iniciativas desde las Administraciones Locales** por tratarse de un contexto especialmente atractivo: fundamentado en los desarrollos sobre ciudades inteligentes de los últimos años, existencia ya de muchos ecosistemas público-privados que han sido promovidos desde las entidades locales, contexto de confianza público-privado, existencia de portales de open data complementarios, etc.
- **Potenciación y promoción de la presencia de pymes** en las iniciativas sectoriales en las condiciones adecuadas para que puedan aprovechar los menores costes de acceso y transacción a los datos para crear nuevos productos, servicios y modelos de negocio.
- **Experimentación en contextos acotados o “sand boxes”** con aquellas aproximaciones que tengan incertidumbres significativas desde el plano legal o regulatorio.
- **Conexión con iniciativas como los eDIHs, los Hubs de IDSA/Gaia-X o los iSpaces de BDVA** para facilitar la experimentación antes de llevar a explotación las aplicaciones, componentes o servicios de manejo de datos.
- **Conexión con iniciativas como los Hubs de IDSA/Gaia-X** para facilitar el desarrollo de cadenas de valor digital y ecosistemas de datos industriales en entornos pre- competitivos.

- **Fortalecimiento del desarrollo de competencias en analítica avanzada de datos e Inteligencia Artificial**, así como en modelos de negocio y monetización de datos, que sirva para maximizar el retorno de las inversiones en el despliegue de los Espacios de Datos.
- **Identificación y validación de componentes (*building blocks*)** que permitan acelerar la construcción de Espacios de Datos y que cubran tanto aspectos generales multi-dominio, como aspectos específicos de un dominio determinado.
- **Potenciar un tejido empresarial tecnológico centrado en la explotación de datos**, que pueda proporcionar servicios sobre los Espacios de Datos construidos y aportar valor a la economía y la sociedad.
- **Realizar campañas de concienciación tanto a empresas como a ciudadanos sobre los datos que disponen**, las oportunidades que supone compartir los datos, los riesgos asociados y las herramientas tecnológicas que permiten minimizar dichos riesgos.

## 8 Referencias

- BDVA. **Towards a European-Governed Data Sharing Space**. Noviembre, 2020. Disponible en: [https://www.bdva.eu/sites/default/files/BDVA%20DataSharingSpaces%20PositionPaper%20V2\\_2020\\_Final.pdf](https://www.bdva.eu/sites/default/files/BDVA%20DataSharingSpaces%20PositionPaper%20V2_2020_Final.pdf)
- Comisión Europea. **A European strategy for data**. Bruselas. Febrero, 2020. Disponible en: <https://eur-lex.europa.eu/legal-content/EN/TXT/?qid=1593073685620&uri=CELEX%3A52020DC0066>
- Comisión Europea. **Digital Europe Programme**. Disponible en: <https://ec.europa.eu/digital-single-market/en/europe-investing-digital-digital-europe-programme>
- Comisión Europea. **Estrategia anual de crecimiento sostenible 2021**. Septiembre, 2020. Disponible en: <https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:52020DC0575&from=EN>
- European Open Science Cloud. **Strategic Research and Innovation Agenda (SRIA) of the European Open Science Cloud (EOSC)**. Noviembre, 2020. Disponible en: <https://www.eosc.eu/sites/default/files/EOSC-SRIA-V09.pdf>
- GAIA-X Project. **GAIA-X: Technical Architecture Release**. June, 2020. Disponible en: [https://www.data-infrastructure.eu/GAIAX/Redaktion/EN/Publications/gaia-x-technical-architecture.pdf?\\_\\_blob=publicationFile&v=5](https://www.data-infrastructure.eu/GAIAX/Redaktion/EN/Publications/gaia-x-technical-architecture.pdf?__blob=publicationFile&v=5)
- Gobierno de España. **España Circular 2030, Estrategia Española De Economía Circular**. Disponible en: [https://www.miteco.gob.es/es/calidad-y-evaluacion-ambiental/temas/economia-circular/espanacircular2030\\_def1\\_tcm30-509532.PDF](https://www.miteco.gob.es/es/calidad-y-evaluacion-ambiental/temas/economia-circular/espanacircular2030_def1_tcm30-509532.PDF)
- Gobierno de España. **Estrategia Nacional de Inteligencia Artificial**. Diciembre, 2020. Disponible en: [https://www.ciencia.gob.es/stfls/MICINN/Ciencia/Ficheros/Estrategia\\_Inteligencia\\_Artificial\\_IDI.pdf](https://www.ciencia.gob.es/stfls/MICINN/Ciencia/Ficheros/Estrategia_Inteligencia_Artificial_IDI.pdf)
- Gobierno de España. **Plan España Digital 2025**. Julio 2020. Disponible en: [https://www.lamoncloa.gob.es/presidente/actividades/Documents/2020/230720-Espa%C3%B1aDigital\\_2025.pdf](https://www.lamoncloa.gob.es/presidente/actividades/Documents/2020/230720-Espa%C3%B1aDigital_2025.pdf)
- Gobierno de España. **Plan España Puede, Plan de Recuperación, Transformación y Resiliencia**. Octubre, 2020. Disponible en: <https://www.lamoncloa.gob.es/presidente/actividades/Paginas/2020/espana-puede.aspx>
- IDS Project. **Reference Architecture Model versión 3**. April, 2019. Disponible en: <https://www.internationaldataspaces.org/wp-content/uploads/2019/03/IDS-Reference-Architecture-Model-3.0.pdf>



