gaia-x Hub España

Data Spaces Guideline

Impulsando la economía del dato española mediante la compartición y explotación de datos de forma confiable, soberana y segura

Fecha: 07/07/2025

Versión 01

Ayuda 27.08.4671.488 financiada por:

GOBIERNO DE ESPAÑA
MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL
SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

ÍNDICE

Contenido

ÍNDICE 1

1. Introducción: Qué es un espacio de datos 8

2. Roles y actores 9

3. Gobernanza de un espacio de datos 10

3.1. Gobernanza de datos 11

3.1.1 Qué es la gobernanza de datos 11

3.1.2 Características principales 12

3.1.3 Principios del gobierno del dato 13

3.1.4 Beneficios 15

3.1.5 Pasos a dar para la implementación de la gobernanza 17

3.2. Configuración de un espacio de datos: Rol de la Autoridad de Gobierno 21

4. Gestión de participantes, identidad y confianza Roles y actores 22

4.1 Gestión de identidad / Credenciales 22

4.1.1 Credenciales Verificables 23

4.1.2 Protocolos para la emisión y verificación (autenticación) de credenciales 24

4.2 Certificados / Clearing Houses 26

4.3 Conformidad con Gaia-X 27

4.4 Onboarding de participantes en un espacio de datos 29

5. Gestión de datos y servicios 29

5.1 Productos de datos 29

5.2 Catálogos 31

5.3 Políticas de acceso y contratos 32

5.3.1 Políticas de uso y acceso (ODRL) 32

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA · MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL · SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

5.3.2 Integración de políticas de acceso y uso en arquitecturas de referencia de espacios de datos....34
5.3.3 Establecimiento de contratos entre proveedor y consumidor....37
5.4 Marketplaces....39
6. Intercambio de datos....40
6.1 Conector EDC....42
6.1.1 Plano de Control....43
6.1.2 Plano de Datos....46
6.1.3 Catálogo Federado....46
6.1.4 Centro de Identidad....47
6.1.4.1. API DCP....47
6.1.4.2. API de Identidad....48
6.1.5 Panel de Datos (Data Dashboard)....48
6.2 Conector FIWARE....48
6.2.1 Visión general....48
6.2.2 Módulo de Autenticación basado en Gestión de Identidad Descentralizada....50
6.2.3 Módulo de Autorización....54
6.2.4 Módulo de Gestión de Productos y Negociación de Contratos (incluyendo soporte a monetización)....55
6.2.5 Soporte a familia de protocolos IDSA DSP....56
6.2.6 Plano de Datos (intercambio)....57
6.2.7 Integración con DOME....58
6.3 Tekniker Dataspace Connector....58
6.3.1 Metadata Manager....60
6.3.2 Base de datos PostgreSQL....63
6.3.3 Dataspace Protocol....63
6.3.4 Data Plane....64
6.3.5 Policy Engine....65
7. Recursos adicionales....65
7.1. Data Spaces Blueprint (DSSC)....65
7.2. Recursos de Gaia-X España....66

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

7.3. Bibliotecas de componentes y conectores..... 66

Créditos y agradecimientos 68

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA
MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL
SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

Lista de figuras

Figura 1. Principios del gobierno del dato. 14

Figura 2. Árbol de roles en la gobernanza del dato. 17

Figura 3. Ejemplo de dataset ofrecido a través de un catálogo de datos (fuente: datos.gob) 31

Figura 4. Plano de Control y Plano de Datos. 42

Figura 5. Arquitectura de alto nivel del conector EDC. 43

Figura 6. Arquitectura de alto nivel del FIWARE Data Space Connector (DSC) 49

Figura 7 - Invocación de servicios mediados via FIWARE DSC por parte de un usuario final 51

Figura 8 - Invocación de servicios mediados via FIWARE DSC por parte de un sistema o dispositivo 53

Figura 9. TDC. Arquitectura General. 60

Figura 10. TDC. Modelo para la descripción del catálogo basado en DCAT. 60

Figura 11. TDC Swagger UI. Creación de un Dataset 61

Figura 12. TDC Swagger UI. Creación de un DataAddress 61

Figura 13. TDC Swagger UI. Enlace de un Dataset a un DataAddress 62

Figura 14. TDC. Extensión de DCAT para la descripción de ContractNegotiation, Agreement y TransferProcess 62

Figura 15. TDC UI. Creación, lectura, actualización y visualización de un Catálogo 63

Figura 16. TDC. Data Plane Pull Aproach 64

Figura 17. Grupos e instancias de conectores (fuente: IDSA) 67

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

Lista de acrónimos

ABAC	Attribute-Based Access Control
API	Application Programming Interface
CAB	Conformity Assessment Body
CDO	Chief Data Officer
CES	Credential Event Service
CKAN	Comprehensive Knowledge Archive Network
CRUD	Create, Read, Update, Delete
DAMA	Data Management Association
DAPS	Dynamic Attribute Provisioning Service
DCAT	Data Catalog vocabulary
DCP	Decentralized Claims Protocol
DID	Decentralized Identifier
DOME	Distributed Open Marketplace for Europe
DPF	Data Plane Framework
DSBA	Data Spaces Business Alliance
DSP	Dataspace Protocol
DSSC	Data Spaces Support Centre
EBSI	European Blockchain Services Infrastructure
EDC	Eclipse Dataspace Components
ETSI	European Telecommunications Standards Institute
EUDI	European Digital Identity Regulation
eIDAS2	electronic IDentification, Authentication and trust Services 2
FC	Federated Catalogue

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

FCC	Federated Catalog Crawler
GXDCH	Gaia-X Digital Clearing House
H2M	Human-to-Machine
IDS-RAM	International Data Space Reference Architecture Model
IDSA	International Data Space Association
ISO	International Standards Organisation
JAAS	Java Authentication and Authorization Service
JWT	JSON Web Token
M2M	Machine-to-Machine
NGSI	Next Generation Service Interfaces
ODRL	Open Digital Rights Language
OID4VC	OpenID for Verifiable Credentials
OID4VCI	OpenID for Verifiable Credential Issuance
OID4VP	OpenID for Verifiable Presentations
OIDC	OpenID Connect
PAP/PRP	Policy Administration/Registration Point
PDP	Policy Decision Point
PEP	Policy Enforcement Point
PIP	Policy Information Point
PXP	Policy eXecution Point
REST	Representational State Transfer
SIOP	Self-Issued OpenID Provider
TCN	Target Catalogue Node
TDC	Tekniker Data Connector
UE	Unión Europea

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

VC	Verifiable Credential
W3C	World Wide Web Consortium
XACML	eXtensible Access Control Markup Language

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

1. Introducción: Qué es un espacio de datos

De acuerdo al glosario del Data Spaces Support Centre[1] (proyecto creado por la Comisión Europea para proporcionar orientación, herramientas y recursos a todas aquellas organizaciones interesadas en crear nuevos espacios de datos o participar en ellos), un **espacio de datos** es un *marco interoperable, basado en principios de gobernanza, **estándares**, prácticas y servicios habilitantes comunes, que permite transacciones de datos **seguras** y **confiables** entre los participantes.*

Técnicamente, un espacio de datos es un sistema digital **distribuido** y **descentralizado** que permite a los participantes transferir datos e invocar servicios de datos de manera confiable y segura. Los espacios de datos descentralizados no deben confundirse con las plataformas de datos centralizadas administradas por una única organización, que también pueden ofrecer marcos para compartir datos.

Una de las características de los espacios de datos es la **soberanía de datos**, entendida como la capacidad de los participantes de poder ejercer el control respecto al intercambio y la compartición de datos a través de servicios de datos, de forma que puedan tomar decisiones informadas sobre con quién y de qué forma se comparten, así como cuáles son las características técnicas específicas y las reglamentaciones europeas o nacionales elegidas a las que deben ajustarse los servicios ofrecidos.

Un espacio de datos es una solución **escalable** para compartir datos y servicios de datos en la que es contractual y técnicamente fácil integrar nuevos participantes y casos de uso. Un **caso de uso** en un espacio de datos se refiere a una situación en la que los participantes crean valor empresarial o social utilizando los mecanismos de intercambio que ofrece el espacio de datos. Un espacio de datos suele admitir múltiples casos de uso, logrando que su implementación sea

[1] https://dssc.eu/space/bv15e/766061530/Introduction+-+DSSC+Glossary

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

rentable y creando efectos de red, ya que los casos de uso a menudo involucran parcialmente a los mismos actores, servicios y/o fuentes de datos.

Las transacciones dentro de un espacio de datos pueden incluir mecanismos que permitan la utilización de los datos sin que estos necesiten ser transferidos, y en cualquier caso todas las transacciones deben de poder ser auditables.

2. Roles y actores

En un espacio de datos existen diversos **actores** que desempeñan **roles específicos** que son cruciales para garantizar el funcionamiento eficiente del mismo. Cada actor tiene responsabilidades específicas e interactúa con el sistema de una manera determinada. A continuación, se detallan los principales roles que los actores que participan en los procesos de negocio de un espacio de datos pueden asumir:

- **Autoridad de Gobernanza**: Es el participante del espacio de datos responsable de crear, desarrollar, operar, mantener y aplicar el marco de gobernanza para el espacio de datos, sin sustituir la función de las autoridades públicas competentes.

- **Solicitante**: Un "solicitante" se refiere a una entidad que presenta una solicitud o petición para participar en el espacio de datos. Este término se utiliza generalmente en contextos donde existe un proceso de selección o aprobación. Los solicitantes se encuentran en la etapa inicial de interacción con el espacio de datos, buscando la aprobación o aceptación, la cual se basa en los criterios y procedimientos establecidos por la Autoridad de Gobernanza del Espacio de Datos.

- **Participante**: Una vez completado el proceso de solicitud, entidad comprometida con el marco de gobernanza de un espacio de datos específico y que puede desempeñar uno o más roles en él.

- **Proveedor de aplicaciones**: Son participantes en los espacios de datos que ofrecen aplicaciones a los consumidores. Típicamente las aplicaciones son consumidas a través de servicios visibles y consumibles en el espacio de datos. Las características de estas aplicaciones pueden estar accesibles a través del catálogo de aplicaciones. El término "aplicación" se utiliza en un sentido amplio en este documento y abarca cualquier tipo de ejecutable, incluyendo aplicaciones, así como algoritmos, como un modelo de IA entrenado que los usuarios pueden utilizar para analizar sus datos. Los proveedores de aplicaciones también pueden definir políticas de control de acceso a sus servicios y facturar a los usuarios por su uso.

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

- **Proveedor de datos**: Esta categoría abarca a todos los actores del espacio de datos que ofrecen datos a los consumidores. Pueden compartir uno o más conjuntos de datos y regular el acceso y el uso de los mismos mediante políticas de control de acceso y acuerdos de uso. Para compensar el uso de los datos, los proveedores de datos también pueden facturar a los consumidores del espacio de datos. Un ejemplo de proveedor de datos puede ser un operador de red energética que comparte datos sobre la carga de la red con las instalaciones de producción de energía (que actúan como consumidores) para aplicaciones de optimización de la producción.

- **Proveedor de infraestructura**: Los proveedores de infraestructura ofrecen recursos y servicios de infraestructura a los consumidores (o posiblemente a cualquier otro tipo de participante) para que puedan procesar los datos proporcionados por los proveedores de datos. Pueden, por ejemplo, iniciar máquinas virtuales o contenedores y ejecutar aplicaciones, algoritmos u otros ejecutables sobre la infraestructura subyacente. De forma similar a los proveedores de datos, los proveedores de infraestructura pueden definir políticas de control de acceso para los recursos de infraestructura y facturar a los usuarios del middleware por su uso.

- **Consumidor**: Este grupo incluye a los participantes que pueden acceder a los datos, las aplicaciones y la infraestructura ofrecidos por proveedores dentro del espacio de datos. Pueden interactuar con el espacio de datos de diversas maneras: pueden buscar datos u otros recursos, como algoritmos y modelos de IA en los espacios de datos; usarlos en línea mediante la infraestructura y los servicios que los proveedores proporcionan en el espacio de datos o, si la política lo permite, descargarlos para uso local. También pueden usar servicios de acceso a datos o de procesamiento de datos, típicamente utilizando algún API REST. Un consumidor está vinculado a una organización participante.

- **Representante**: Se refiere al "usuario final" humano que realiza acciones en nombre de una entidad, en su rol como consumidor, proveedor o autoridad de gobernanza.

3. Gobernanza de un espacio de datos

En los últimos años, debido al creciente reconocimiento de la importancia estratégica de los datos y a los avances tecnológicos en el campo de la gestión y procesamiento de datos, se ha visto la necesidad de tratar a los datos como un activo fundamental en cualquier organización. La gobernanza del dato ha evolucionado considerablemente a medida que las necesidades y las herramientas han permitido aumentar el control sobre los datos.

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

Existen varias organizaciones internacionales que han desarrollado estándares y marcos de trabajo relacionados con la gobernanza del dato. Algunas de las más destacadas son las siguientes:

- DAMA International (*Data Management Association International*): DAMA International es una organización sin ánimo de lucro dedicada a la gestión de datos. Han desarrollado el *Data Management Body of Knowledge* (DMBOK), que es un marco de trabajo ampliamente reconocido para la gestión de datos, que incluye la gobernanza del dato como uno de sus componentes principales.

- ISO (Organización Internacional de Normalización): La ISO ha publicado estándares relacionados con la gestión de datos y la gobernanza del dato. Por ejemplo, ISO 8000 establece los requisitos para la calidad de los datos, mientras que ISO 27001 se enfoca en la seguridad de la información y puede ser aplicado a la gestión de datos.

- CDO (Chief Data Officer) Council: El CDO Council es una organización global que reúne a ejecutivos de datos y líderes en gobernanza del dato. Aunque no emite estándares específicos, proporciona orientación y recursos valiosos para la gobernanza del dato y el papel del CDO.

Estas organizaciones y estándares proporcionan orientación y mejores prácticas para la gobernanza del dato a nivel internacional. Sin embargo, es importante destacar que los estándares y enfoques pueden variar dependiendo de la industria, la región y las necesidades específicas de cada organización.

Por otro lado, la Unión Europea (UE), a través del Data Governance Act ha establecido el Data Governance Act, un marco normativo que busca promover la confianza y la interoperabilidad en el uso de datos en la Unión Europea. El acto aborda temas como el acceso a los datos, la portabilidad y la gobernanza de los datos.

Es recomendable que las organizaciones consideren los estándares como referencias y los adapten en su enfoque de gobernanza del dato según su contexto y requisitos individuales, siguiendo el marco regulatorio de la UE.

3.1. Gobernanza de datos

3.1.1 Qué es la gobernanza de datos

La gobernanza de datos (data governance) supone la definición de normas y el control sobre la gestión de datos en términos de planificación, ejecución y seguimiento. La gestión de datos

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

implica, por tanto, el desarrollo e implementación de arquitecturas, políticas, procesos y procedimientos que gestionan de manera adecuada todas las necesidades del ciclo de vida de los datos, en el contexto de una organización. Es, por tanto, la capacidad de gestionar los datos como un verdadero activo empresarial. En otras palabras, al igual que la gestión de cualquier otro activo, por ejemplo, un edificio, una maquinaria, un servicio innovador o, en general, cualquier otro elemento que tenga valor para la empresa, la gestión de los datos también necesita unas reglas básicas para que los datos puedan producir y traducirse en un valor económico.

3.1.2 Características principales

Todo buen gobierno del dato debe cumplir una serie de características básicas. A continuación, se detallan:

1. **El negocio patrocina el proceso del gobierno**. El área o áreas de negocio de una empresa son lo que alimentan y al mismo tiempo consumen los datos, y es por ello que el negocio de la empresa debe ser quien fomente su desarrollo.

2. **El gobierno del dato es un proceso continuo**. Es importante tener en cuenta que el gobierno del dato es una tarea que debe desarrollarse durante toda la vida del sistema. No puede tener una fecha final ya que a medida que se avanza en él, se transforma en una actividad diaria y de continua revisión y mejora.

3. **Mejora continua de los datos, procesos y herramientas**. En la evolución y con el paso del tiempo debe tenerse en cuenta la mejora de la calidad del dato para que aporte valor al negocio de la empresa. Esto incluye la mejora de los procesos y las actualizaciones de las herramientas o, incluso, la sustitución de las mismas por otras más potentes o modernas.

4. **Los beneficios del gobierno del dato deben ser reales y cuantificables**. El gobierno del dato necesita contar con medidas o métricas claras que permitan medir el beneficio de su implementación. Estas medidas son amplias y se deberán elegir aquellas que demuestren fehacientemente la mejora en la administración y la gestión de los datos en una empresa. Ejemplos de estas medidas pueden ser los siguientes:

    - Reducción de *"time to market"*, es decir, del tiempo necesario para que un producto llegue al mercado.
    - Incremento en la disponibilidad de datos a ser explotados para fines comerciales.
    - Eficiencias de recursos debido a la simplificación de procesos.

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

- Reducción de tiempo dedicado a la limpieza de datos o conciliación de la información.
- Reducción de reclamaciones de clientes por error de la información o uso indebido de sus datos.
- Reducción de multas del regulador por incumplimiento de normas de gestión de datos.

5. **Alineación de toda la organización con la estrategia de gobierno.** Todas las áreas, departamentos, personas, etc. de una empresa u organización deben estar implicadas en la estrategia del gobierno del dato. Aunque como se explica anteriormente, debe ser negocio quien lidere todo el proceso, la responsabilidad es compartida entre todos los actores implicados en el proceso. Es muy importante por ello definir claramente las responsabilidades de los diferentes actores de una organización para una gobernanza efectiva.

6. **Reglas y lenguaje común.** Deben establecerse reglas comunes para toda la organización para el uso y tratamiento de los datos. Además, se debe establecer un lenguaje común para facilitar el entendimiento, por ejemplo, un diccionario de datos.

7. **Cumplir los requerimientos regulatorios.** El gobierno del dato debe cumplir con la regulación vigente ya sea en la protección de datos, ética, transparencia, etc. La mejora continua, explicada anteriormente, debe tener en cuenta cada cambio en la legislación para que sea aplicada al gobierno del dato de la organización. Dependiendo de los datos que se vayan a almacenar la legislación puede ser más estricta, ya que no es lo mismo un conjunto de datos de producción industrial, que los datos económicos o de salud de un grupo de personas.