## 9 Anexos

### 9.1 Caso GAIA-X

#### ELEMENTOS CENTRALES DE LA ARQUITECTURA.

1. El primero de los elementos y uno de los más importantes de la arquitectura GAIA-X son los Activos. Un **Activo** puede ser cualquiera de los componentes el Ecosistema: un **Nodo**, un **Servicio**, una **Instancia de Servicio** o un **Activo de Datos**.

- **Servicios y Nodos** constituyen el primer elemento de la arquitectura GAIA-X. En este entorno, un **Nodo** es un recurso computacional y de carácter genérico en el que se puede implementar diferentes Servicios. Los Nodos a través de su Auto-descripción exponen atributos funcionales y no funcionales permitiendo a los Consumidores seleccionarlos en base a sus requerimientos. GAIA-X permite las jerarquías de Nodos y por tanto un Nodo puede contener otros Nodos.

Por su parte un **Servicio** GAIA-X es todo aquel servicio que se oferta en la nube, y estos pueden ser individuales o contruidos en conjunto con otros servicios GAIA-X convirtiéndolos en redes de servicios más complejas. Los Servicios son ofrecidos por un Proveedor de servicios y utilizado por un Consumidor de servicios. Cuando nos referimos a una **Instancia de Servicio** GAIA-X, estamos hablando de la ejecución de un Servicio en los Nodos. Cada Servicio puede utilizar un solo Nodo o ser ejecutado en varios. Las instancias de servicio pueden consumir otras instancias de servicio de las que dependen.

- Un **Activo de datos** GAIA-X es un conjunto de datos que se pone a disposición de los Consumidores a través de un Servicio. De esta forma, los datos que proporciona o consume un Servicio GAIA-X se alojan en un GAIA-X Nodo. Consumidores y Proveedores también pueden alojar datos que no estén disponibles dentro de GAIA-X.

2. De otro lado, como segundo elemento del modelo encontramos a los **Participantes**, y agrupa a todas las personas físicas o jurídicas que puede asumir uno o varios de los siguientes roles: Proveedor, Consumidor, Propietario de Datos o Visitante. La combinación de múltiples roles por un participante GAIA-X depende del respectivo Caso de negocio.

- **Usuarios:** Los usuarios son cuentas técnicas derivadas de un participante. Por ejemplo, si una empresa se convierte en Participante, puede haber muchos empleados o usuarios de esa empresa con cuentas individuales.
- **Proveedor:** Todos los Nodos, Servicios e Instancias de servicio tienen un Proveedor asociado.
- **Proveedor de Instancia de Servicio:** Proveen Instancias de Servicios y suelen ser también consumidores de Nodos y Servicios.
- **Propietario de Datos**

3. El tercer elemento lo constituyen las **Autodescripciones** GAIA-X, éstas expresan características y describen propiedades tanto de los **Activos** como de los **Participantes**. Estas Auto-descripciones se vinculan al respectivo Activo o Participante mediante un **Identificador**. En el caso de los Activos, los Proveedores del mismo son los responsables de generar las Autodescripciones.

La **Figura 14** muestra una vista general de las interacciones entre los **Activos** y los **Participantes** del modelo.

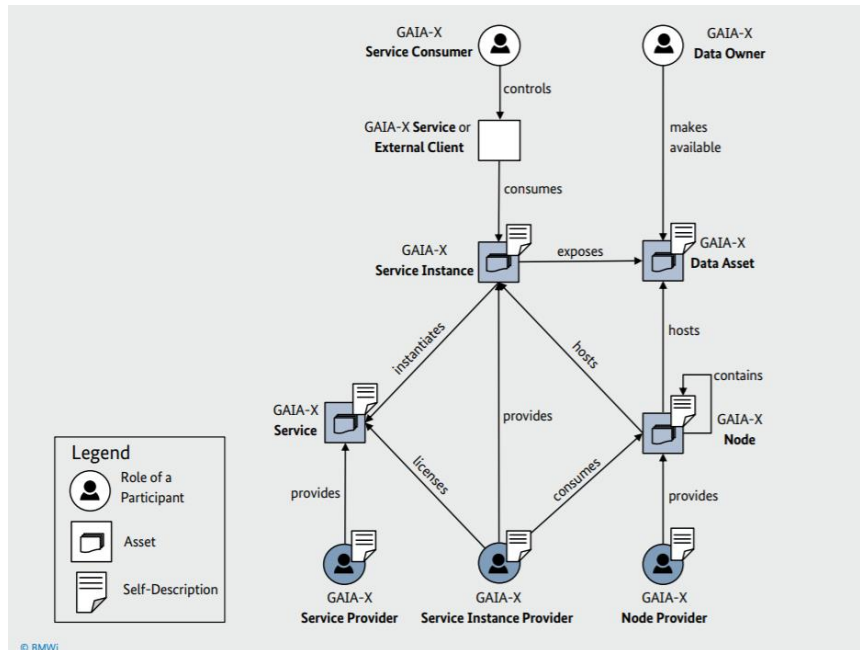


Figura 14. Relaciones entre los Activos y los Participantes de GAIA-X. Fuente: GAIA-X Project

4. Los **Catálogos** son el bloque principal a través del cual se publican y descubren Activos, siendo la base de los **Catálogos Federados** GAIA-X son las Auto-descripciones. Para el proceso de búsqueda de un Activo en el catálogo, el modelo implementa un algoritmo abierto que asegura la transparencia en la oferta de Servicios ya que no adjudica preferencia alguna a cada uno de ellos. De esta forma, si un Usuario desea realizar la búsqueda de un Activo puede realizarlo a través de una interfaz fácil de usar, verificar las Auto-descripciones y tomar autónomamente la decisión, GAIA-X no actúa como intermediario en esta Actividad. Para fines de integración, los Catálogos brindan acceso a través de una interfaz de programación de aplicaciones (API), de esa manera los integradores, los Proveedores o cualquier interesado es libre de integrar las ofertas de GAIA-X en su propia oferta.
5. Un elemento importante en la arquitectura de GAIA-X, son las **políticas** que definen un conjunto de restricciones dentro de la misma. Describe invariantes que deben asegurarse en un entorno de ejecución de software basado en la información de las autodescripciones de activos y participantes.
6. Los datos y la infraestructura deben poder combinarse de formas casi arbitraria para permitir el movimiento de Datos y Servicios en una arquitectura federada. Independientemente de sus ubicaciones virtuales abstractas, los Servicios y los Datos tienen una ubicación física y por tanto GAIA-X aborda aspectos de comunicación por lo que los aspectos de **interconexión y redes** toman gran relevancia en la arquitectura.