3.1.3 Principios del gobierno del dato

Todo ámbito del gobierno del dato debe regirse por unos principios que regulen el tratamiento de la información almacenada por la organización. Cumplir estos principios es importante para alcanzar los objetivos que una organización desea cuando quiere implementar la estrategia del gobierno de datos en, por ejemplo, un espacio de datos. Además, son fundamentales para garantizar una gestión efectiva y responsable de los datos en una organización. A continuación, se detallan estos principios:

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

Figura 1. Principios del gobierno del dato.

1. **Responsabilidad**: Este principio establece que debe haber una clara asignación de roles y responsabilidades en la gestión de los datos. Se deben designar responsables de la calidad de los datos, la seguridad, la privacidad y el cumplimiento normativo. Es fundamental establecer una cultura de responsabilidad y rendición de cuentas en relación con los datos.

2. **Accesibilidad**: La accesibilidad se refiere a la capacidad de los usuarios autorizados para acceder y utilizar los datos de manera oportuna y eficiente. Es decir, que los datos deben estar accesibles, reutilizables y deben compartirse de forma transparente.

3. **Disponibilidad**: Este principio se refiere a la capacidad de acceder a los datos cuando se necesiten. Esto implica asegurar que los datos estén disponibles y accesibles en el momento adecuado y en el formato correcto. Para ello deben tenerse en cuenta políticas de entre otras, de respaldo, escalabilidad, monitoreo, recuperación, etc.

4. **Calidad de los datos**: La calidad de los datos es esencial para tomar decisiones informadas y confiables. Este principio se refiere a la necesidad de establecer estándares, procesos y controles para garantizar la exactitud, integridad, coherencia y actualidad de los datos. Esto implica la implementación de políticas de control de calidad, validación, corrección de

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

errores y monitoreo continuo de los datos. Estas políticas deben tener en cuenta la función o el propósito de su uso, ya que en dos procesos diferentes los datos almacenados pueden tener funciones diferentes.

Por ejemplo, un proceso puede tener una tolerancia a datos erróneos que otro proceso no puede tener e implicaría que las políticas para cada proceso sean más o menos restrictivas.

5. **Privacidad y seguridad**: El principio de privacidad y seguridad establece que se deben implementar medidas adecuadas para proteger los datos de accesos no autorizados, pérdidas o mal uso. Esto implica establecer políticas de seguridad, controles de acceso, encriptación, anonimización y cumplimiento de las regulaciones de protección de datos aplicables, como el Reglamento General de Protección de Datos.

6. **Cumplimiento normativo**: Este principio se refiere a la necesidad de cumplir con las regulaciones y leyes relacionadas con la gestión de datos, como las normativas de protección de datos, las regulaciones sectoriales o las leyes de privacidad. Es fundamental comprender y cumplir con los requisitos legales y regulatorios aplicables para evitar riesgos legales y sanciones.

7. **Estandarización:** Para facilitar la comprensión y el intercambio de datos, se deben establecer estándares y metadatos consistentes. Los estándares definen las estructuras, formatos y vocabularios comunes para asegurar la interoperabilidad de los datos, mientras que los metadatos proporcionan información descriptiva y contextual sobre los datos, facilitando su búsqueda, descubrimiento y uso adecuado. Dicho de otra forma, debe establecerse un lenguaje común que permita de forma unívoca identificar un dato por los diferentes miembros de una organización.

8. **Trazabilidad**: Este principio hace referencia al ciclo de vida de los datos, es decir, a la necesidad de establecer procesos y políticas para gestionar los datos a lo largo de su ciclo de vida. Esto incluye aspectos como la captura, almacenamiento, procesamiento, retención, eliminación y archivo de datos. Se deben definir las reglas y las políticas para gestionar eficientemente los datos desde su creación hasta su eliminación.

9. **Colaboración**: Este principio promueve la colaboración y la transparencia en la gestión de los datos. Implica fomentar la comunicación y cooperación entre los actores involucrados en la organización, así como compartir información relevante sobre los datos de manera transparente. La colaboración y la transparencia facilitan la toma de decisiones basadas en datos y fomentan la confianza y la participación de los interesados.

3.1.4 Beneficios

La implementación efectiva de una estrategia del gobierno del dato en una organización puede proporcionar una serie de beneficios significativos. Teniendo en cuenta los principios y las

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

características descritas en esta memoria, se pueden definir estos beneficios como los principales a la hora de implementar el gobierno del dato:

- Toma de decisiones informada: El gobierno del dato asegura que los datos estén disponibles, sean confiables y estén en consonancia con los estándares de calidad establecidos. Esto permite a los tomadores de decisiones basar sus elecciones en datos precisos y oportunos, lo que a su vez mejora la calidad de las decisiones empresariales.

- Eficiencia operativa: Ayuda a estandarizar los procesos de gestión de datos y promueve la eficiencia operativa. Al establecer políticas y procedimientos claros para la captura, almacenamiento, acceso y uso de los datos, se reducen los errores, los tiempos de búsqueda de datos y los esfuerzos redundantes, lo que ahorra tiempo y recursos.

- Cumplimiento normativo: Facilita el cumplimiento de las regulaciones y normativas relacionadas con la gestión de datos, como la protección de datos personales, la privacidad y la seguridad de la información. Al implementar políticas y controles adecuados, la organización puede mitigar los riesgos legales y proteger la reputación de la empresa.

- Gestión de riesgos mejorada: El gobierno del dato ayuda a identificar y mitigar los riesgos asociados con la gestión de datos. Al establecer controles de calidad, seguridad y privacidad, se reducen los riesgos de pérdida de datos, violaciones de seguridad y mal uso de la información. Esto contribuye a la protección de activos valiosos de la organización.

- Confianza y credibilidad: La implementación adecuada del gobierno del dato fomenta la confianza en los datos utilizados en la organización. Los datos de alta calidad y precisión generan mayor confianza en los informes, análisis y resultados basados en ellos. Esto fortalece la credibilidad de la organización ante sus empleados, clientes y socios comerciales.

- Innovación y colaboración: Facilita la innovación y la colaboración al proporcionar un acceso seguro y controlado a los datos. Al tener una gestión eficaz de los datos, se pueden realizar análisis avanzados, descubrir nuevas oportunidades y fomentar la colaboración entre diferentes áreas y equipos de la organización.

- Optimización de recursos: Al tener una gestión estructurada de los datos, se evita la duplicación y la redundancia, lo que conduce a una mejor utilización de los recursos. Se maximiza el valor de los datos existentes y se evita invertir recursos en la adquisición de datos innecesarios o duplicados.

- Mejora de la calidad de los productos y servicios: Contribuye a la mejora de la calidad de los productos y servicios de la organización. Al tener datos confiables y precisos sobre

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

los clientes, preferencias de compra, patrones de comportamiento, entre otros, se pueden personalizar y optimizar los productos y servicios ofrecidos.

3.1.5 Pasos a dar para la implementación de la gobernanza

La implementación exitosa de la gobernanza de datos en una organización requiere un enfoque planificado y estructurado. A continuación, se presentan diez pasos clave para llevar a cabo una implementación efectiva de la gobernanza de datos:

1. **Definir el alcance y los objetivos**: Para empezar, hay que definir claramente el alcance de la gobernanza de datos en tu organización. Esto implica identificar los datos críticos, los procesos y las áreas de la organización que se verán afectados por la gobernanza. Además, hay que establecer los objetivos claros y medibles que se desean lograr con la implementación.

2. **Crear un equipo de gobernanza de datos**: Se debe formar un equipo dedicado a la gobernanza, compuesto por representantes de diferentes áreas de la organización, como tecnologías de la información, legal, y las unidades de negocio pertinentes. Este equipo será responsable de desarrollar y supervisar las políticas, procesos y controles relacionados con la gobernanza de datos. A continuación, se muestran los roles que son necesarios cubrir y las responsabilidades que debe tener cada uno de ellos. Hay que tener en cuenta que, dependiendo de los recursos asignados al proyecto, una persona puede tener varios roles asignados.

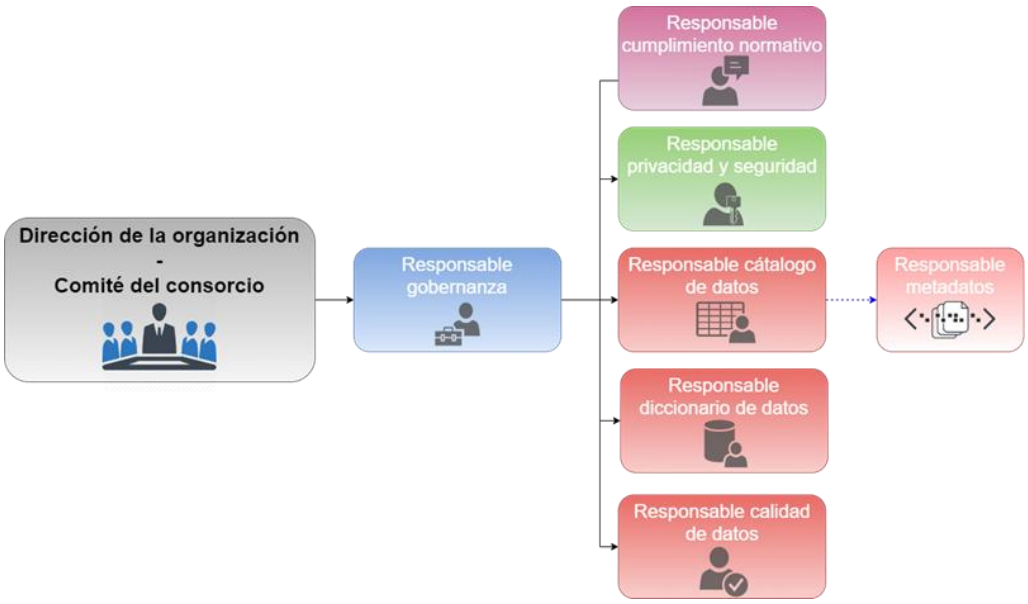


Figura 2. Árbol de roles en la gobernanza del dato.

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

- ***Responsable de la gobernanza***: Este rol es, quizás, el más importante debido a las tareas que tiene por delante y la importancia de las mismas. Es necesario designar a la persona para este rol desde un principio.
    - Definir y establecer la estrategia de gobernanza del dato, comunicándolo al resto de la organización. Esto implica definir los objetivos, los principios, las políticas y los estándares que guiarán todas las actividades relacionadas con la gestión de datos a nivel organizacional.
    - Supervisar la implementación de políticas y estándares de datos: Este rol debe asegurarse de que las políticas y los estándares de datos definidos se implementen de manera efectiva en la organización. Esto implica monitorear y evaluar el cumplimiento de las políticas y los estándares, y tomar medidas correctivas si es necesario.
    - Coordinar y colaborar con otros roles y equipos: Este responsable debe trabajar estrechamente con otros roles y equipos para garantizar la alineación y la coordinación en todas las actividades relacionadas con la gobernanza del dato.
    - Realizar auditorías y evaluaciones periódicas: Este rol debe llevar a cabo auditorías y evaluaciones periódicas para evaluar la eficacia de la gobernanza del dato, identificando áreas de mejora, acciones correctivas y asegurando la mejora continua de las prácticas de gobernanza del dato.
    - Promover la conciencia y la capacitación en gobernanza del dato: Este responsable debe fomentar la conciencia y la comprensión de la importancia de la gobernanza del dato entre los empleados de la empresa. Para ello puede llevar a cabo la organización de sesiones de capacitación, la elaboración de materiales educativos y buenas prácticas, etc.
    - Mantenerse actualizado sobre las mejores prácticas, tendencias, nuevas regulaciones, herramientas y tecnologías emergentes, etc., en el ámbito de la gobernanza del dato.
- ***Responsable del catálogo de datos***:
    - Mantener y administrar el catálogo de datos corporativo.
    - Clasificar los conjuntos de datos y mantener una estructura organizada en el catálogo.
    - Documentar los metadatos relevantes de los conjuntos de datos, como nombre, descripción, origen, propietario, permisos de acceso y calidad. Esta tarea puede ser extensa y podría, dependiendo del detalle que se desee alcanzar, dar lugar a un nuevo rol como *responsable de metadatos*. Esta función la puede desempeñar otra persona.
    - Actualizar y enriquecer continuamente el catálogo con nuevos conjuntos de datos y cambios en los existentes.

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

- Garantizar la accesibilidad y la usabilidad del catálogo de datos para los usuarios y los responsables de datos.

- ***Responsable del diccionario de datos***:
  - Mantener y administrar el diccionario de datos corporativo.
  - Definir y documentar las definiciones y descripciones claras de los elementos de datos utilizados en la plataforma.
  - Establecer estándares y convenciones de nomenclatura para los elementos de datos.
  - Colaborar con los responsables de los conjuntos de datos para asegurar que los elementos de datos estén correctamente definidos y documentados.
  - Actualizar y enriquecer continuamente el diccionario de datos con nuevos elementos de datos y cambios en los existentes.
  - Facilitar la comprensión y la colaboración en torno a los elementos de datos, promoviendo la consistencia y la uniformidad en su uso.
- ***Responsable de la calidad de los datos***:
  - Definir y establecer estándares y métricas de calidad de datos para evaluar y mantener la integridad y la precisión de los conjuntos de datos.
  - Realizar pruebas y análisis de calidad de datos, identificar problemas y proponer acciones correctivas.
  - Supervisar y realizar auditorías periódicas para garantizar el cumplimiento de los estándares de calidad de datos.
  - Colaborar con los equipos responsables de los conjuntos de datos para mejorar continuamente la calidad de los datos.
- ***Responsable de Cumplimiento Normativo***:
  - Monitorear y asegurar el cumplimiento de las regulaciones y normativas aplicables a la gestión de datos en el sector industrial.
  - Establecer políticas y procedimientos para garantizar que los datos se manejen de acuerdo con las regulaciones, como protección de datos personales o requisitos específicos del sector afectado.
  - Realizar auditorías internas y preparar informes de cumplimiento normativo.
  - Mantenerse actualizado sobre los cambios en las regulaciones y actualizar las políticas y los procesos en consecuencia.
- ***Responsable de la privacidad y la seguridad de los datos***:
  - Definir y aplicar políticas y medidas de seguridad de datos para proteger la confidencialidad, la integridad y la disponibilidad de los conjuntos de datos.

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

- Establecer controles de acceso y permisos adecuados para garantizar que los datos sean accesibles sólo para usuarios autorizados.
- Implementar tecnologías y prácticas de seguridad, como encriptación de datos, monitoreo de acceso y prevención de brechas de seguridad.
- Realizar evaluaciones de riesgos y gestionar incidentes de seguridad de datos.

3. **Realizar un inventario de datos**: Realizar un inventario exhaustivo de los datos de la organización, identificando su ubicación, formato, calidad, propietarios y usuarios. Esto ayudará a comprender el panorama de datos de la organización y a establecer una base sólida para la gobernanza. Es conveniente empezar por un glosario empresarial, es decir, un vocabulario con definiciones de algunos de los datos más relevantes para la empresa, identificando algunos procesos clave que utilizan estos datos, para después ir incrementando el detalle y la complejidad. Esta labor tiene que ser llevada a cabo entre los roles *responsable del catálogo de datos* y *el responsable del diccionario de datos*.

4. **Establecer políticas y estándares de datos**: Se deben desarrollar políticas claras y estándares para la gestión de los datos. Estas políticas deben abordar aspectos como la calidad de los datos, la seguridad, la privacidad, el cumplimiento normativo y la gestión del ciclo de vida de los datos. Estas políticas deben ser comprensibles y aplicables a todos los empleados de la organización. Esta labor tiene que ser llevada a cabo por el rol de *responsable de la gobernanza* con la participación del resto de roles orientados a datos.

5. **Definir roles y responsabilidades**: Establecer roles y responsabilidades claras para la gobernanza de datos. Se debe definir quiénes serán los responsables de la calidad de los datos, la seguridad, la privacidad y el cumplimiento normativo. Se asigna a personas o equipos específicos la responsabilidad de monitorear y hacer cumplir las políticas de gobernanza de datos. Esta labor tiene que ser llevada a cabo por el rol de *responsable de la gobernanza*.

6. **Implementar procesos y controles**: Se debe definir y documentar los procesos y controles necesarios para asegurar el cumplimiento de las políticas de gobernanza de datos. Esto puede incluir procesos para la captura, validación, almacenamiento, acceso y eliminación de datos, así como controles para garantizar la seguridad y la privacidad de los datos. Esta labor tiene que ser llevada a cabo conjuntamente entre los roles de *responsable de la gobernanza* y *responsable de la privacidad y la seguridad de los datos*.

7. **Establecer un marco de calidad de datos**: Implementar un marco de calidad de datos que aborde la evaluación, la mejora y el monitoreo continuo de la calidad de los datos. Se deben definir métricas y criterios claros para medir y mejorar la calidad de los datos, y establecer mecanismos de retroalimentación para corregir y prevenir problemas de calidad. Esta labor tiene que ser llevada a cabo por el rol de *responsable de la calidad de los datos* y su equipo (si lo tuviese).

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

8. **Capacitar y concienciar**: Hay que proporcionar capacitación adecuada a los empleados sobre los conceptos y prácticas de la gobernanza de datos, asegurándose de que comprendan la importancia de la gobernanza y su responsabilidad en la gestión adecuada de los datos. Esta labor tiene que ser llevada a cabo por el rol de *responsable de la gobernanza*.

9. **Monitorear y auditar**: Se deben establecer mecanismos de monitoreo y auditoría para asegurar de que las políticas y los controles de gobernanza de datos se estén aplicando y cumpliendo de manera efectiva. Hay que llevar a cabo evaluaciones periódicas de la gobernanza de datos para identificar áreas de mejora y ajustar las políticas y procesos según sea necesario. Esta labor tiene que ser llevada a cabo por el rol de *responsable de la gobernanza*.

10. **Mejora continua**: La gobernanza de datos es un proceso continuo. A medida que evoluciona la organización y sus necesidades de datos, es importante revisar y mejorar constantemente los enfoques y las prácticas de gobernanza de datos. El equipo o persona responsable debe mantenerse al tanto de las mejores prácticas y las nuevas tecnologías relacionadas con la gobernanza de datos y realizar ajustes según corresponda. Esta labor tiene que ser llevada a cabo por el rol de *responsable de la gobernanza*.

Es fundamental que para una implementación exitosa de la gobernanza de datos se consiga el compromiso y el apoyo de toda la organización. Es un esfuerzo conjunto que requiere la colaboración de diferentes áreas y la participación activa de los empleados en la gestión responsable de los datos.

3.2. Configuración de un espacio de datos: Rol de la Autoridad de Gobierno

Los espacios de datos se rigen por un conjunto de principios, procesos y prácticas, conocido como el Marco de Gobernanza del espacio de datos, que asegura un liderazgo, control y supervisión responsables. Este marco define las funciones del espacio de datos y los roles de la autoridad de gobernanza y los participantes. Las responsabilidades incluyen la asignación de la autoridad de gobernanza y la formalización de los poderes de decisión de los participantes. Además, especifica los procedimientos para aplicar el marco de gobernanza y resolver conflictos, abarcando tanto aspectos comerciales como tecnológicos2.

[2] https://dssc.eu/space/BVE/357073747/2+Core+Concepts

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

La Autoridad de Gobernanza o el Cuerpo de Gobernanza de un espacio de datos está compuesto por participantes comprometidos con el Marco de Gobernanza. Son responsables de desarrollar, mantener, operar y hacer cumplir este marco. Aunque la Autoridad de Gobernanza maneja tanto las funciones de gobierno como las ejecutivas, no reemplaza a las autoridades públicas de regulación y cumplimiento. Establecer el marco de gobernanza inicial está fuera de su alcance. Además, dependiendo de la forma legal y el tamaño del espacio de datos, estas funciones pueden ser realizadas por una o diferentes partes2.

Para una descripción detallada sobre los aspectos de gobernanza, se recomienda revisar la versión 2.0 del "Data Spaces Blueprint" desarrollado por el DSSC. Los bloques de construcción de gobernanza, junto con los bloques de construcción empresariales, legales y técnicos, proporcionan un marco necesario para que los espacios de datos funcionen de manera eficiente y segura3. De la misma forma, se recomienda la lectura del documento "Rulebook Model for a Fair Data Economy (version 3.0)", publicado por el Fondo de Innovación de Finlandia Sitra4. Si la arquitectura del espacio de datos se basa en la definida por IDSA, se recomienda revisar el "IDSA Rulebook"5. Finalmente, el Marco Regulatorio del espacio de datos Catena-X6 proporciona un ejemplo práctico de definición de gobernanza. Este documento incluye información sobre la soberanía de los datos, los requisitos obligatorios de casos de uso y otras consideraciones legales relevantes para sus participantes.

4. Gestión de participantes, identidad y confianza Roles y actores

4.1 Gestión de identidad / Credenciales

Para crear confianza y mejorar la seguridad, es necesario que la autoridad de gobernanza del espacio de datos establezca mecanismos para la gestión y verificación de las identidades de los

[3] https://dssc.eu/space/BVE2/1071253634/Governance+Building+Blocks
[4] Rulebook model for a fair data economy (version 3.0) - Sitra
[5] https://internationaldataspaces.org/idsa-rulebook/
[6] https://catenax-ev.github.io/docs/next/regulatory-framework/governance-framework

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

participantes en el espacio de datos para garantizar la confianza en los intercambios entre los participantes. Con este fin, un espacio de datos debe poseer ciertas capacidades.[7]

4.1.1 Credenciales Verificables

Las Credenciales Verificables (VCs)[8] son una especificación desarrollada por el Consorcio World Wide Web (W3C) para permitir la creación, el intercambio y la verificación de credenciales digitales de forma segura y respetuosa con la privacidad. Estas credenciales pueden representar diversos tipos de información, como documentos de identidad, certificados educativos, licencias profesionales o roles dentro de una determinada aplicación. La característica principal de las VCs es que están diseñadas a prueba de manipulaciones y son criptográficamente verificables, lo que significa que su autenticidad e integridad pueden confirmarse sin depender de una autoridad central.

El estándar define cómo se pueden emitir, almacenar y presentar las VCs de forma que se garantice su verificación por cualquier parte. Este modelo admite una amplia gama de casos de uso, desde la simple verificación de identidad hasta interacciones complejas entre múltiples partes. Por ejemplo, una universidad podría emitir un diploma digital a un graduado, quien posteriormente podrá presentarlo a posibles empleadores. Los empleadores pueden verificar la autenticidad del diploma sin contactar directamente a la universidad.

Dentro de un espacio de datos, se puede emitir una VC para la identificación e indicar que un participante o alguna entidad (e.g., usuario final, aplicación, dispositivo) cumple con algún requisito, lo que se denomina atestado (*attestation*). Un ejemplo de este tipo de atestado es indicar que una organización participa en un espacio de datos y cumple con todas las normas y obligaciones del mismo. Otro ejemplo de atestado puede indicar que un servicio en particular cumple con características específicas (por ejemplo, el cumplimiento de ciertas especificaciones en cuanto a las API compatibles). Por último, son ejemplos de atestado los que indican que cierto

[7] https://dssc.eu/space/bv15e/766068449/Identity+and+Attestation+Management
[8] https://www.w3.org/TR/vc-data-model-2.0/

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

usuario final cuenta con los privilegios ligados a un determinado rol en relación con una determinada aplicación o un conjunto de datos concreto.

4.1.2 Protocolos para la emisión y verificación (autenticación) de credenciales

Existen varios protocolos para la emisión y verificación de credenciales, basados en el estándar de VCs del W3C, que presentan cierta superposición de funcionalidades, pero también diferencias. Además, algunos espacios de datos operativos aún utilizan enfoques más tradicionales para la emisión de credenciales, como los certificados X.509, en combinación con un servicio DAPS (tal como se especificó en los modelos iniciales de arquitectura de referencia de IDSA).

Se prevé un proceso de convergencia tecnológica en el mercado respecto al uso de estos protocolos (y otros relacionados). Por ahora, los más utilizados son: OID4VC y DCP.

Eclipse Dataspace Decentralized Claims Protocol (DCP)

El Decentralized Claims Protocol (DCP)[9], que está siendo desarrollado por Eclipse, define una superposición interoperable a las especificaciones del Dataspace Protocol (DSP) para transmitir identidades de organizaciones y establecer confianza de una manera que preserva la privacidad y limita la posibilidad de interrupciones en la red, contribuyendo a la definición de reglas y procedimientos para la gobernanza del espacio de datos. Permite múltiples Trust Anchors y permite a los participantes gestionar y verificar presentaciones sin necesidad de sistemas de terceros fuera de su control.[10]

Las especificaciones de DCP incluyen:

- Un formato para tokens de identidad autoemitidos
- Un protocolo para almacenar y presentar credenciales verificables y otros recursos relacionados con la identidad de organizaciones

[9] https://projects.eclipse.org/projects/technology.dataspace-dcp/governance
[10] https://github.com/eclipse-dataspace-dcp/decentralized-claims-protocol

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

Un protocolo para que las partes soliciten credenciales a un emisor de credenciales

Este protocolo se basa en Identificadores Descentralizados (DIDs)[11] y VCs, que eliminan la necesidad de autoridades centrales para verificar identidades, ya que los DIDs permiten a las entidades mantener el control sobre sus identificadores, mejorando la privacidad y la confianza. El Framework EDC utiliza sistemas criptográficos para verificar las reclamaciones, lo que asegura aún más las identidades de los participantes.

OpenID for Verifiable Credentials (OID4VC)

Open ID for Verifiable Credentials (OID4VC)[12] comprende una serie de protocolos que están siendo estandarizados por la OpenID Foundation[13]. En concreto, comprende tres protocolos para gestionar la emisión y verificación de las credenciales:

Open ID for Verifiable Credential Issuance (OID4VCI)[14] define una API y los correspondientes mecanismos de autorización basados en OAuth para la emisión de VCs.

OpenID for Verifiable Presentations (OID4VP)[15] define un mecanismo basado en OAuth 2.0 que permite la presentación de VCs para su verificación.

Self-Issued OpenID Provider v2 (SIOPv2)[16] permite a los usuarios finales utilizar los proveedores de OpenID que controlan.

Los VCs emitidos e intercambiados vía OID4VC pueden estar firmados por entidades identificadas a través de identificadores descentralizados de acuerdo con las especificaciones sobre DIDs[17] de W3C, pero no es estrictamente obligatorio.

[11] https://www.w3.org/TR/did-1.0/
[12] https://openid.net/sg/openid4vc/specifications/
[13] https://openid.net/
[14] https://openid.net/specs/openid-4-verifiable-credential-issuance-1_0.html
[15] https://openid.net/specs/openid-4-verifiable-presentations-1_0.html
[16] https://openid.net/specs/openid-connect-self-issued-v2-1_0.html
[17] https://www.w3.org/TR/did-1.0/

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

OID4VC forma parte de la Arquitectura y Marco de Referencia de EUDI Wallet[18]. Con eIDAS2, todas las personas físicas y jurídicas de la UE podrán recibir una identidad digital almacenable en cualquier monedero conforme a esta arquitectura (EUDI Wallet) para 2026.

OID4VC, en combinación con mecanismos de confianza tales como registros de participantes en un espacio de datos, permite que usuarios (e.g., empleados, usuarios clientes, aplicaciones, dispositivos) ligados a organizaciones participantes puedan autenticarse ante proveedores para consumir los servicios de acceso y procesamiento de datos que éstos ofrezcan en el espacio de datos, sin necesidad registrarse en proveedores de identidad centralizados o asociados a los proveedores. En este sentido, es importante destacar que OID4VC puede emplearse para interacciones tanto H2M como M2M en los procesos de autenticación. También que una organización consumidora de los servicios de un determinado proveedor puede emitir distintas credenciales para los usuarios que ella misma gestiona, gobernando qué usuarios pueden acceder a qué servicios y con qué potestades.

4.2 Certificados / Clearing Houses

Los llamados Gaia-X Digital Clearing House (GXDCH) son nodos donde se ejecutan componentes software desarrollados por el Lab de Gaia-X (de acuerdo a los documentos de especificación y reglas formuladas por los grupos de trabajo de la iniciativa dentro del marco de trabajo del Trust Framework) que permiten la generación de VCs que identifican tanto a los participantes en el espacio de datos como a las ofertas de servicios disponibles en el mismo.

Cada nodo es operado por un proveedor de acuerdo a las reglas definidas por Gaia-X AISBL, que delega la provisión de servicios adicionales de valor añadido en las diferentes instancias de Clearing House. Este enfoque operativo descentralizado, abierto y transparente permite escalar en la creación de espacios de datos basados en los mecanismos de confianza que establece Gaia-X.

[18] https://eu-digital-identity-wallet.github.io/eudi-doc-architecture-and-reference-framework/1.1.0/arf/

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

Los servicios que se ejecutan en los nodos que conforman la red de GXDCHs son los especificados por Gaia-X AISBL, y evolucionan a medida que se completan las fases de desarrollo en el Lab, ofreciendo un creciente número de funcionalidades y servicios de interés para espacios de datos.

En la versión denominada "Tagus" (finalizada en el cuarto trimestre de 2023), existen tres componentes obligatorios (Gaia-X Compliance, Gaia-X Registry y Gaia-X Notarisation Service) encargados de proporcionar la funcionalidad mínima necesaria para generar y verificar la validez de las descripciones de participantes y servicios, así como la información contenida en ellas.

La evolución de estos componentes queda recogida en la denominada versión "Loire", que se prevé esté operativa en el segundo trimestre de 2025, en la que el nivel de conformidad con Gaia-X de los servicios ofrecidos - como se detalla en los siguientes apartados - viene definido por el cumplimiento de una serie de criterios de certificación y buenas prácticas por parte del proveedor en lo referente a aspectos como el marco contractual, protección de datos, portabilidad del servicio, ciberseguridad, normativa europea o sostenibilidad.

Además de los tres servicios mencionados, el GXDCH también puede ofrecer funcionalidades y herramientas adicionales, como un generador de credenciales verificables, un catálogo para poner a disposición nuestra oferta de servicios en el ecosistema o un *wallet* para almacenar credenciales verificables.

4.3 Conformidad con Gaia-X

La intención de las reglas de la política de Gaia-X es identificar controles claros para demostrar los valores europeos fundamentales de Gaia-X: apertura, transparencia, protección de datos, seguridad y portabilidad. La Conformidad Básica define el conjunto mínimo de requisitos para poder participar en el Ecosistema Gaia-X.

Dada la situación en que se encuentran los estándares de Gaia-X, muchos en proceso de concreción, y en algunos casos de definición, las empresas deben valorar un acompañamiento a lo largo del proceso de sus proyectos (al menos hasta final de 2025), que permita el cumplimiento a la fecha de la entrega del mayor valor posible de adecuación a dichos estándares.

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

El cumplimiento de estos objetivos de las reglas de política se puede lograr mediante el cumplimiento de los estándares, certificaciones y códigos de conducta establecidos.

Gaia-X define un marco para el etiquetado de la oferta de servicios presentes en el ecosistema, que introduce un **Cumplimiento Estándar** para los servicios (que se aplica a todos los servicios de todos los proveedores), así como una serie de etiquetas opcionales de servicios en tres niveles de confianza (que representan diferentes grados de cumplimiento con respecto a los objetivos de transparencia, autonomía, protección de datos, seguridad, interoperabilidad, portabilidad, sostenibilidad y control europeo):

Etiqueta Gaia-X nivel 1: Cumplimiento básico con las normas de protección y seguridad de datos de acuerdo con la legislación europea: Una oferta de servicio o producto está certificada como etiqueta Gaia-X nivel 1 si la oferta cumple con el estándar Gaia-X y cumple con las normas europeas adicionales relacionadas con la protección de datos.

Etiqueta Gaia-X nivel 2: La oferta de servicios debe cumplir con los estándares de seguridad y protección de datos de alto nivel, conforme a la legislación europea y debe estar basada en una certificación.

Etiqueta Gaia-X nivel 3: La oferta de servicios debe cumplir el nivel más alto de cumplimiento para servicios que requieren un manejo de datos, seguridad y control legal excepcionales, solo para proveedores europeos.

El Cumplimiento Estándar y las Etiquetas de Nivel 1 son accesibles mediante declaraciones. Las Etiquetas de Nivel 2 y de Nivel 3 necesitan la certificación de terceras partes denominadas Organismos de Evaluación de la Conformidad (*Conformity Assessment Bodies*, CAB).

Las etiquetas Gaia-X se emitirán de acuerdo a los criterios y evaluaciones de forma federada. Las etiquetas emitidas permitirán a la empresa reutilizar las certificaciones existentes para los atributos de servicio subyacentes siempre que sea posible, reduciendo así el costo y la complejidad de adoptar el etiquetado de Gaia-X, especialmente para los servicios existentes ya certificados. Los Procesos de Evaluación definidos por la propia Gaia-X también se basarán en una federación de responsabilidades.

La conformidad de Gaia-X se debe aplicar a todas las ofertas de servicio y deberá haber una Credencial Gaia-X para todas las entidades definidas como parte del modelo empresarial:

- Participante, incluidos el Consumidor y el Proveedor

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

- Oferta de servicios
- Recurso

4.4 Onboarding de participantes en un espacio de datos

Los pasos que debe seguir un participante para incorporarse a un espacio de datos (*onboarding*) que siga la arquitectura Gaia-X, y que utilice por tanto su Trust Framework, son los siguientes:

1. Crear una credencial verificable con un número válido de registro legal, que debe ser firmada por el servicio de *Notarization* de un Digital Clearing House de Gaia-X.
2. Crear y firmar una credencial verificable de participante legal
3. Crear y firmar una credencial verificable de los términos y condiciones de Gaia-X
4. Enviar las credenciales para validación al servicio de *Compliance* de un Digital Clearing House de Gaia-X, y obtener una proof-of-compliance
5. Si el participante es un proveedor de datos, aplicaciones o infraestructura, deberá crear y firmar una credencial verificable con su oferta de servicios
6. Si el participante es un proveedor de datos, aplicaciones o infraestructura, deberá publicar su oferta de servicios en el catálogo del espacio de datos, utilizando el Credential Event Service (CES) de un GXDCH. Este catálogo podrá ser federado.

5. Gestión de datos y servicios

5.1 Productos de datos

En el contexto de los espacios de datos, un **producto de datos** engloba un conjunto de datos o servicios (servicios de acceso a datos o de procesamiento de datos) que constituye la unidad objeto de compartición entre un proveedor y cualquier consumidor, e incluye toda la información necesaria para su descubrimiento, negociación y uso efectivo.

Los datos, que son generados por una persona, organización, dispositivo o software, se intercambian entre proveedores y usuarios de datos a través de productos de datos que incluyen, entre otros, metadatos que describen el producto de datos y, en el contexto de transacciones de datos confiables, los términos de la licencia de datos definidos por el proveedor

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

de datos. De la misma manera, en un espacio de datos, los servicios que forman parte de un producto de datos incluyen metadatos que describen dichos servicios como son las políticas de autorización que se aplican en la invocación a los servicios, o los términos de licencia definidos por el proveedor de aplicaciones.

Es importante señalar que un producto de datos puede variar desde un simple conjunto de datos ofrecidos para su descarga, a una aplicación compleja que integra múltiples servicios de acceso a datos (tanto estáticos como en tiempo cuasi-real) o de procesamiento de datos. Es habitual que los productos de datos combinen *datasets* (conjuntos de datos) y servicios de datos que facilita su acceso, uso e integración por parte de los consumidores.

Un **conjunto de datos** se refiere a una colección estructurada de datos, que puede existir en múltiples distribuciones según su serialización para su transferencia, intercambio o almacenamiento. Cada distribución puede tener un formato específico (p. ej., JSON) y estar optimizada para casos de uso o requisitos técnicos específicos. Un conjunto de datos muy grande puede dividirse en varios conjuntos de datos, conocidos colectivamente como una serie de conjuntos de datos, para una mejor gestión. Los conjuntos de datos de una serie comparten características comunes. Una descripción de alta calidad de un conjunto de datos es esencial, ya que debe capturar exhaustivamente todos estos aspectos.

Un **servicio de datos** proporciona operaciones para la selección, extracción, combinación, procesamiento o transformación de conjuntos de datos, que pueden estar alojados local o remotamente. El resultado de cualquier solicitud a un servicio de datos es una representación de una parte o de la totalidad de un conjunto de datos. El servicio puede estar vinculado a conjuntos de datos específicos, o sus datos fuente pueden configurarse dinámicamente en tiempo de solicitud o ejecución. Esta flexibilidad permite que el servicio se adapte a diferentes necesidades de datos, según la naturaleza de la solicitud. La descripción de un servicio de datos es fundamental para garantizar que se pueda descubrir y utilizar eficazmente en los espacios de datos. La descripción del servicio aclara cómo se accede, transforma y devuelve a los datos, proporcionando todos los detalles esenciales para su correcta invocación. Describe cómo los usuarios pueden seleccionar, extraer o transformar datos y qué resultados esperar, garantizando así que los participantes en el espacio de datos comprendan exactamente cómo funciona el servicio.