Como último elemento, pero no menos importante dentro de este modelo federado encontramos las actividades **monitoreo y medición**. Sin embargo, el acceso para llevar a cabo este seguimiento suele ser un obstáculo técnico para conseguir idealmente un acoplamiento débil de Servicios y

Nodos de diferentes Proveedores. Por esta razón, GAIA-X provee estándares de *mecanismos de monitoreo* que permitan el desarrollo de Ecosistemas de Infraestructura y Datos. De hecho, en GAIA-X las *Auto-descripciones* debe contener la descripción de las capacidades de monitoreo de forma que un Consumidor pueda elegir Servicios y Nodos de acuerdo a sus necesidades en este aspecto. El monitoreo en GAIA-X se aborda desde 3 posibles escenarios: Registro y auditoría, Supervisión y alerta de estado o Medición. Es importante resaltar que, en GAIA-X cualquier tercera parte puede monitorizar la disponibilidad de Servicios en nombre de otro Participante, por ejemplo, supervisar KPIs del nivel de Servicio si se requiere para un acuerdo contractual o certificación.

## INCORPORACIÓN Y CERTIFICACIÓN.

Para poder ofrecer Servicios y Nodos en el entorno de GAIA-X, el Proveedor del mismo debe registrarse previamente. Durante el proceso de ***incorporación de un Proveedor***, deberá aportar la Auto-descripción de los Servicios y/o Nodos que desea incorporar a través del uso de una herramienta disponible para tal fin en el Portal GAIA-X. Teniendo en cuenta que un Proveedor puede ofrecer infinidad de Activos, esta información debe ser registrada como *“datos maestros del proveedor”* para asegurar la consistencia y facilitar el proceso de actualización. El órgano rector de GAIA-X o una institución designada por el mismo deberá realizar una verificación inicial de las Auto-descripciones aportadas, y una vez se ha superado este primer paso siempre que el resultado de la verificación sea positivo, el Proveedor deberá aceptar y firmar los términos y condiciones (T&C) de GAIA-X.

Por su parte, la ***incorporación de un Consumidor*** en GAIA-X es muy simple, limitándose a aceptar los principios y acuerdos de servicio durante el proceso de registro en línea. El Consumidor también deberá aportar una Auto-descripción que será revisada en términos de integridad y honestidad.

Los Proveedores registrados exitosamente podrán ofrecer Servicios o Nodos en GAIA-X y por tanto deberá realizar el proceso de ***incorporación de cada Servicio/Nodo***. Para realizar esta incorporación debe aportar la información organizativa, legal, técnica y la Autodescripción de cada uno de los Activos. Es importante resaltar que, para que un Servicio o Nodo pueda incorporarse en el Ecosistema debe contar una certificación de seguridad que como mínimo asegure un *“nivel básico de garantía de seguridad”*. Un Servicio o Nodo puede obtener la certificación de seguridad básico o alto.

- Para solicitar el ***Nivel de seguridad básico***, el Proveedor debe presentar la solicitud de aplicación al órgano rector de GAIA-X (o en su defecto al órgano de seguimiento designado) aportando la documentación requerida y firmar un contrato que especifique sus obligaciones (honorarios, notificación de cambios en el Nodo / Servicio, etc.). La solicitud será examinada por un revisor cualificado e independiente con base en el manual de directrices que describe el proceso de evaluación. Este revisor envía un informe que deberá ser aprobado por el órgano rector de GAIA-X para emitir la *declaración básica de seguridad y por tanto incluir el Nodo/Servicio en el Catálogo GAIA-X*.
- Para solicitar el ***Nivel de seguridad considerable o Alto*** se requiere una certificación basada en terceros. Por tanto, la auditoría debe ser realizada por un organismo de evaluación de la conformidad (CAB) de acuerdo con el Programa de evaluación de la conformidad GAIA-X. La organización solicitante presentará el certificado obtenido al órgano de gobierno de GAIA-X o un designado organismo de seguimiento. Después de superar una prueba extendida de seguridad y vulnerabilidad el *Nodo/Servicio será incluido en el Catálogo GAIA-X*.

## SEGURIDAD DE LA INFORMACIÓN Y PROTECCIÓN DE DATOS

La **seguridad de la información** es uno de los principios básicos de GAIA-X. Para garantizar el más alto nivel de protección de datos, seguridad, transparencia y portabilidad en todos los Servicios, GAIA-X define pautas, políticas, una arquitectura de referencia, y determina los Servicios Federados de forma que se implementen de la misma forma por parte de todos los Participantes. Con el propósito de crear, mantener y fortalecer la confianza entre los Participantes, GAIA-X brindará total transparencia en cuanto a la implementación técnica y el nivel de seguridad de los Servicios Federaos de GAIA-X. Además, para hacer parte del Ecosistema cada Activo debe estar registrado en el *Catálogo Federado de GAIA-X*, aportando su **Auto-descripción la cual deberá firmar criptográficamente durante el proceso de incorporación del Activo**.

En cuanto a la responsabilidad y **protección de datos**, teniendo en cuenta que GAIA-X es un sistema federado de proveedores autónomos, basa su modelo a partir de los principios de un modelo de responsabilidad compartida. Por esto cada Participante GAIA-X es responsable por los servicios y datos que están bajo su control, cada Proveedor de Servicios/Nodos es responsable de su seguridad, así como cada Proveedor de datos será responsable de la correspondiente protección de datos. Sin embargo, GAIA-X proporciona Servicios técnicos de la Federación para que los Participantes pueden realizar esta tarea de una forma más automatizada.

La base de un Ecosistema basado en datos son precisamente los datos, en ese sentido, un aspecto que cobra importancia es el acceso a los mismos. Al respecto, los **mecanismos de control de acceso** son obligatorios y deben ser parte de las negociaciones entre el proveedor de datos y el Consumidor de datos. El Proveedor (o propietario) de los datos proporcionan acceso a los datos, pero debe seguir teniendo control sobre su uso para poder cumplir con sus responsabilidades en cuanto a protección y seguridad de los mismos.

GAIA-X aplica **políticas de uso** que permiten la evaluación de atributos, Auto-descripciones e incluyen atributos como la identidad del usuario. Así, un Visitante podrá navegar por todo el catálogo GAIA-X pero podría no estar habilitado para ver todos los atributos de la Auto-descripción de un Activo específico. GAIA-X no implementará mecanismos de acceso central para controlar el acceso de cualquier Consumidor a un determinado Activo, la responsabilidad recae en el Proveedor de este activo. Sin embargo, GAIA-X proporcionará una API que permite al Proveedor y al Consumidor consultar y verificar la identidad y la Autodescripción de la otra parte basándose en firmas criptográficas

Una vez el acceso a los datos ha sido otorgado, los datos pueden ser alterados, copiados o difundidos arbitrariamente por el destinatario, por esta razón el **control de uso de datos** es necesario para controlar futuros usos de datos (también conocidas como obligaciones del Consumidor). Este control aborda la especificación y aplicación de restricciones que regulan lo que no debe suceder con los datos. El control de uso es relevante en el contexto de la protección de la propiedad, cumplimiento de la normativa y gestión de derechos digitales.

Respecto del **cumplimiento normativo**, GAIA-X plantea *Servicios de Cumplimiento* para asegurar que todos los Participantes, Servicios o Nodos cumplen con los requerimientos internos, así como con las regulaciones externas y políticas (como por ejemplo los requisitos de seguridad de la información y protección de datos).

## 9.2 Caso International Data Spaces (IDS)

### CAPA DE PROCESO

Esta capa especifica las interacciones que tienen lugar entre los diferentes componentes del IDS y por tanto, proporciona una vista dinámica del modelo de referencia. Dentro del modelo existen 3 procesos fundamentales relacionados estrechamente con su propuesta de valor e involucran la mayoría de los roles descritos en la capa de negocios (capa 1).

- Proceso de Incorporación:** Este proceso involucra diferentes subprocesos. El primer paso para que un Usuario se una al IDS como Proveedor de datos o Usuario de datos es **adquirir una identidad** proporcionada por un Organismo de certificación. En segundo lugar, deberá **solicitar un Conector** (componente técnico a instalar) a un Proveedor de Software. Una vez cuenta con estos dos requisitos recibirá un certificado con el que se asegura el cumplimiento de las especificaciones necesarias. El proceso general de "incorporación" se ilustra en la *Figura 15*.