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

5.2 Catálogos

Si bien durante la sección 3.3 hemos hecho especial énfasis en listar los principales motivos por los que cualquier actor podrían estar interesado en participar en un espacio de datos, aún no se ha llegado a responder a la siguiente cuestión:

"Si la accesibilidad a los datos es tan importante, pero a la misma vez no debería de tener acceso a ellos a menos que el propietario me autorice, ¿cómo puedo conocer quién ofrece qué y bajo qué condiciones?"

Esta pregunta no es trivial: un espacio de datos no podría considerarse seguro si se tiene acceso indiscriminado a todos los datos de sus participantes, pero, por otro lado, no tener información suficiente puede limitar mi capacidad para conocer si me interesa adquirir un dataset, o cualquier otro servicio que se ofrezca.

Para solucionar este problema, en los espacios de datos se utilizan los denominados **catálogos**. Al igual que el que podrías encontrar en cualquier tienda física, el catálogo es, en esencia, un listado con los diferentes productos cuyo derecho de uso puedes adquirir, además de información adicional que será importante para conocer con mayor detalle cada producto antes de decidir usarlo. Esta información del producto, en el contexto digital, se denomina **metadatos**.

Gracias a los catálogos, un proveedor puede compartir con el resto de los participantes del espacio de datos información (metadatos) tanto de qué tenemos (datos y servicios) como cualquier otra información adicional importante para que el resto de usuarios entiendan el valor de lo que se está ofreciendo.

Un ejemplo de catálogo de datos real, es el de la iniciativa de datos abiertos del Gobierno de España (https://datos.gob.es/es/catalogo). Si bien no forma parte de un espacio de datos, es de gran utilidad para entender conceptualmente qué es un catálogo.

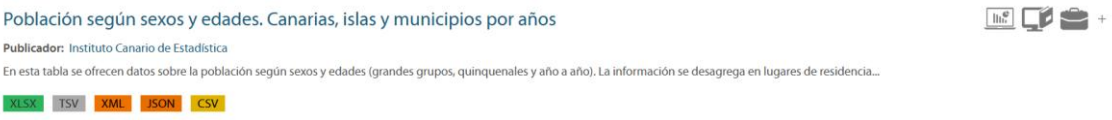


Figura 3. Ejemplo de dataset ofrecido a través de un catálogo de datos (fuente: datos.gob)

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

Por ejemplo, en la figura anterior vemos que podemos acceder a datos de población de las diferentes Islas Canarias. Además, tenemos información adicional (metadatos) que nos indican quién ha publicado el dataset (Instituto Canario de Estadística), una breve descripción, los diferentes formatos en los que se puede obtener, y que son datos relacionados con demografía, educación y empleo.

5.3 Políticas de acceso y contratos

5.3.1 Políticas de uso y acceso (ODRL)

Open Digital Rights Language (ODRL)[19] es un framework estándar para expresar políticas relacionadas con el uso de contenido y servicios digitales definido por W3C. Permite a los titulares de los derechos sobre contenidos/servicios digitales definir reglas y restricciones sobre cómo se pueden acceder, distribuir o consumir dichos activos digitales. ODRL proporciona un formato estructurado y legible por máquinas para especificar permisos, prohibiciones y obligaciones asociadas con el contenido digital, apoyando la gestión de derechos digitales de manera clara e interoperable en diferentes plataformas y servicios.

Una regla ODRL[20] [21] puede tener una o varias de las siguientes partes:

1. Permiso (*permission):* permite que se realice una acción sobre un activo si se cumplen las condiciones y obligaciones.
2. Obligación (*obligation):* una acción requerida que debe ser cumplida, con consecuencias si no se completa.
3. Prohibición (*prohibition)*: prohíbe una acción sobre un activo, con remedios específicos requeridos si se viola.

La especificación ODRL también describe cómo definir perfiles (ODRL *Profiles*[22]), que son extensiones de la especificación ODRL básica adaptadas a casos de uso o industrias específicas. Mientras ODRL proporciona un lenguaje flexible y genérico para expresar derechos digitales, los perfiles se utilizan para personalizar este lenguaje para necesidades particulares definiendo un

[19] https://w3c.github.io/poe/model/
[20] https://w3c.github.io/poe/model/#dfn-rule
[21] https://w3c.github.io/poe/model/#rule
[22] https://www.w3.org/TR/odrl-model/#profile

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

vocabulario, restricciones y tipos de políticas más específicos. Cualquier nuevo ODRL Profile debería seguir las mejores prácticas [23] definidas por el grupo comunitario ODRL [24].

Ejemplos de perfiles existentes de ODRL incluyen:

- Gaia-X ODRL VC Profile[25], que describe cómo incorporar referencias a contenidos de VCs en la parte izquierda de las reglas ODRL
- ODRL Profile[26]: Big Data con un ejemplo de acuerdo de Machine Learning (ML)[27]
- ODRL Temporal Profile[28]

Por ejemplo, imagina que la Empresa A quiere compartir un conjunto de datos con la Empresa B y permitirles entrenar un modelo de ML. Sin embargo, la Empresa A no quiere que la Empresa B comparta el conjunto de datos con nadie.

```
{
  "@context": "http://www.w3.org/ns/odrl.jsonld",
  "@type": "Agreement",
  "uid": "http://example.com/policy:1010",
  "duc:provider": "http://oem.com/ids#me",
   "permission": : [{
          "duc:consumer": "http://supplier.com/"",
          "target": "http://oem.com/ids/inventory/dataset-1",
          "action": "duc:train"",
   }],

  "obligation": [{
          "@type": "Duty",
          "target": "http://oem.com/ids/inventory/dataset-1",
          "action": "inform",
          "consequence": [{
          "target": "http://oem.com/ids/inventory/dataset-1",
                "action":[{
                      "rdf:value": { "@id": "duc:KillJob" },
                      "refinement": [{
```

[23] https://w3c.github.io/odrl/profile-bp/
[24] https://www.w3.org/community/odrl/
[25] https://gitlab.com/gaia-x/lab/policy-reasoning/odrl-vc-profile
[26] https://w3c.github.io/odrl/profile-bigdata/
[27] https://w3c.github.io/odrl/profile-bigdata/#x2-how-to-represent-an-obligation
[28] https://w3c.github.io/odrl/profile-temporal/

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

```
                                "leftOperand": "recipient",
                                "operator": "eq",
                                "rightOperand": {
                                      "@value": "flinkid00",
                                      "@type":
    "http://orion.fiware.org/Job"
                                }
                          }]
                    }]
              }]
        }],

        "prohibition": [{
              "assignee": "http://supplier.com/",
              "target" : "http://oem.com/ids/inventory/dataset-1",
              "action": "print"
        }]
}
```

5.3.2 Integración de políticas de acceso y uso en arquitecturas de referencia de espacios de datos

5.3.2.1 IDS-RAM 4.0

La arquitectura de IDS ("*IDS-RAM*"[29]) admite tanto el control de acceso como el control de uso[30]. Mientras que el control de acceso determina quién puede acceder a los datos, el control de uso aplica restricciones sobre cómo se utilizan los datos después de que se otorga el acceso. Esto permite regular tanto el acceso como el uso posterior, como las obligaciones en el procesamiento de datos o el cumplimiento de requisitos legales y de propiedad intelectual.

Data Access Control aplica los requisitos de provisión de recursos para los Participantes IDS. En el paradigma de *Attribute-Based Access Control* (ABAC)[31], las políticas se representan mediante atributos que describen el sujeto, la acción, el recurso y el contexto de una regla. Según la Arquitectura de Seguridad de IDS, los conectores pueden usar cualquier implementación (por ejemplo, XACML, JAAS) del modelo ABAC para restringir la provisión de datos.

[29] https://docs.internationaldataspaces.org/ids-knowledgebase/ids-ram-4

[30] https://docs.internationaldataspaces.org/ids-knowledgebase/ids-ram-4/perspectives-of-the-reference-architecture-model/4_perspectives/4_1_security_perspective/4_1_6_usage_control

[31] https://www.okta.com/blog/2020/09/attribute-based-access-control-abac/

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

Data Usage Control gestiona las obligaciones de recursos para los participantes IDS. La aplicación de reglas de uso abarca desde soluciones técnicas hasta no técnicas (por ejemplo, reglas organizacionales, contratos legales). IDS utiliza su propio *Usage Control Language*, una extensión del *Open Digital Rights Language* (ODRL), para proporcionar una especificación formal de las reglas. Específicamente, las *IDS Policy Classes* (plantillas de políticas) definen qué operandos, operadores y tipos de datos se pueden usar para una restricción de uso de datos dada. Por ejemplo, la clase de política restringida por tiempo solo acepta valores de tipo xsd:dateTimeStamp[32].

5.3.2.2 Dataspace Protocol (DSP)

En el Dataspace Protocol, el término política (*policy*) se define de la siguiente manera: [33]

> *Un conjunto de reglas, deberes y obligaciones que definen los términos de uso para un Dataset. También se conoce como "Usage Policy".*

Aquí, el término Dataset implica un conjunto de datos publicado por un proveedor en el Data Space. Según el Dataspace Information Model[34], una política de uso se puede asociar con un Dataset de las siguientes maneras:

- ODRL Offer[35], con al menos una requerida por Dataset, o
- ODRL Agreement[36], que resulta de cualquier negociación exitosa entre un proveedor y un consumidor de los participantes.

5.3.2.3 Eclipse Dataspace Components (EDC)

Eclipse Dataspace Components (EDC) es un marco integral (concepto, arquitectura, código, ejemplos) que proporciona un conjunto básico de características (funcionales y no funcionales) que las implementaciones de espacios de datos pueden reutilizar y personalizar aprovechando las API definidas y garantizando la interoperabilidad desde el diseño. Se basa en las especificaciones del Trust Framework de Gaia-X AISBL y en el DSP de IDSA. En EDC las políticas

[32] https://www.w3.org/TR/xmlschema11-2/#dateTimeStamp

[33] https://docs.internationaldataspaces.org/ids-knowledgebase/dataspace-protocol/overview/terminology#policy

[34] https://docs.internationaldataspaces.org/ids-knowledgebase/dataspace-protocol/overview/model#id-2.1.1-context-of-the-dataspace-protocol

[35] https://www.w3.org/TR/odrl-vocab/#term-Offer

[36] https://www.w3.org/TR/odrl-vocab/#term-Agreement

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

de uso y acceso se definen utilizando ODRL. En concreto, las políticas ODRL se pueden utilizar para los siguientes objetivos[37]:

- Definir los requisitos de acceso a datos para datasets en un catálogo
- Actuar como acuerdos contractuales que otorgan acceso a los datos
- Habilitar el control de acceso

Para representar un conjunto reutilizable de políticas ODRL, EDC las agrupa como una *PolicyDefinition*. Las políticas son evaluadas por el motor de políticas de EDC[38]. El siguiente ejemplo usa el namespace edc para representar una duty.

```
{
 "@context": {
   "edc": "https://w3id.org/edc/v0.0.1/ns/"
 },
 "@type": "PolicyDefinition",
 "policy": {
   "@context": "http://www.w3.org/ns/odrl.jsonld",
   "@id": "8c2ff88a-74bf-41dd-9b35-9587a3b95adf",
   "duty": [
     {
       "target": "http://example.com/asset:12345",
       "action": "use",
       "constraint": {
         "leftOperand": "headquarter_location",
         "operator": "eq",
         "rightOperand": "EU"
       }
     }
   ]
 }
}
```

[37] https://eclipse-edc.github.io/documentation/for-adopters/control-plane/#policies-and-policy-definition
[38] https://eclipse-edc.github.io/documentation/for-adopters/control-plane/policy-engine/

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

5.3.2.4 FIWARE Data Space Components

Los componentes para la gestión de autorizaciones en FIWARE, por ejemplo los utilizados en el FIWARE Data Space Connector (DSC), implementan la aplicación de políticas de autorización que se definen en ODRL.

Cabe destacar que se soporta el perfil ODRL VC[39] definido por Gaia-X, lo que permite definir políticas de acceso en función de las afirmaciones (claims) ligadas a credenciales verificables de los usuarios que tratan de acceder a datos o servicios de datos. Así mismo, se soportan perfiles ODRL que permiten establecer condiciones sobre los campos de datos que viajan en las peticiones.

5.3.3 Establecimiento de contratos entre proveedor y consumidor

Una vez conocemos el producto de datos que nos pueda interesar, así como el proveedor dentro del espacio de datos, y estemos de acuerdo con sus políticas de uso, estamos a un paso de conseguir un acuerdo entre proveedor y consumidor.

Es importante saber que la confiabilidad que nos otorga un espacio de datos no viene solo por la una gobernanza del dato efectiva, también nos proveen de una capa legal que nos permite:

- Asegurarnos que el acuerdo entre proveedor y consumidor cumple las legislaciones relevantes.
- Utilizar un marco contractual robusto.

Es decir, como participantes de un espacio de datos, tenemos la certeza de que hay una organización que vela porque se cumpla la legislación, y que además tenemos suficientes herramientas para poder ejercer nuestros derechos y deberes sobre los productos que ofrecemos o consumimos de otros participantes.

[39] https://gitlab.com/gaia-x/lab/policy-reasoning/odrl-vc-profile

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

Específicamente, este contrato entre proveedor y consumidor se denomina *"acuerdo de producto de datos"*, o *"data product contract"* en inglés. Si bien este contrato puede tener gran cantidad de cláusulas, las de mayor importancia son las enumeradas por el Data Spaces Support Centre:

- **Identificación**: el contrato debe incluir un identificador inequívoco de los datos sobre los que se establece el contrato.
- **Usos aceptables**: como proveedor, puedo definir para qué fines puede ser utilizado un producto de datos.
- **Restricciones de uso**: una forma análoga a la anterior cláusula es definir los usos específicos en los que se prohíbe el uso de nuestros datos.
- **Derechos de uso**: se pueden definir derechos que se otorgan al consumidor, como por ejemplo la posibilidad de distribuir o modificar los datos.
- **Rescisión del contrato**: incluiría las condiciones por las que el contrato puede ser terminado, esencial para proteger a ambas partes.
- **Condiciones relacionadas con el uso del material derivado**: si se desarrolla un producto derivado, se pueden establecer reglas sobre cómo se pueden utilizar o comercializar.
- **Tarifas y monetización**: define cómo se va a monetizar el producto de datos.
- **Seguridad de los datos**: lista medidas de seguridad que se necesiten implementar para proteger los datos.
- **Cláusulas relacionadas con los derechos de terceros**: indica cómo se protegerán los derechos e intereses de terceros, asegurando el cumplimiento de leyes de protección de datos, respetando derechos de propiedad intelectual y manteniendo la confidencialidad de la información sensible.

Como vemos, tener un marco contractual supone un cambio de paradigma en la monetización del dato. Gracias a los espacios de datos, podemos tener la certeza a nivel contractual de que mis productos de datos se van a explotar de la forma que defina como participante. Por otro lado, como consumidor voy a conocer específicamente cuáles son las cláusulas que debo seguir para la explotación de datos de otros participantes.

Pongamos varios ejemplos ficticios para entender la importancia de este contrato:

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

Como proveedor de un producto de datos anonimizados relacionados con el rendimiento de estudiantes presentados a una oposición, podré permitir el desarrollo de subproductos como modelos de inteligencia artificial en el caso de que vayan a ser utilizados por el consumidor de datos. Bajo ningún concepto permitiré la comercialización del subproducto.

Como proveedor de datos generados por robots humanoides industriales, estableceré en el contrato que el cliente puede desarrollar y comercializar subproductos basados en estos datos, como modelos de inteligencia artificial. Sin embargo, exigiré que se reconozca la fuente de los datos en cualquier subproducto comercializado para mejorar visibilidad y reputación de mi empresa.

Desde un punto de vista técnico, este establecimiento de contratos dependerá de la arquitectura y los estándares adoptados para implementar los mecanismos de contratación de productos por el espacio de datos en el que quieras participar. Iniciativas como Gaia-X e IDSA ofrecen marcos y protocolos que facilitan el intercambio seguro y soberano de datos. En las secciones posteriores conoceremos con mayor profundidad cómo se negocian estos contratos desde un punto de vista técnico.

5.4 Marketplaces

Dentro de un espacio de datos, el establecimiento de contratos entre proveedores de productos de datos y consumidores de esos productos puede establecerse directamente (peer-to-peer) o mediante la intermediación de un *marketplace* desplegado como servicio global.

Un *marketplace* integra las funciones de catálogo descritas anteriormente e incorpora funciones adicionales que permiten a consumidores cerrar contratos en torno a ofertas sobre productos de datos que proveedores hacen visibles a través del marketplace. El marketplace notifica al proveedor de un producto de datos cuando un contrato ha sido suscrito por una determinada organización consumidora, de forma que dicho proveedor sepa que, a partir de ese momento, debe permitir acceso a los datos y servicios de aplicación ligados al producto tanto a programas como a usuarios vinculados con dicha organización consumidora. Un marketplace puede ofrecer servicios de cobro a consumidores y pago a proveedores de productos de datos, liberando a

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

estos últimos de la complejidad de implementar la infraestructura y cumplir las condiciones requeridas para ello.

TM Forum especifica un conjunto completo de APIs (TM Forum Open APIs) que facilitan a proveedores de productos de datos registrar especificaciones de sus productos, así como ofertas alrededor de los mismos en un catálogo. Empleando dichas APIs los consumidores pueden navegar por el catálogo y negociar la contratación de un determinado producto de datos. Dichas APIs permiten así mismo que proveedores de productos de datos conozcan cuando un determinado producto es contratado y por quién, así como generar registros de uso de los productos por parte de consumidores (base para la implementación de esquemas de pago por uso) o gestionar el ciclo de vida de productos y ofertas alrededor de los mismos. En este sentido, los APIs de TM Forum constituyen un estándar de referencia a la hora de conseguir la interoperabilidad en relación con funciones de gestión del ciclo de vida de productos de datos y ofertas alrededor de los mismos, así como la gestión de la contratación de productos de datos.