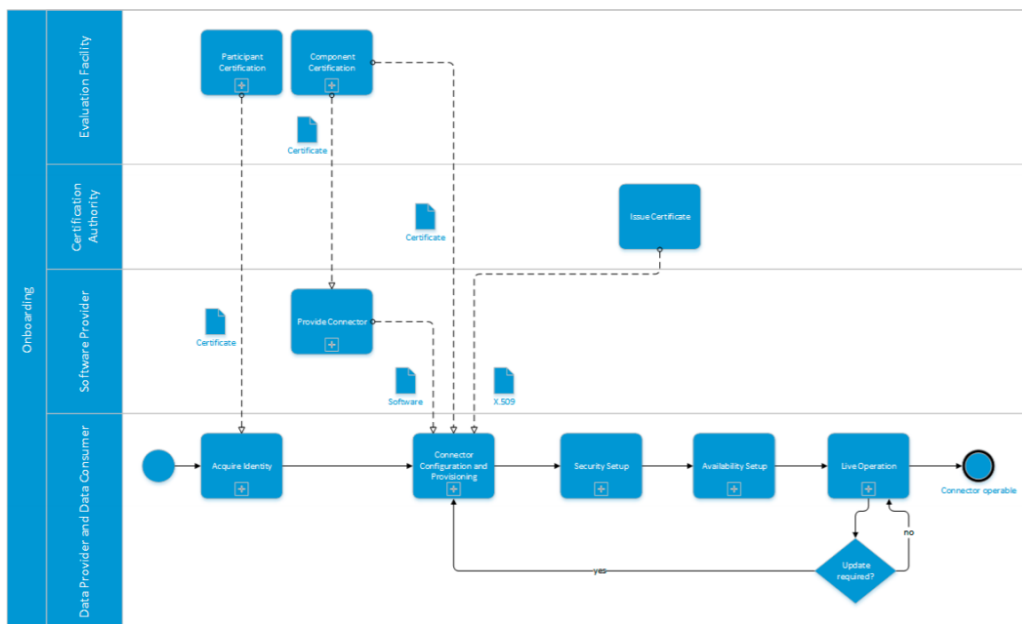


Figura 15. Proceso de incorporación

- Intercambio de datos:** El proceso general consta de dos sub procesos. El primero aborda al Consumidor de datos que busca un Proveedor de datos, en este caso el Consumidor de datos puede contactar directamente al Proveedor de datos si le conoce o utilizar los servicios de un Proveedor de brokering para encontrarlo. El segundo subproceso es la invocación de una operación de datos como, por ejemplo, la carga o descarga, transformación o consulta de datos.
- Publicación y uso de Apps de datos:** Los conectores pueden utilizar **Aplicaciones de datos** para tareas específicas de procesamiento o transformación de datos. A nivel conceptual, estas Aplicaciones se pueden tratar de la misma manera que la oferta de datos en los IDS. Por tanto, de la misma forma en que los *Datos* son proporcionados por un *Proveedor de Datos* mediante un *Conector* asociado a un *Broker*, en este caso las *Aplicaciones* son creadas por un *Proveedor de Aplicaciones* y registradas en una *tienda de Aplicaciones* (usando el Conector del Proveedor

como medio para usar la App store). En ese sentido un Proveedor de aplicaciones también debe registrarse, sin embargo, en vez de asociar su conectar a un bróker deberá registrar sus aplicaciones en una tienda de aplicaciones. Resaltando que algunas aplicaciones pueden necesitar una certificación especial antes de ser registradas.

### CAPA DE INFORMACIÓN

Esta capa especifica el **Modelo de Información y el lenguaje común** a utilizar en los IDS. El Modelo de Información facilita la compatibilidad y la interoperabilidad entre los participantes y los componentes del IDS. El propósito principal de este modelo es permitir (semi) automatizar el intercambio de recursos digitales dentro del ecosistema, sin embargo, es un modelo genérico no especializado en un dominio particular. El Modelo de Información se ha especificado en tres niveles de formalización, en el que cada nivel corresponde a una representación digital (conceptual, declarativa o programática) como se muestra en la *Figura 16*.