A través del proyecto DOME (Distributed Open Marketplace for Europe)[40], la Comisión Europea ha puesto en marcha una infraestructura que permite la federación de marketplaces, de forma que un proveedor de productos de datos, además de otros productos digitales, pueda registrar sus productos una sola vez y que dichos productos sean ofertados automáticamente en todos aquellos marketplaces dentro de la federación que cumplan determinados criterios establecidos por el proveedor. En este punto, cabe destacar que la integración de marketplaces y proveedores con dicha infraestructura se apoya en el uso de los APIs de TM Forum.

6. Intercambio de datos

La diversidad de estándares e implementaciones de estos estándares para la compartición de datos puede generar incompatibilidades que den lugar a silos de datos dentro de un espacio de

[40] https://dome-marketplace.org/

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

datos. Por ello, es fundamental la convergencia y asegurar la interoperabilidad en el intercambio de datos.

En este contexto, los conectores surgen como la pieza fundamental para garantizar la interoperabilidad en el intercambio de datos promoviendo la confianza entre las entidades participantes y la soberanía de los datos siguiendo estándares comunes largamente aceptados.

La iniciativa que se ha convertido en el estándar de facto para el intercambio de datos interoperable, confiable y soberano a través de conectores es el DSP de IDSA.

El DSP, especificado por IDSA y en camino hacia su estandarización por parte de la Eclipse Foundation, es un conjunto de protocolos y esquemas para facilitar el intercambio interoperable de datos entre entidades regidas por el control de uso. Cabe destacar que el DSP está compuesto de 3 protocolos a su vez:

- Catalog Protocol: Es un conjunto de protocolos y esquemas para publicar un Catálogo de Conjuntos de Datos siguiendo el ampliamente aceptado Data Catalog Vocabulary (DCAT).
- Contract Negotiation: Es un conjunto de protocolos y esquemas para negociar acuerdos de uso de conjuntos de datos específicos dentro de un catálogo.
- Transfer Process Protocol: Es un conjunto de protocolos y esquemas para acceder a los conjuntos de datos acordados.

Tomando como referencia el DSP el conector como concepto se divide en dos planos: el plano de control y el plano de datos tal y como se muestra en la siguiente figura:

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

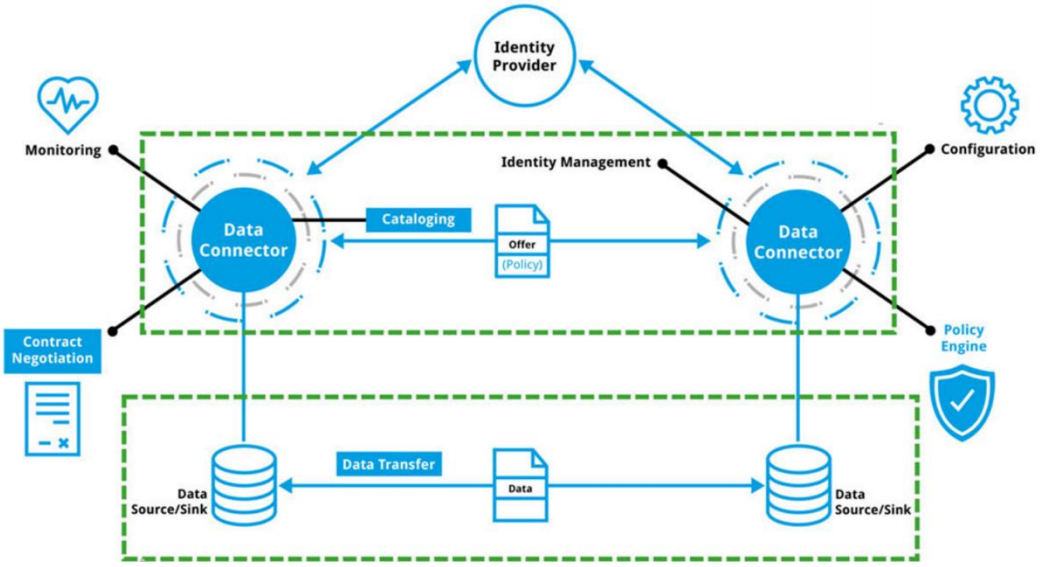


Figura 4. Plano de Control y Plano de Datos.

Mientras que el plano de control se encarga de la gestión estandarizada de los procesos de transferencia entre conectores, el plano de datos está orientado a abordar la transferencia de los datos en cada caso de uso en función de los requerimientos específicos. De esta forma, ofreciendo interoperabilidad en el intercambio de datos a la vez que flexibilidad y escalabilidad.

Sobre esta base, son varios los conectores, que con el objetivo de facilitar la compartición de datos interoperabilidad, confiable y soberana entre organizaciones han implementado el DSP. A continuación, se describen las implementaciones de referencia.

6.1 Conector EDC

Como mencionado anteriormente, el proyecto Eclipse Dataspace Components (EDC) [41] [42] proporciona los siguientes componentes clave: Conector, estructurado en dos módulos: plano de control y plano de datos, Catálogo Federado e Identity Hub. Además, ofrece un Panel de Datos, una aplicación de usuario que permite interactuar con la API de Gestión del Conector EDC. Estos componentes están en desarrollo y se encuentran disponibles bajo licencia Apache 2.0.

[41] https://eclipse-edc.github.io/documentation/
[42] https://eclipse-edc.github.io/documentation/for-adopters/control-plane/

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

La arquitectura del Conector (Figura 5) se diseñó de manera extensible para soportar protocolos alternativos e integrarse en varios ecosistemas. La página web del proyecto EDC en Github ofrece documentación, ejemplos y componentes para implementar un Espacio de Datos Mínimo Viable.

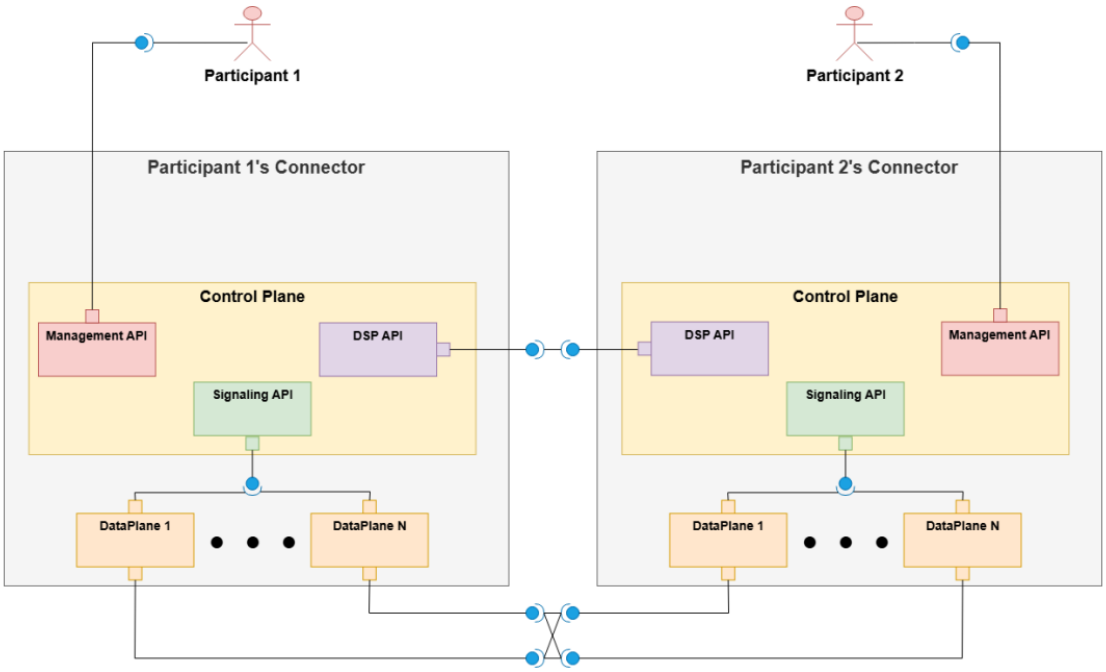


Figura 5. Arquitectura de alto nivel del conector EDC.

6.1.1 Plano de Control

El plano de control es responsable de gestionar las interacciones dentro del espacio de datos, facilitando el intercambio seguro y controlado de datos entre los participantes. Garantiza que el intercambio de datos cumpla con las políticas predefinidas, permitiendo la negociación de contratos, el control de acceso y la orquestación de transferencias. Como se ilustra, el Plano de Control está compuesto por tres APIs principales: API de Gestión, API de Señalización y API DSP.

6.1.1.1 API de Gestión

Esta API proporciona una interfaz para que un participante interactúe con su Plano de Control para realizar tareas administrativas, como gestionar entidades EDC, negociar contratos con otros participantes y gestionar procesos de transferencia.

Las entidades EDC consisten en Activos, Políticas y Definiciones de Contrato. Un Activo sirve como un descriptor de datos en lugar de una entidad de datos física, representando cualquier tipo de datos que se puedan poner a disposición. Esto va más allá de un solo archivo o una colección de archivos: un activo puede ser un flujo continuo de datos, un punto final de API, o

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

incluso un conjunto de cálculos a ejecutar en el futuro. Los activos en EDC se definen utilizando JSON-LD, lo que permite una representación estructurada y semánticamente rica de los metadatos. Cada activo tiene un identificador único, lo que permite que se haga referencia a él en las definiciones de contrato.

Las políticas definen los deberes, derechos y obligaciones asociados con el uso de datos. En EDC, estas políticas se expresan utilizando el Lenguaje de Derechos Digitales Abiertos (ODRL). Cada política tiene un identificador único, lo que permite que se haga referencia a ella al especificar los requisitos de uso para los activos o al controlar el acceso. Esta separación de políticas de los activos añade flexibilidad, permitiendo la creación de políticas corporativas estandarizadas que se pueden reutilizar en diferentes contextos y escenarios.

Las Definiciones de Contrato vinculan activos y políticas especificando qué políticas se aplican a un conjunto determinado de activos. Incluyen dos tipos clave de políticas: una política de acceso, que determina si un consumidor puede acceder a un activo, y una política de contrato, que describe los derechos y obligaciones del consumidor para un activo y es visible para el consumidor si tiene acceso a él.

Negociación de Contratos: La API de Gestión permite iniciar una negociación de contrato, mientras que la negociación en sí se lleva a cabo en la API DSP. El consumidor puede enviar una solicitud para iniciar la negociación a través de la API de Gestión, y el progreso de la negociación se puede monitorear a través de la misma API. Sin embargo, la negociación real es manejada y ejecutada por la API DSP, con la API de Gestión proporcionando una forma de observar su progreso.

Procesos de Transferencia: Una vez que se ha finalizado una negociación de contrato, el consumidor puede iniciar un proceso de transferencia de datos a través de la API de Gestión. Después de que el proceso se inicia, el Plano de Control se comunica con el Plano de Datos a través de la API de Señalización para gestionar la transferencia. La transferencia de datos real ocurre en el Plano de Datos.

6.1.1.2 API DSP

La API DSP se utiliza para la comunicación directa entre los participantes en el espacio de datos. Sus principales responsabilidades se centran en la negociación de contratos y en agilizar el proceso de transferencia.

Negociación de Contratos

La negociación de contratos tiene dos fases: la presentación del catálogo y luego la negociación propiamente dicha. Los proveedores de datos utilizan la API DSP para publicar catálogos en formato Data Catalog Vocabulary (DCAT), enviandolos directamente con otros participantes.

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

Cuando un consumidor solicita un catálogo, el plano de control del proveedor responde dinámicamente, en función de la identidad y las credenciales del consumidor.

Un catálogo contiene información sobre los datos que el proveedor pretende poner a disposición del consumidor solicitante. Cada conjunto de datos incluye una oferta, que define las políticas de uso, describiendo los derechos, deberes y obligaciones que el proveedor impone al consumidor para otorgar acceso a los datos. Además, los conjuntos de datos también contienen distribuciones, que especifican los métodos por los cuales se puede acceder al conjunto de datos, incluidos los protocolos disponibles para la transferencia de datos. Estas distribuciones detallan los mecanismos a través de los cuales los datos se ponen a disposición del consumidor. Cada distribución está asociada con un servicio de acceso, que sirve como el punto final donde se pueden negociar los términos para acceder a los datos, asegurando que ambas partes acuerden las condiciones antes de que los datos sean compartidos.

La API de Gestión proporciona operaciones CRUD para interactuar con las entidades EDC. Cuando se recibe una solicitud de catálogo a través de la API DSP, el Plano de Control transforma las entidades EDC en un catálogo. Primero, evalúa las políticas de acceso del consumidor en todas las definiciones de contrato. Basado en esta evaluación, se seleccionan los activos apropiados a los que el consumidor tiene acceso. Luego, se genera dinámicamente un catálogo que contiene los conjuntos de datos relevantes y las ofertas asociadas. Finalmente, el catálogo generado se presenta al consumidor a través de la API DSP.

Después de recibir un catálogo, un consumidor puede iniciar una negociación de contrato a través de la API de Gestión, pero la solicitud real al proveedor se envía a través de la API DSP. El proveedor reconoce la solicitud y la negociación progresa de manera asincrónica a través de estados DSP predefinidos. Ambas partes pueden desencadenar transiciones, con el Plano de Control iniciador enviando un mensaje DSP a la contraparte. Si la negociación se finaliza con éxito, el consumidor recibe un acuerdo de contrato a través de un mensaje de acuerdo de contrato DSP, que luego se puede usar para iniciar un proceso de transferencia de datos.

Proceso de Transferencia

Durante el proceso de transferencia, la API DSP se utiliza para la comunicación entre el consumidor y el proveedor. El consumidor envía una solicitud para comenzar la transferencia, y el proveedor responde, confirmando que la transferencia ha comenzado. Esto asegura que ambas partes estén sincronizadas antes de que ocurra la transferencia de datos real.

API de Señalización

La API de Señalización facilita la comunicación entre el Plano de Control y el Plano de Datos con respecto a los eventos de transferencia, asegurando la sincronización durante todo el proceso de transferencia de datos. Cuando un consumidor solicita una transferencia de datos, el proveedor recibe el evento REQUESTED. Si el proveedor reconoce la solicitud y pone los datos a

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

disposición para la transferencia, se desencadena el evento STARTED. Durante la transferencia, cualquiera de las partes puede pausar el proceso emitiendo un evento SUSPENDED, deteniendo temporalmente la transmisión de datos, y luego reanudarlo con un evento RESUMED. Una vez que la transferencia se completa con éxito, se envía el evento COMPLETED. Si la transferencia necesita ser detenida forzosamente antes de completarse, se emite un evento TERMINATED.

6.1.2 Plano de Datos

El Plano de Datos es responsable de transmitir datos según las instrucciones del Plano de Control, utilizando un protocolo de comunicación específico. EDC admite dos tipos de transferencia de datos: Extracción por el Consumidor (Consumer Pull) y Envío por el Proveedor (Provider Push).

En una transferencia consumer pull, el proveedor proporciona al consumidor información sobre dónde se pueden obtener los datos, y el consumidor ejecuta la transferencia de datos. Un ejemplo típico es cuando un consumidor envía una solicitud a un punto final del servidor y recibe una respuesta con los datos.

En contraste, las transferencias Provider Push implican que el proveedor envíe los datos al consumidor. Por ejemplo, esto ocurre cuando un consumidor solicita que los datos se entreguen a un punto final de almacenamiento de objetos que controla. En este caso, el consumidor inicia el proceso de transferencia proporcionando al proveedor un token de acceso, que el proveedor utiliza para enviar los datos una vez que están listos.

Un Plano de Datos puede adoptar diversas formas, desde una simple función sin servidor hasta una plataforma de transmisión de datos o una API a la que acceden los clientes. Un solo Plano de Control puede gestionar múltiples Planos de Datos, cada uno especializado en el tipo de datos que se transmiten o en el protocolo de comunicación requerido por el consumidor de datos.

El proyecto EDC proporciona el Marco del Plano de Datos (DPF), una plataforma para construir Planos de Datos personalizados. Un Plano de Datos puede ser creado desde cero o extender el DPF, dependiendo de los requisitos. El DPF está diseñado con la extensibilidad en mente, permitiendo el soporte para diversas fuentes y destinos de datos. Cuando se registra un Plano de Datos, el Plano de Control utiliza metadatos sobre sus capacidades, como los protocolos de comunicación soportados, tipos de transferencia y tipos de fuentes de datos, para determinar qué Plano de Datos utilizar para un proceso de transferencia dado y para generar catálogos con los tipos de transferencia apropiados para los conjuntos de datos.

6.1.3 Catálogo Federado

El Catálogo Federado (FC) es responsable de agregar y almacenar en caché los metadatos de varios participantes. En su núcleo, hay dos componentes clave: el Rastreador del Catálogo Federado (FCC, *Federated Catalog Crawler*) y el Nodo del Catálogo de Destino (*Target Catalogue*

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

Node, TCN). El Rastreador del FC es responsable de recuperar metadatos de fuentes externas y actualizar la caché local, asegurando que el FC se mantenga actualizado con la información más reciente del catálogo. El TCN, por otro lado, es la entidad que almacena y responde a las solicitudes de metadatos, típicamente implementado como un Conector EDC. Cada participante en el espacio de datos aloja su propio TCN, haciendo que la arquitectura sea inherentemente descentralizada.

Cuando se inicia el tiempo de ejecución del catálogo, su caché está inicialmente vacía. A medida que el Rastreador del FC comienza a recopilar datos de los TCN, la caché se llena gradualmente con la información del catálogo recuperada, permitiendo a los clientes realizar consultas sin acceder directamente a cada participante individual.

Tanto el Rastreador del FC como los TCN son componentes conceptuales más que físicos, lo que significa que pueden integrarse como extensiones de tiempo de ejecución, desplegarse de manera independiente o incluso operar en entornos separados. Una sola instancia de tiempo de ejecución también puede alojar tanto un FCC como un TCN simultáneamente.

Para que el FC funcione, necesita una lista de TCN disponibles, que es proporcionada por el Directorio de Nodos de Destino (*Target Node Directory*). Antes de que comience cada rastreo, se consulta este directorio para identificar qué nodos acceder. Este directorio puede ser tan simple como un archivo de configuración estático o tan sofisticado como un registro centralizado que rastrea y gestiona dinámicamente a los participantes.

6.1.4 Centro de Identidad

El Centro de Identidad es responsable de gestionar la identidad de un participante en un espacio de datos. Como se menciona en la Sección 3.2.1, el Centro de Identidad implementa el DCP y almacena y gestiona DIDs y VCs, siendo responsable de la presentación y emisión de estos últimos. Está compuesto por dos API principales: la API DCP, que es responsable de implementar el protocolo de comunicación definido por la especificación DCP, y la API de Identidad, que es utilizada por los usuarios finales para gestionar recursos.

6.1.4.1. API DCP

La API DCP define los mecanismos para la generación de Presentaciones Verificables (VPs), así como los flujos de presentación y emisión dentro del Centro de Identidad. Permite intercambios de credenciales seguros y descentralizados, asegurando confianza, privacidad e interoperabilidad en un espacio de datos.

Esta API permite a los participantes solicitar y presentar VCs sin depender de una autoridad centralizada. El flujo de presentación asegura que un consumidor pueda probar reclamaciones específicas a un proveedor, mientras que el flujo de emisión gobierna cómo se obtienen las credenciales de manera asincrónica de un emisor. Además, el proceso de generación de VP

facilita la creación de pruebas verificables criptográficamente basadas en consultas de credenciales.

6.1.4.2. API de Identidad

La API de Identidad permite a los usuarios gestionar recursos relacionados con la identidad dentro del Centro de Identidad, incluyendo contextos de participantes, DIDs, VCs y pares de claves.

Un contexto de participante define la identidad de una organización dentro de un espacio de datos, actuando como un límite de seguridad para los recursos. Soporta operaciones CRUD y transiciones entre los estados CREADO, ACTIVADO y DESACTIVADO.

Los documentos DID se generan y publican cuando se activa un contexto de participante, haciendo que la identidad de la organización sea resoluble. Actualmente, solo se admite el did:web.

Los recursos de pares de claves se utilizan para firmar y verificar credenciales, presentaciones y otros activos relacionados con la identidad. Las claves privadas se almacenan de forma segura, mientras que las claves públicas se publican en el documento DID, asegurando su verificabilidad. Los pares de claves soportan activación, rotación y revocación para mantener la seguridad.

Los VCs pueden añadirse a través de la API de Identidad. Sigue siendo la única forma de almacenar VCs en el Centro de Identidad porque, como se mencionó anteriormente, el flujo de emisión aún está en fase de desarrollo.

6.1.5 Panel de Datos (Data Dashboard)

El Panel de Datos es una aplicación de interfaz de usuario (UI) de front-end que tiene como objetivo interactuar con la API de Gestión de EDC. Actualmente, solo está destinado a fines de demostración. Dado que la API de Gestión carece de autenticación y autorización integradas, el control de acceso debe manejarse externamente. Esto significa que, en su estado actual, el Panel de Datos no está asegurado por defecto y no debe usarse en producción sin medidas de seguridad adicionales.

6.2 Conector FIWARE

6.2.1 Visión general

El producto FIWARE Data Space Connector (FIWARE DSC) es un conector de espacios de datos, desarrollado como software de código abierto por miembros de la Comunidad FIWARE. La siguiente figura muestra los principales módulos que forman parte del FIWARE DSC, marcando tanto aquellos módulos que están ya disponibles como nuevos módulos que se contempla desarrollar como parte de su roadmap.

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

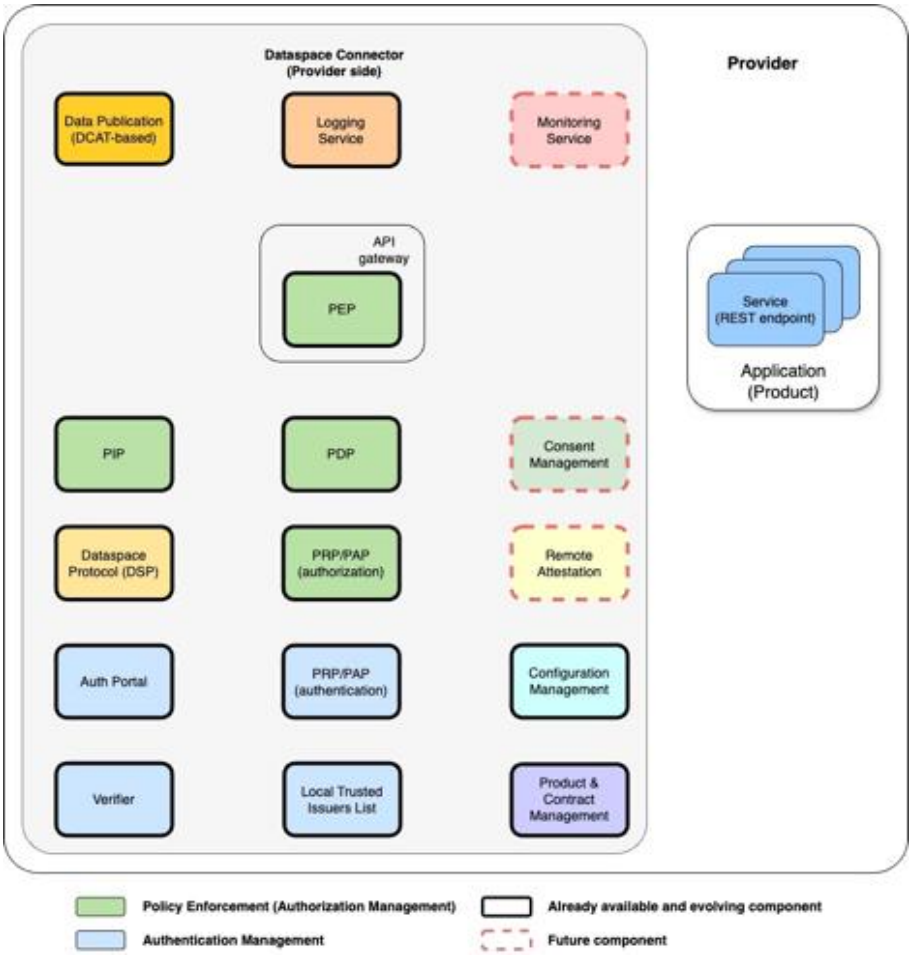


Figura 6. Arquitectura de alto nivel del FIWARE Data Space Connector (DSC)

El FIWARE DSC está concebido fundamentalmente como la integración de dichos componentes basada en recetas (Helm Charts) para su despliegue en entornos Kubernetes, ya sea en servidores de los participantes (desplegados a su vez en entornos Cloud o en propia infraestructura, i.e., "on premise") o en entornos de agentes intermediarios que implementen el concepto de Data Space Connector as a Service (DSCaaS) o Data Space as a Service (DSaaS).

La arquitectura de referencia del FIWARE DSC contempla una serie de módulos integrados para:

- **Autenticación (Authentication)** de usuarios (usuarios finales y aplicaciones);
- **Autorización (Authorization)**, que determina en función de políticas establecidas si un usuario previamente autenticado puede realizar el acceso a un determinado servicio de datos, evaluando información disponible del usuario, servicio solicitado, campos/propiedades de los datos a acceder o procesar, así como características del entorno;
- **Gestión de Productos y Negociación de Contratos (Product and Contract Management)** que gestiona por un lado todo lo relativo a registro de especificaciones

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

de productos (entendidos como conjunto de servicios y recursos necesarios para la ejecución de los mismos) y ofertas alrededor de los mismos y que, por otro lado, gestiona los procesos de contratación de productos por parte de consumidores;

- **Publicación de Catálogo (Catalog Publication)** que registra el conjunto de servicios de datos, así como ofertas alrededor de los mismos en el formato estándar DCATv3;
- **Dataspace Protocol (DSP)** que soporta los protocolos Catalog, Contract Negotiation y Transfer Process, parte del DSP especificado por IDSA;
- **Gestión del Consentimiento (Consent Management)** que facilite la verificación del consentimiento por parte de usuarios finales de procesamiento de datos enviados a servicios ofrecidos a través del conector, particularmente datos personales
- **Registro (Logging)** que almacena información sobre los distintos pasos ejecutados en el curso de las transacciones mediadas por el conector
- **Monitorización (Monitoring)** que permite monitorizar la operación del conector, manejando conceptos propios del mismo vs bajo nivel;
- **Atestación Remota (Remote Attestation)** que permite verificar el software desplegado dentro conector

Estos módulos cubren las funcionalidades básicas que debe soportar cualquier componente conector para espacios de datos de acuerdo con las especificaciones IDS RAM 4.0 y han sido desarrollado en completo alineamiento con las últimas recomendaciones de convergencia tecnológica[43] publicadas por la alianza DSBA (Data Spaces Business Alliance)[44] creada por Gaia-X, FIWARE Foundation, IDSA y BDVA. A continuación, se describen los módulos más relevantes desde el punto de la interoperabilidad en espacios de datos.

6.2.2 Módulo de Autenticación basado en Gestión de Identidad Descentralizada

Este módulo soporta mecanismos de autenticación basados en una gestión de identidad descentralizada siguiendo los estándares W3C (DID, VCs) e implementando la familia de protocolos OID4VC definidos por OpenID Foundation para el intercambio de VCs (SIOPv2, OID4VP)[45]. Este módulo permite que los usuarios (aplicaciones o usuarios finales) que quieran acceder a servicios de datos disponibles en el espacio de datos y cuenten con los VCs requeridos

[43] https://data-spaces-business-alliance.eu/download/33968

[44] https://data-spaces-business-alliance.eu/

[45] OpenID también define el protocolo OID4VCI para la emisión de VCs, pero este protocolo no tiene por qué estar soportado a nivel del módulo de autenticación que todo conector debe soportar.

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

para ello, puedan autenticarse frente al conector que media el acceso a dichos servicios y obtener un token válido (token JWT) con el que poder invocarlos.

Al estar basado en la familia de protocolos OIDC, el módulo de autenticación soporta esquemas de interacción H2M (*Human-to-Machine*) y M2M (*Machine-to-Machine*). Esta familia de protocolos ha sido identificada como de obligado cumplimiento en el marco de la iniciativa EU Digital Identity[46] (ver sección 4.2.1 "*Interfaces and protocols*" dentro de las especificaciones del *EUDI Wallet Architecture and Reference Framework*).

La siguiente figura ilustra cómo un individuo, desde una aplicación móvil o vía web, invoca servicios de acceso a datos o de procesamiento de datos publicados por un determinado proveedor que participa en el espacio de datos, tras autenticarse usando VCs (escenario H2M). Estas VCs habrán sido asignadas por una organización que haya ganado la facultad de emitir VCs siguiendo los pasos que veremos en la sección ligada al módulo de Gestión de Productos y de Negociación de Contratos.

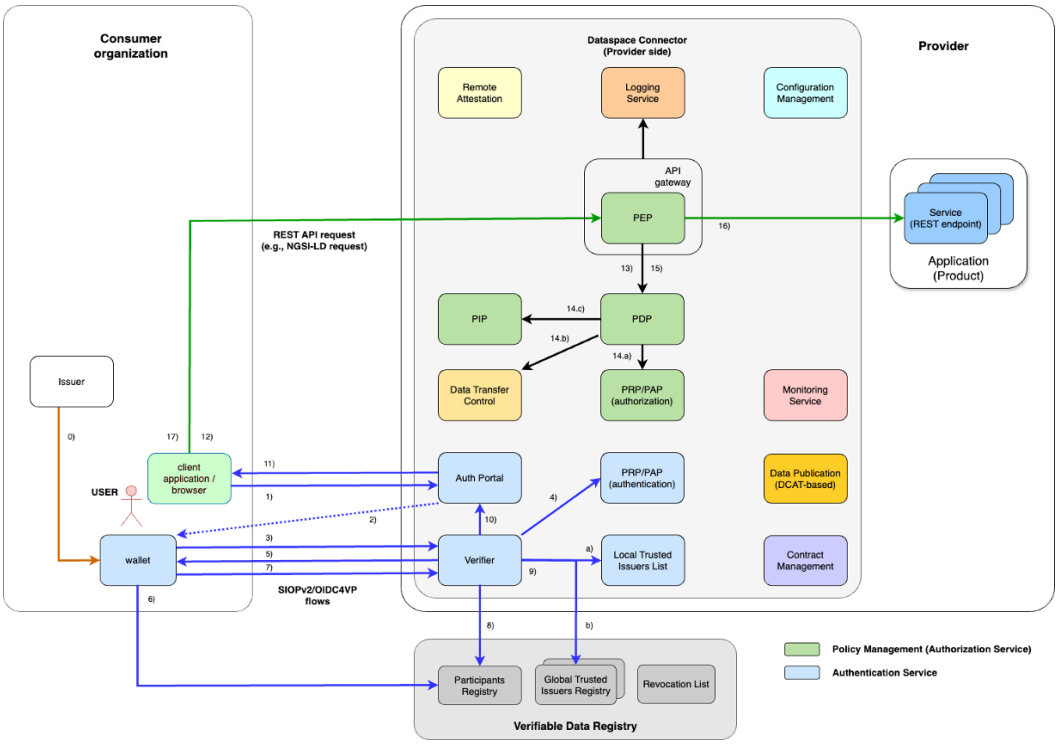


Figura 7 - Invocación de servicios mediados via FIWARE DSC por parte de un usuario final

[46] https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-digital-identity_en

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

A continuación, se explican los pasos reflejados en el diagrama:

- Un usuario que desea usar los servicios ligados a un determinado producto que un proveedor ofrece a través del FIWARE DSC inicia el proceso de autenticación contra un portal del proveedor que típicamente presenta un código QR que el usuario escaneará con su billetera electrónica - (digital wallet). Como resultado de efectuar este escaneado, el wallet lanza una petición al componente Verifier que reside en el conector (pasos 1-3).
- El componente Verifier responde solicitando al wallet del usuario aquellos credenciales verificables (Verifiable Credentials - VCs) que acrediten a) que el usuario posee las credenciales necesarias para solicitar acceso a los servicios del producto, además de otros VCs que se utilizarán para personalizar la experiencia (pasos 4-5). El wallet chequea que el Verifier que está solicitando las VCs es un Verifier ligado a un participante fiable del espacio de datos (paso 6 que siempre es necesario para evitar que alguien pueda interactuar con el wallet usurpando la identidad del proveedor del servicio) y, en caso de que el usuario confirme su acuerdo, envía los VCs solicitados al Verifier enviando una petición contra un punto que éste le ha indicado en su petición (paso 7).
- El Verifier chequea que las VCs del usuario fueron firmadas por una organización que, además de ser un participante fiable del espacio de datos (paso 8), es una organización que había sido dada de alta como emisora fiable de los VCs requeridos para solicitar acceso a servicios ligados al producto ofrecido por el proveedor (paso 9.a). Si solicitó que se enviaran otros VCs que sólo pueden firmar emisores de VCs globales al espacio de datos, chequeará también que esos otros VCs vienen firmados por emisores de VCs marcados como fiables en el espacio de datos (paso 9.b)
- Si todas las verificaciones se completaron con éxito, el Verifier genera un token JWT (paso 10) que transmite a la aplicación del ciudadano a través del portal de autenticación (paso 11)
- Empleando el token entregado, el usuario a través de su aplicación o a través del navegador invocará el servicio, por ejemplo, ligado a la invocación de un API REST, el acceso a una página web, la descarga de un fichero contenido datos, etc (paso 12) y obtendrá una respuesta (paso 14). Dicho token contará con un tiempo de validez, transcurrido el proceso de autenticación debe repetirse.

La siguiente figura ilustra el proceso de autenticación en el escenario M2M.

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

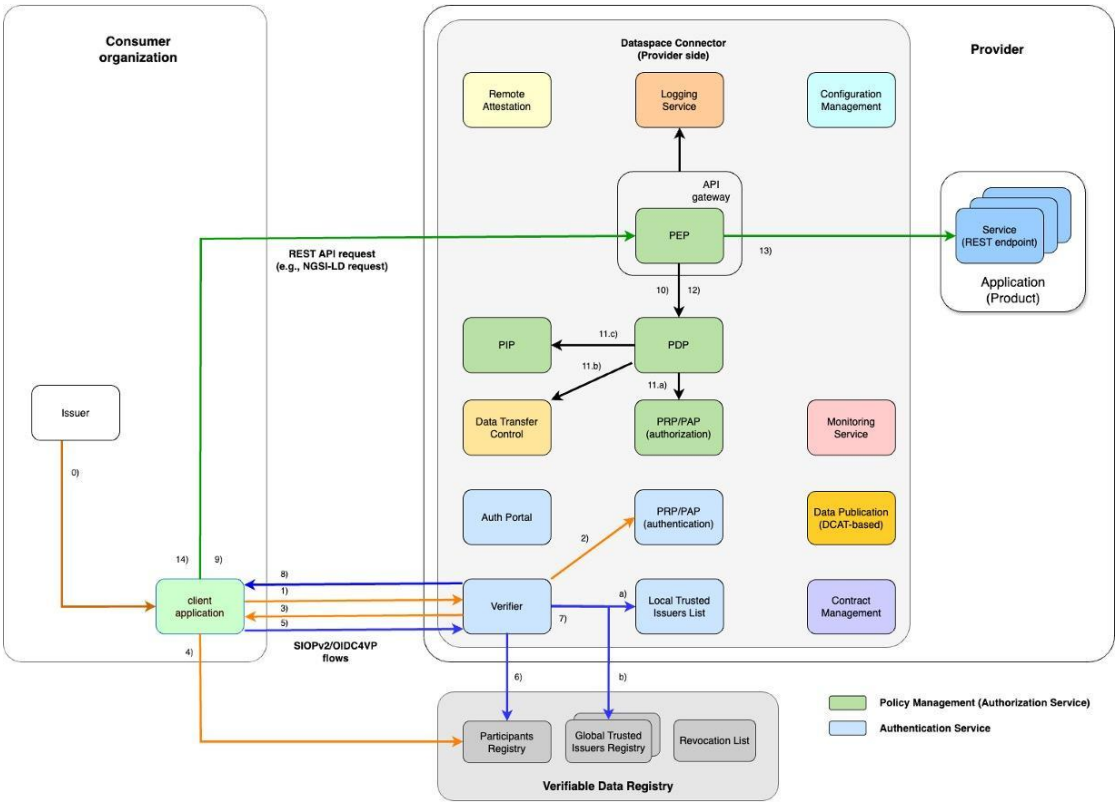


Figura 8 - Invocación de servicios mediados via FIWARE DSC por parte de un sistema o dispositivo

El FIWARE DSC está diseñado para ser compatible con un registro/embarque (onboarding) de organizaciones como participantes en el espacio de datos basado en los mecanismos de confianza definidos por Gaia-X. Dicho registro/embarque supone efectuar los siguientes pasos:

que la organización se auto-emita una credential verificable del tipo LegalParticipantCredential (cuyo formato está especificado por Gaia-X) que describa la organización y presente dicha credencial a un Gaia-X Digital Clearing House (GXDCH) oficial reconocido, para que éste emita para dicha organización una credencial verificable del tipo CompliancyCredential.