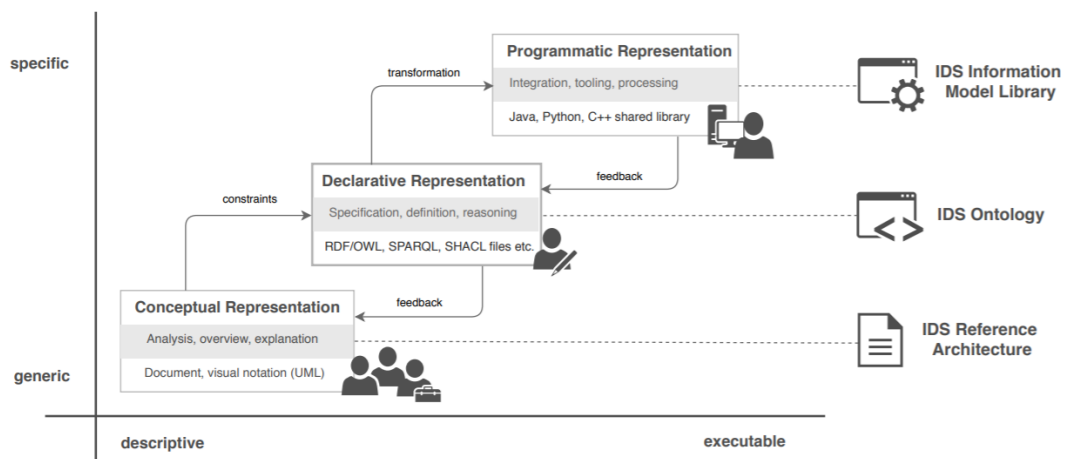


Figura 16. Modelo de Información

El IDS expresa su modelo de información como una ontología basada en un esquema RDF con el fin de proporcionar identificadores inequívocos y definiciones formales de sus conceptos y relaciones. Teniendo en cuenta que el intercambio de datos es el núcleo del IDS, **el vocabulario** requerido para las descripciones de los datos y las invocaciones es muy básico.

En relación a las Aplicaciones y de forma similar en como un Conector proporciona información sobre la identidad, una Aplicación proporciona información sobre sí misma mediante un archivo de descripción que contiene detalles sobre su uso previsto, el propósito, nivel de seguridad y modelo de licencia que aplica. Sin embargo, un Proveedor de datos puede describir una aplicación de datos con vocabularios fuera de la Ontología principal del IDS.

### 3. CAPA DEL SISTEMA

En esta capa los roles especificados en la capa de negocios (capa 1) se asignan a una arquitectura concreta de datos y servicios para cumplir con los requisitos especificados en la capa funcional (capa 2) dando lugar así al núcleo técnico de los IDS. A partir de los requisitos identificados en la capa funcional, se establecen **tres componentes técnicos** principales: **el Conector**, **el Corredor (bróker)** y la **App Store**. La forma en cómo interactúan entre sí se muestra en la *Figura 17*.

Una red distribuida como el IDS se basa en la conexión de diferentes nodos miembros donde se alojan Conectores u otros componentes centrales. El Conector es entonces el responsable del intercambio de datos ya que ejecuta el proceso completo.

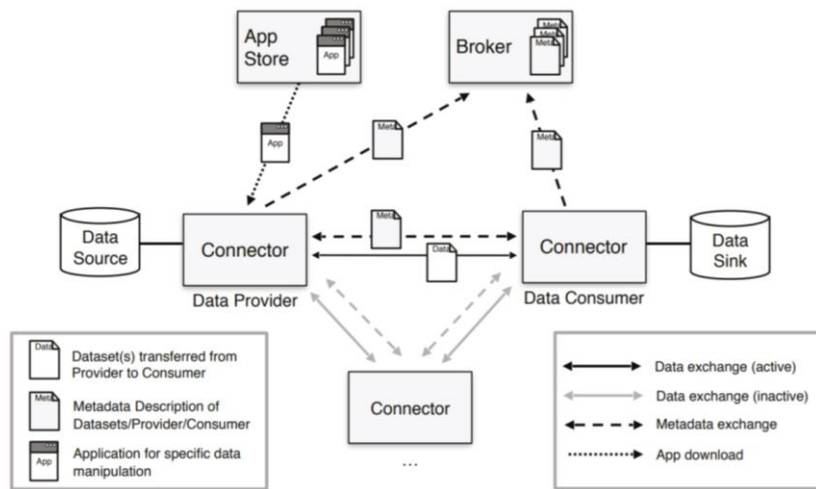


Figura 17. Interacción de componentes técnicos

### PERSPECTIVA DE SEGURIDAD.

La Arquitectura de Seguridad de IDS proporciona medios para identificar a los Participantes, proteger las comunicaciones y transacciones de intercambio de datos y controlar el uso de datos después ha sido intercambiado. **El desarrollo de la Arquitectura de seguridad** sigue dos principios generales: por un lado, **el uso de las normas y mejores prácticas** ya existentes y por otro lado la **escalabilidad de los niveles de seguridad**. La *Arquitectura de Seguridad* aborda siete conceptos clave de seguridad:

- 1) comunicación segura,
- 2) gestión de identidad,
- 3) gestión de confianza,
- 4) plataforma de confianza,
- 5) control de acceso a datos,
- 6) control de uso de datos y
- 7) seguimiento de procedencia de datos.