A continuación, la organización acudirá a un servicio de registro/embarque en el espacio de datos aportando ambas credenciales con el propósito de registrarse en el espacio de datos. Dicho servicio a) confirmará que la ComplaincyCredential fue efectivamente emitida por un GXDCH oficial (lo que requiere que dicho servicio acuda al servicio Registry de cualquier GXDCH), y b) chequeará si la organización cuya descripción se extrae de la credencial LegalParticipantCredential cumple con las políticas establecidas para la aceptación de organizaciones en el espacio de datos, en cuyo caso registrará la organización en el registro de participantes del espacio de datos (Participants Registry). En este punto, cabe señalar que los FIWARE DSC podrán configurarse para ofrecer este servicio de registro/embarque o asumir la existencia de un servicio de registro/embarque global al espacio de datos.

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

El registro de participantes (Participants Registry) en el espacio de datos sobre el que el módulo de autenticación del FIWARE DSC se apoya puede ser un registro que soporte el *Trusted Issuers Registry* API especificado en EBSI[47] o un registro que soporte el Gaia-X Registry API especificado en Gaia-X[48]. Esta opción se decide por configuración.

6.2.3 Módulo de Autorización

Este módulo se basa en el producto software de código abierto Apache APISIX e implementa una arquitectura que integra componentes que realizan funciones de *Policy Enforcement Point* (PEP), *Policy Decision Point* (PDP), *Policy Information Point* (PIP) y *Policy Administration/Registration Point* (PAP/PRP).

Dentro del módulo de Autorización del FIWARE DSC, el componente PDP toma la decisión de si autorizar o no una determinada petición, aplicando políticas definidas en el lenguaje W3C ODRL que facilita un control de acceso basado en atributos (ABAC). En efecto, estas políticas se formulan en función de: a) campos dentro de las credenciales verificables de los usuarios que viajan con el token de acceso con el que se invocan los servicios, b) el servicio que es invocado, y campos específicos de los datos a acceder o procesar (p.ej., propiedades de entidades cuando se está invocando el API NGSIv2 o NGSI-LD). En este punto, es importante destacar que FIWARE DSC soporta en la actualidad la extensión (profile) para soportar VCs en ODRL especificada por Gaia-X[49].

La Figura 7 y la Figura 8 de la sección anterior mostraban el acceso a servicios ligados a productos de un proveedor mediados a través del FIWARE DSC en escenarios H2M y M2M respectivamente. En ambos escenarios, una vez completada la fase de autenticación y obtenido un token de acceso, los pasos que implementa el módulo de Autorización son los siguientes:

- El componente PEP del módulo de autorización verifica si el consumidor del servicio, contando con los atributos (claims) que se incluyen dentro de los VCs extraídos del token de acceso puede ser autorizado a realizar la petición en cuestión, teniendo en cuenta las políticas definidas por el proveedor del servicio y chequeando si el proceso de transferencia/conversación en el que se enmarca la petición está activa (pasos 13-15 en el escenario H2M y 10-12 en el escenario M2M)

[47] https://hub.ebsi.eu/apis/conformance/trusted-issuers-registry

[48] https://docs.gaia-x.eu/technical-committee/architecture-document/latest/gx_services/#gaia-x-registry

[49] https://gitlab.com/gaia-x/lab/policy-reasoning/odrl-vc-profile

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

- Si se determina que la petición es autorizable, ésta se propaga al servicio (paso 16 en el escenario H2M y 13 en el escenario M2M) y la correspondiente respuesta se retorna al consumidor (paso 17 en el escenario H2M y 14 en el escenario M2M)

6.2.4 Módulo de Gestión de Productos y Negociación de Contratos (incluyendo soporte a monetización)

A través de este módulo es posible gestionar la contratación de productos ofrecidos por un proveedor a través del FIWARE DSC. Típicamente, un producto engloba un conjunto de servicios y contempla la necesaria provisión de una serie de recursos para el funcionamiento de esos servicios. Ejemplos de estos productos pueden ser:

- Productos simples como pueda ser el asociado a un conjunto de datos (dataset) que comprende un servicio para su descarga.
- Productos sofisticados como pueda ser un sistema de gestión de calidad del aire, que abarca todo un conjunto variado de servicios (portal web, servicios API REST de acceso a medidas de calidad del aire en tiempo real o a histórico de medidas, servicios API REST para la invocación de servicios de procesamiento para predicción de parámetros de calidad del aire, etc) y que requieren, para un determinado cliente, la provisión y activación de una serie de recursos tanto en el lado backend (capacidad de memoria, procesamiento y almacenamiento requeridos para atender peticiones del cliente, etc) como en campo (dispositivos IoT para captura de medidas de calidad del aire).

El módulo de Gestión de Productos y Contrataciones permite registrar especificaciones de servicios, recursos y productos que implementa y ofrece un proveedor en el espacio de datos. Así mismo, permite especificar ofertas alrededor de los productos especificados, y gestionar los procesos de adquisición de los derechos de uso de esos productos en función de esas ofertas (contratación) así como la gestión del ciclo de vida de los productos una vez son contratados (contemplando fases iniciales de provisión y activación de los mismos). La adquisición del derecho a usar un producto conlleva el derecho a usar todos los servicios correspondientes. Todas estas funciones se realizan invocando operaciones que el módulo de Gestión de Productos y Contratos implementa cumpliendo las especificaciones TM Forum Open APIs[50]. Esta circunstancia favorecerá la plena integración de espacios de datos que empleen el FIWARE DSC con la infraestructura que implementa el proyecto DOME.

Es importante hacer notar que la lógica de negocio asociada a un determinado producto típicamente contempla la definición de atributos que los usuarios deben poseer para poder invocar los servicios asociados al producto. Así, por ejemplo, un producto para la Gestión de

[50] https://www.tmforum.org/oda/open-apis/directory

Parques y Jardines en ciudades puede incluir servicios de acceso a datos actuales o futuros (modo predictivo) de temperatura y calidad del aire tanto a personal de mantenimiento como a visitantes de esos parques y jardines, pero también servicios de soporte al mantenimiento de jardines (incluyendo servicios para activación de riego) que obviamente sólo pueden ser accesibles para el personal de mantenimiento. En los espacios de datos que utilizan el FIWARE DSC, los atributos de usuarios (en el ejemplo, visitantes y jardineros) se representan usando credenciales verificables. El módulo de Gestión de Productos y Contratos facilita que la organización que contrate un producto sea aceptada como un emisor fiable de las credenciales verificables definidas para ese producto. A partir de este punto, la organización contratante puede asignar, según su criterio, esas credenciales verificables a aplicaciones o usuarios finales (empleados o clientes) de la organización, para que puedan acceder a los servicios del producto contratado. Volviendo al ejemplo, una ciudad que contrate el producto de Parques y Jardínes se convertirá en emisor fiable de las credenciales de jardinero y visitante, que por ejemplo podrá decidir emitir para su personal de mantenimiento y para ciudadanos/turistas, respectivamente.

Una funcionalidad destacada del módulo de Gestión de Productos y de Gestión de la Contratación es el soporte a proveedores en relación con la monetización de productos cuyo acceso es mediado a través del FIWARE Data Space Connector (y, por tanto, la monetización de los servicios asociados a esos productos). En efecto, el FIWARE DSC se puede configurar para soportar funciones de monetización. Esto significa que a la hora de especificar las características de un producto cuyos servicios sean accesibles a través del FIWARE DSC, es posible establecer distintos modelos de monetización (pago una sola vez, pago vía suscripción, pago por uso, etc). El FIWARE DSC cuenta con módulos que permiten llevar la contabilidad en el uso de servicios y la integración con distintas pasarelas de pago.

6.2.5 Soporte a familia de protocolos IDSA DSP

En la actualidad, la creación de servicios de datos y procesamiento de datos, de productos que engloban diversos servicios o de ofertas alrededor de productos a través del módulo de Gestión de Productos y Contrataciones, usando los TM Forum Open APIs, supone la creación de entradas en un catálogo que implementa el FIWARE DSC y que cumple las especificaciones del estándar W3C DCATv3[51]. Las entradas de este catálogo son accesibles a través del protocolo *Catalog Protocol* del IDSA DSP[52], de forma que organizaciones consumidoras pueden consultar que servicios ofrece un determinado proveedor a través del FIWARE DSC. Este API/protocolo es bastante limitado en cuanto a capacidades que ofrece (por ejemplo, no soporta capacidades de cosechado - harvesting). Por ello, dentro del roadmap del FIWARE se plantea soportar APIs

[51] https://www.w3.org/TR/vocab-dcat-3/

[52] https://docs.internationaldataspaces.org/ids-knowledgebase/dataspace-protocol/catalog/catalog.protocol

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

compatibles con DCAT-AP v3 que facilitate la publicación y cosechado (harvesting) por parte de las plataformas de publicación de datos más populares (e.g., CKAN).

El protocolo *Transfer Process Protocol* del IDSA DSP[53] también está implementado en el FIWARE DSC. Cuando quiera que la orden de contratación de un determinado producto (adquisición de derecho a usar el producto visto en el punto anterior) se completa con éxito, se crea un objeto *Agreement* en el catálogo. Empleando este *agreement*, y la identidad de un consumidor, es posible establecer un plano de control bajo cuyo paraguas se realiza el intercambio de datos (o la invocación de servicios), de acuerdo con las especificaciones Transfer Process. En el proceso de autorización de cualquier petición, uno de los aspectos que se verifica es si el proceso de transferencia (conversación) está en estado activo y no ha sido suspendido o terminado.

Por último, en relación con el protocolo *Contract Negotiation* del IDSA DSP[54], el FIWARE DSC soporta un binding basado en las APIs Quote y Product Ordering de TM Forum, que implementa el módulo de Gestión de Productos y Negociación de Contratos descrito en el punto anterior.

6.2.6 Plano de Datos (intercambio)

Los sistemas participantes en muchos espacios de datos emplearán APIs REST para el intercambio de datos. Un ejemplo de este tipo de APIs es el API ETSI NGSI-LD[55], que facilita el intercambio en tiempo cuasi-real de datos ligados a propiedades que describen entidades del mundo real. Este API soporta diversos modelos de intercambio (petición/respuesta, subscripción/notificación y streaming) y emplea JSON-LD[56] como mecanismo de serialización de datos. Sin embargo, el FIWARE DSC está desarrollado con el objetivo de funcionar con cualquier API REST, por ejemplo, APIs como S3 para el intercambio de grandes volúmenes de datos, etc.

A la hora de definir los modelos de datos que describen los distintos tipos de entidades sobre los que los sistemas conectados a un espacio de datos intercambiarán información, la DSBA recomienda reutilizar los modelos publicados en el marco de la iniciativa Smart Data Models[57]. En este punto, el FIWARE DSC se muestra especialmente potente para mediar en escenarios donde, además de utilizar un API REST, la información que se intercambia se define en torno a entidades y sus propiedades, ya que el módulo de autorización puede explotar entonces políticas en un grano fino, considerando no sólo las credenciales de usuarios, variables de

[53] https://docs.internationaldataspaces.org/ids-knowledgebase/dataspace-protocol/transfer-process/transfer.process.protocol
[54] https://docs.internationaldataspaces.org/ids-knowledgebase/dataspace-protocol/contract-negotiation/contract.negotiation.protocol
[55] https://www.etsi.org/committee/cim
[56] https://www.w3.org/TR/json-ld11/
[57] https://smartdatamodels.org/

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

entorno y operaciones invocadas, sino también valor de las propiedades de entidades cuyos datos se intercambian.

6.2.7 Integración con DOME

Es importante señalar que los estándares recomendados por la DSBA en relación con mecanismos de confianza, gestión de identidad y lenguaje de definición de políticas de autorización, así como los estándares TM Forum, han sido adoptados en la arquitectura de DOME. Esto facilita la publicación de productos mediados a través del FIWARE DSC en marketplaces federados con DOME o directamente con el Marketplace principal de DOME.

6.3 Tekniker Dataspace Connector

El Tekniker Dataspace Connector (TDC), considerado como un Participant Agent tal y como se define en el DSSC Blueprint 2.0, es una herramienta de referencia publicada en el DSSC Toolbox desarrollada por Tekniker que, siguiendo estándares ampliamente aceptados, tiene como objetivo permitir a las empresas participar en un espacio de datos proporcionando funcionalidades para publicar y buscar un catálogo de conjuntos de datos, negociar acuerdos de uso para estos y transferirlos de acuerdo a tres pilares principales: interoperabilidad, confianza y soberanía.

En primer lugar, la compartición de datos igualitaria se asegura siguiendo el enfoque en dos planos propuesto para los conectores. Por un lado, el plano de control permite gestionar las transferencias de datos proporcionando las interfaces necesarias para publicar y buscar un catálogo de conjuntos de datos, negociar acuerdos de uso y acceder a los conjuntos de datos acordados de forma estandarizada implementando el Dataspace Protocol 2025. Por otro lado, el plano de datos permite transferir datos desde distintas APIs considerando diferentes tipos de datos desde formatos en texto plano como JSON o XML hasta formatos binarios como PDF validando la correctitud de los datos de forma sintáctica y semántica.

En segundo lugar, la confianza, dependiente de los requerimientos de cada espacio de datos en materia de tecnologías e identidades, se establece soportando diferentes enfoques de gestión de identidades federadas que se pueden dividir en aquellos que utilizan identidades centralizadas y aquellos que utilizan identidades distribuidas. Por un lado, en relación con las identidades centralizadas, una entidad central en el papel de proveedor de identidades emite, mantiene y gestiona las identidades. Siguiendo este enfoque, actualmente se da soporte al

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

estándar OpenID Connect el cual, construido sobre OAuth 2.0, permite a los conectores verificar la identidad de otro conector basada en Json Web Token (JWTs). Por otro lado, en relación con las identidades descentralizadas, una entidad central en el papel de proveedor de identidades emite identidades y el conector las mantiene y autogestiona. Siguiendo este enfoque, actualmente se da soporte al Decentralized Claims Protocol (DCP), un conjunto de protocolos y esquemas basados en OpenID4VC para emitir y presentar identidades basadas en las especificaciones del W3C de Verifiable Credentials v1.1 y Decentralized Identifiers v1.0.

En tercer lugar, la soberanía del dato, una vez establecidos los acuerdos de uso de cada conjunto de datos a través del Contract Negotiation Protocol definido en el Dataspace Protocol 2025 se garantiza mediante la implementación de un motor de políticas cuyo objetivo es la aplicación de estas desencadenado por el plano de datos a lo largo del ciclo de vida del dato.

Con el objetivo de garantizar la interoperabilidad, confianza y soberanía en el intercambio de datos de forma modular, El TDC se divide en 5 componentes principales: Metadata Manager, Base de datos PostgreSQL, Dataspace Protocol, Data Plane y Policy Engine. Partiendo de estos componentes cabe destacar que el TDC se complementa con cuatro herramientas adicionales que facilitan la integración, gestión y el despliegue del mismo en entornos operacionales: Nginx, Swagger UI, PgAdmin y Dozzle. Estos se describen a continuación:

- Nginx: Es una herramienta de alto rendimiento, flexible y ampliamente adoptada para mejorar el rendimiento, la seguridad y la escalabilidad de los servicios en entornos de producción.

- Swagger UI: Se trata de una potente herramienta que permite a los desarrolladores visualizar, explorar e interactuar con los puntos finales de la API en una interfaz interactiva y fácil de usar.

- PgAdmin: Es una herramienta de administración y gestión para bases de datos PostgreSQL.

- Dozzle: Se trata de un visor de registros en tiempo real para contenedores Docker que proporciona una interfaz intuitiva basada en web para supervisar los registros y solucionar problemas en tiempo real, lo que facilita el seguimiento del comportamiento y el rendimiento de los servicios.

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

Tomando como referencia los componentes principales del TDC y las herramientas adicionales, la arquitectura general del TDC se muestra en la siguiente figura, y sus componentes principales se describen en las siguientes subsecciones.

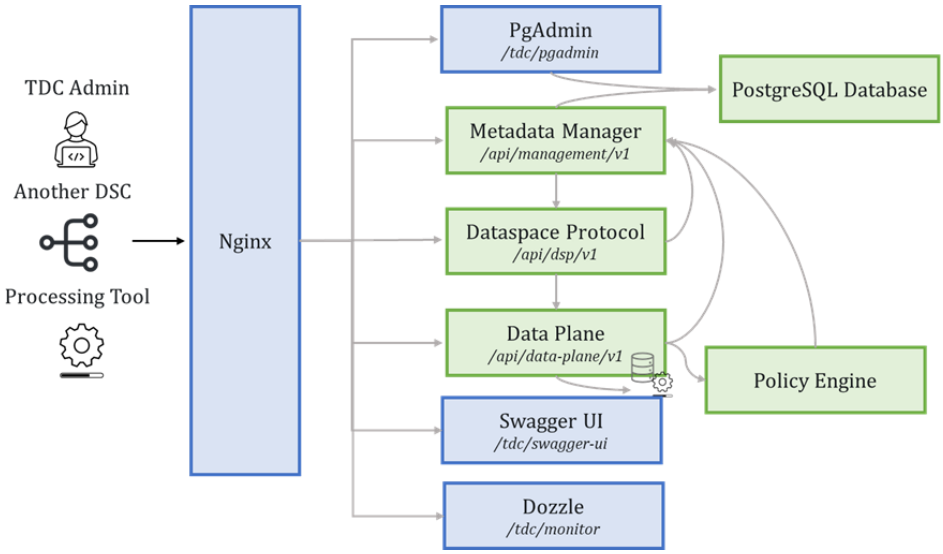
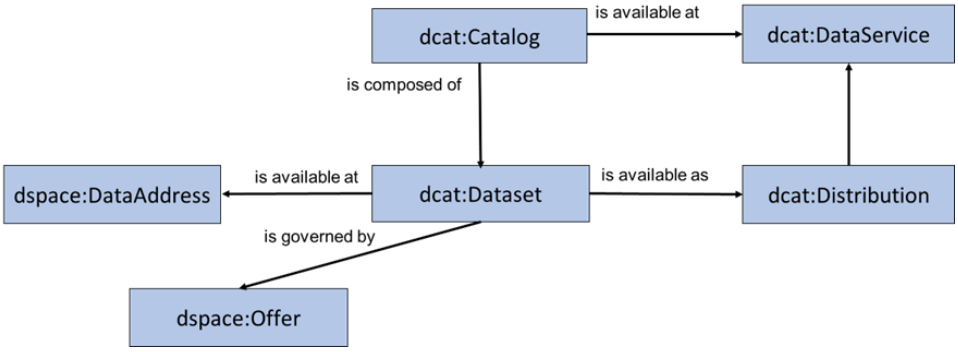


Figura 9. TDC. Arquitectura General.

6.3.1 Metadata Manager

El componente Metadata Manager permite al administrador del TDC gestionar el catálogo de datos ofrecido y los conjuntos de datos solicitados a través del espacio de datos.

En primer lugar, para gestionar tanto la descripción del catálogo de datos ofrecido como el conjunto de datos solicitados es esencial entender el modelo Data Catalog Vocabulary (DCAT) ya que este es utilizado por el componente Metadata Manager para la creación, lectura, actualización y visualización de los diferentes metadatos del catálogo tal y como se ve en la siguiente figura:



Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA
MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL
SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

Figura 10. TDC. Modelo para la descripción del catálogo basado en DCAT.

Este modelo describe cómo un catálogo (dcat:Catalog) está disponible externamente a través de un DataService (dcat:DataService). Cada catálogo está compuesto por una lista de conjuntos de datos (dcat:Dataset), y cada conjunto de datos está asociado a una dirección de datos (dspace:DataAddress), controlada por una oferta (odrl:Offer) y está disponible como distribución (dcat:Distribution) que está disponible externamente en un DataService.

Partiendo de este modelo, el componente Metadata Manager ofrece unas interfaces que, disponibles de forma sencilla a través del Swagger UI, permiten la creación, lectura, actualización y borrado de cada uno de estos conceptos, así como los enlaces entre ellos ofreciendo expresividad y flexibilidad en la descripción de los conjuntos de datos. En las figuras que siguen, se muestra un ejemplo de creación de un Dataset, un DataAddress y en enlace entre ambos.

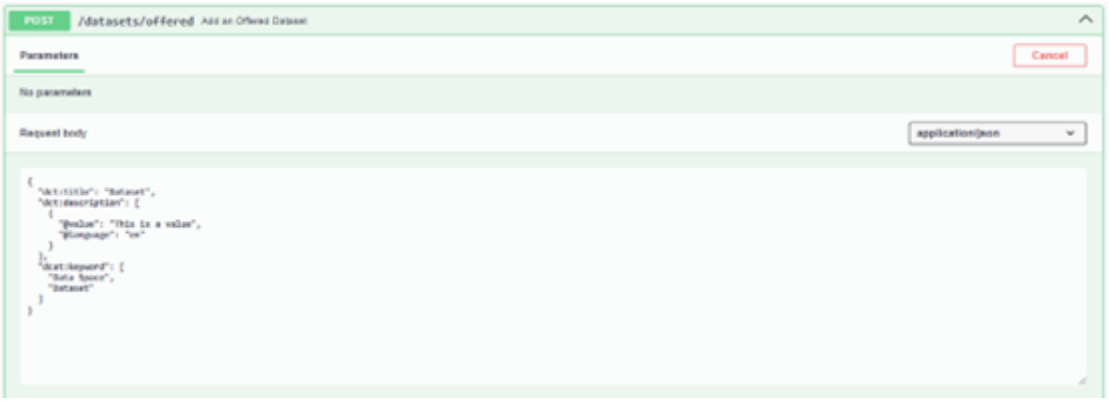


Figura 11. TDC Swagger UI. Creación de un Dataset

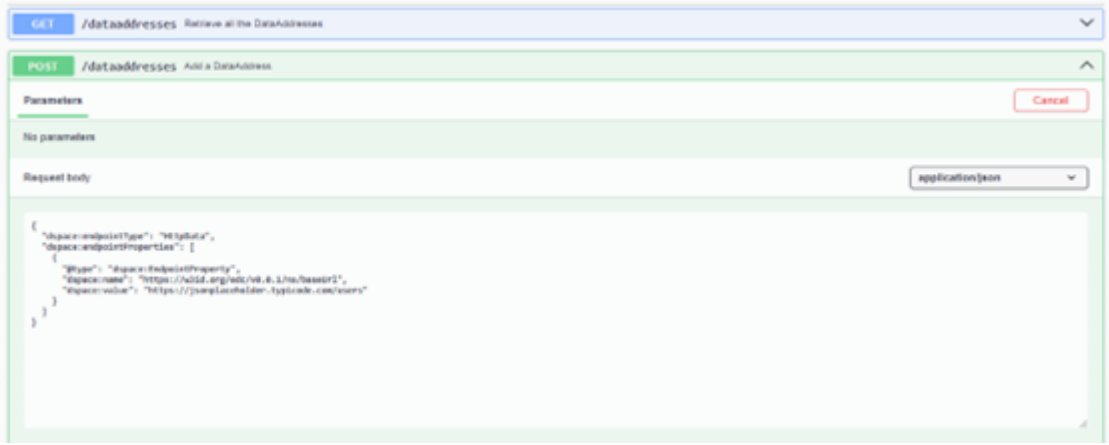


Figura 12. TDC Swagger UI. Creación de un DataAddress

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA
MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL
SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

Figura 13. TDC Swagger UI. Enlace de un Dataset a un DataAddress

En segundo lugar, el componente Metadata Manager ofrece también las interfaces para interactuar con otros conectores, es decir, para poder consultar catálogos de datos a través de un espacio de datos indicando la URL del conector objetivo, negociar las políticas para un conjunto de datos solicitado y acceder a un conjunto de datos cuyas políticas han sido acordadas. Estos procesos se realizan a través del componente Dataspace Protocol que se verá más adelante.

En tercer lugar, como consecuencia de los procesos de negociación y transferencia, son varios los metadatos que se incluyen en los conjuntos de datos y que, ofrecidos por el componente Metadata Manager, pueden ser monitorizados por el administrador del TDC. Estos son los procesos de negociación, los acuerdos resultantes y los procesos de transferencia para cada uno de los acuerdos.

En este sentido, el modelo utilizado por el TDC para la descripción de los conjuntos de datos es extendido con los conceptos de ContractNegotiation, Agreement y TransferProcess tal y como se ve en la siguiente figura:

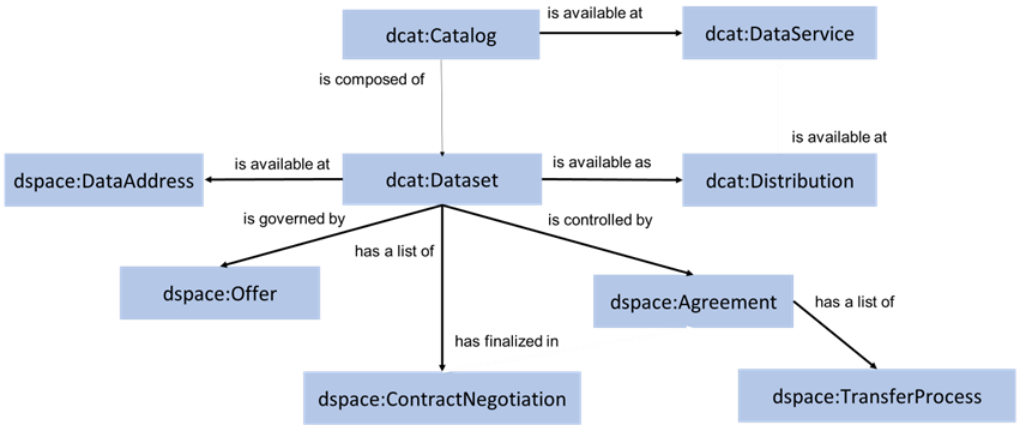


Figura 14. TDC. Extensión de DCAT para la descripción de ContractNegotiation, Agreement y TransferProcess

Del mismo modo que ocurre con la descripción de los conjuntos de datos en sí, el componente Metadata Manager permite la creación, lectura, actualización y borrado de cada uno de estos

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA
MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL
SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

conceptos, así como los enlaces entre ellos ofreciendo expresividad y flexibilidad. Hay que mencionar que mientras que la creación y actualización de estos conceptos es resultante de la operación del TDC, la lectura sirve para su monitorización y el borrado no está permitido.

Finalmente remarcar que, pese a que estas interfaces del componente Metadata Manager fácilmente gestionables desde el Swagger UI están orientadas a los desarrolladores para, por ejemplo, automatizar la configuración de conjuntos de datos, el TDC consta de una interfaz de usuario tal y como se muestra en la siguiente figura, que permite gestionar los conjuntos de datos de forma más amigable tal y como se muestra a continuación.

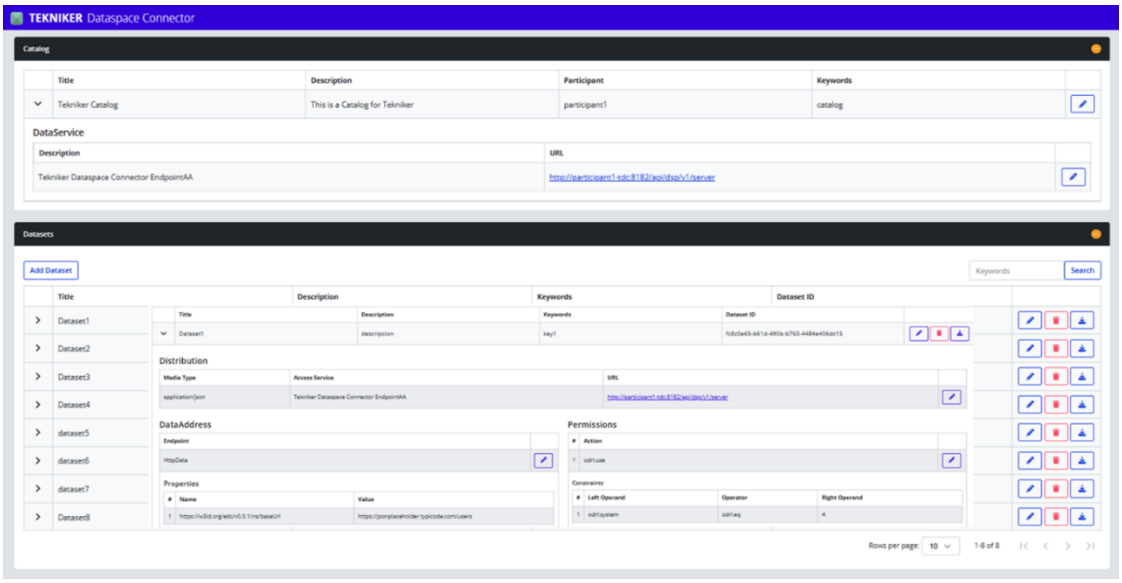


Figura 15. TDC UI. Creación, lectura, actualización y visualización de un Catálogo

6.3.2 Base de datos PostgreSQL

La base de datos PostgreSQL contiene un esquema SQL que le permite al TDC mantener de forma estructurada todos los metadatos necesarios para su correcta operación. Estos metadatos van desde los aquellos relativos a los conjuntos de datos descritos previamente hasta metadatos de configuración del propio TDC.

6.3.3 Dataspace Protocol

El componente Dataspace Protocol ofrece las interfaces que, siguiendo el Dataspace Protocol 2025 permiten la implementación del cliente y el servidor necesarios para la descripción de un catálogo de conjuntos de datos, la negociación del uso de los datos y el acceso a estos a través del Catalog Protocol, Contract Negotiation Protocol y Transfer Process Protocol de forma estandarizada.

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

En el caso del Contract Negotiation Protocol y el Transfer Process Protocol, el componente Dataspace Protocol implementa las interfaces y esquemas para la adopción de los diagramas de estados de negociación y transferencia correspondientes. Actualmente, mientras que el proceso de negociación se limita a los estados que permiten un proceso de negociación simple donde se aceptan las políticas definidas por el propietario de los datos, el proceso de transferencia no contempla la suspensión de la transferencia de datos.

6.3.4 Data Plane

El componente Data Plane ofrece las interfaces para la transferencia de los datos siguiendo los enfoques pull y push una vez esta es solicitada por el usuario de los datos a través del componente Metadata Manager vía el componente Dataspace Protocol mediante la implementación Transfer Process Protocol.

Por un lado, en la transferencia de datos vía push, durante el Transfer Process Protocol, el TDC en el lado del propietario de los datos genera y envía al TDC en el lado del usuario de los datos un token que, a través de una interfaz del componente Data Plane en el lado del propietario de los datos, le permitirá al usuario de los datos acceder de forma controlada y temporal a los datos. Para acceder a los datos, el componente Data Plane en el lado del usuario de los datos ofrece una interfaz que a partir del identificador del conjunto de datos de forma automatizada le permita parametrizar el acceso a la fuente de datos obteniendo el token temporal a través del Metadata Manager. De esta forma, tal y como se muestra a continuación, los datos se transfieren entre los Data Planes de los correspondientes conectores pudiendo validar la correctitud de los datos y el cumplimiento de las políticas en ambos lados. Es importante remarcar que este proceso es transparente para el usuario de los datos limitando su solicitud a su correspondiente Data Plane indicando el conjunto de datos a consumir.

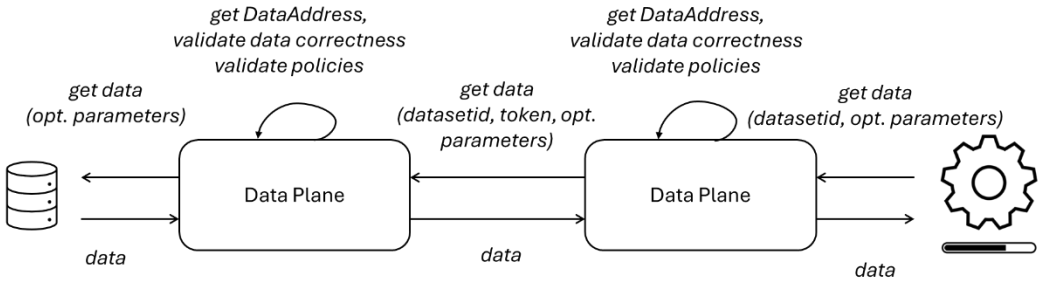


Figura 16. TDC. Data Plane Pull Aproach

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

Por otro lado, en la transferencia de datos vía pull, el usuario de los datos solicita la transferencia de los datos a un endpoint determinado a través del componente Metadata Manager vía el componente Dataspace Protocol. En este caso, el componente Data Plane en el lado del propietario de los datos ofrece una interfaz para, dado un conjunto de datos, obtener los datos y transferirlos al endpoint establecido.

6.3.5 Policy Engine

El Policy Engine, interceptado el intercambio de datos por el Policy Enforcement Point (PEP) del Data Plane, se encarga de realizar el enforcement de las políticas, por un lado, a través del Policy Decision Point (PDP) tomando decisiones sobre el permiso y prohibición del uso de los datos y, por otro lado, a través del Policy eXecution Point (PXP) ejecutando acciones derivadas del permiso sobre el uso de los datos.

Para realizar el enforcement de las políticas, almacenadas como acuerdos en la base de datos PostgreSQL, estas son solicitadas a través del Metadata Manager en función del conjunto de datos solicitado.

7. Recursos adicionales

7.1. Data Spaces Blueprint (DSSC)

Delivery Plan - DSSC Delivery Plan - Summary of assets publication - DSSC Delivery Plan - Data Spaces Support Centre

Starter Kit - Starter Kit for Data Space Designers | Version 1.0 | March 2023 - Starter Kit - Data Spaces Support Centre (dssc.eu)

Conceptual Model - Conceptual Model of Data Spaces - Blueprint v1.0 - Data Spaces Support Centre (dssc.eu)

Blueprint - Data Spaces Blueprint v2.0 - Home - Blueprint v2.0 - Data Spaces Support Centre

- Introducción - Introduction - Key Concepts of Data Spaces - Blueprint v2.0 - Data Spaces Support Centre
- Glosario de términos - Glossary - Blueprint v2.0 - Data Spaces Support Centre

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

- Componentes - Building Block Overview - Blueprint v2.0 - Data Spaces Support Centre
- Método de cocreación - Co-Creation Method - Blueprint v2.0 - Data Spaces Support Centre

7.2. Recursos de Gaia-X España

Listado de entidades, proyectos y recursos de Gaia-X España - www.gaiax-spain.com/listado-de-entidades-proyectos-demostradores-casos-de-uso/

7.3. Bibliotecas de componentes y conectores

DSSC toolbox - toolbox.dssc.eu

IDSA Data Connector Report - Data Connector Report - International Data Spaces

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

Figura 17. Grupos e instancias de conectores (fuente: IDSA)

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA
MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL
SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL

Créditos y agradecimientos

Esta guía se ha creado con las aportaciones voluntarias de los miembros del Grupo de Trabajo de Tecnologías Habilitadoras de Gaia-X España.

Autores

Enrique Areizaga Sánchez	Tecnalia
Kiko Cisneros de los Rios	Utopia Consulting
Manuel Escuin Gil	Aire Networks
Helena Fernández López	Gradiant
Gonzalo Gil Inchaurza	Tekniker
Asier Gravina Alfonso	Gaia-X España
Juanjo Hierro	Seamware
Alejandro Martínez Molina	ITI
Mario Petruzella Malagón	Arsys
Marco Rodrigues Ribeiro	ITI
Francisca Rubio Berenguel	Gaia-X España
Valentín Sánchez Peláez	Tecnalia
Jesus María Santamaría Yugueros	Tecnalia
Rizkallah Touma	i2CAT
Eros Zaupa	i2CAT

Si quieres unirte al Grupo de Trabajo de Tecnologías Habilitadoras de Gaia-X España o ampliar información, escríbenos a: tecnologias@gaiax-spain.com

Ayuda 27.08.467I.488 financiada por:
GOBIERNO DE ESPAÑA MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL SECRETARÍA DE ESTADO DE DIGITALIZACIÓN E INTELIGENCIA ARTIFICIAL