### PERSPECTIVA DE CERTIFICACIÓN

La soberanía de datos es uno de los valores fundamentales del IDS, entendida como la capacidad e una persona natural o una entidad jurídica de poseer el control total de sus datos. Por esta razón, cualquier organización o individuo que solicite acceso al IDS debe estar certificada, al igual que los componentes de software (Conector).

Si bien, la certificación de organizaciones e individuos se centra en la seguridad y confianza, la certificación de componentes se refiere al cumplimiento de los requisitos técnicos que garantizan la interoperabilidad.

Para asegurar un proceso consistente en la certificación de participantes y componentes, el IDS utiliza un *Esquema de Certificación* que comprende todos los procesos, reglas y estándares que rigen el proceso de certificación. La *Figura 18*. describe este esquema junto con los roles involucrados en este proceso.

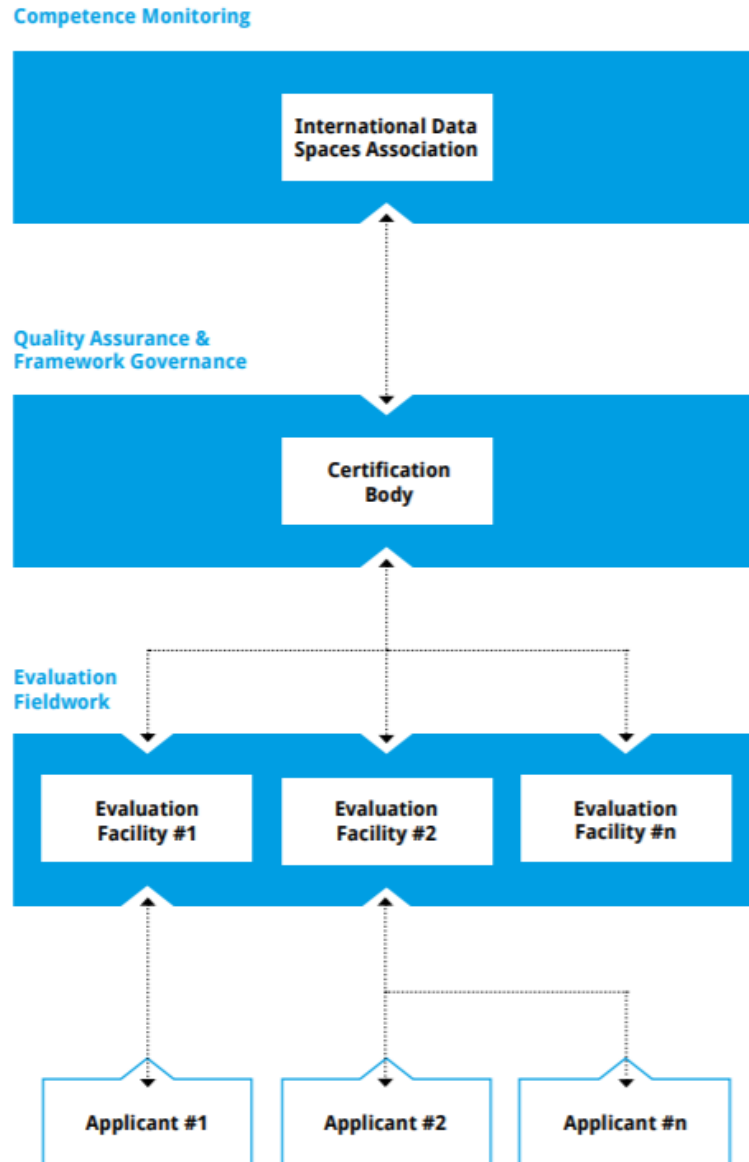


Figura 18. Esquema de certificación



