ANJANA DATA

THE DATA GOVERNANCE SOLUTION FOR THE BIG DATA & MULTI-CLOUD ERA

Gobierno del Dato Proactivo y Preventivo

2021

Problemas para compartir datos

Lleva a

ESTO

- No se sabe dónde encontrar datos de calidad ni quiénes tienen el conocimiento de su localización
- No se han establecido políticas y procedimientos para la compartición de datos o estos no han sido comunicados
- Sólo se disponen de mecanismos y herramientas ineficientes y poco efectivos para compartir datos
- No se sabe qué reglas de privacidad o seguridad aplican sobre qué datos o qué regulación o normativas se tienen que cumplir

ESTO

- Los usuarios no usan datos de calidad porque no saben dónde buscar lo que necesitan o porque usan lo primero que encuentran
- Los usuarios no saben cómo compartir datos así que no los comparten (pérdida de valor para negocio) o lo hacen mal
- Los usuarios comparten datos enviando ficheros por e-mail o aplicaciones de mensajería (pensando en la era digital)
- Los usuarios no saben qué datos se pueden compartir a quién y para qué por lo que no los comparten o no cumplen la normativa

FALTA DE GOBIERNO DEL DATO

Diferentes enfoques del Gobierno del Dato

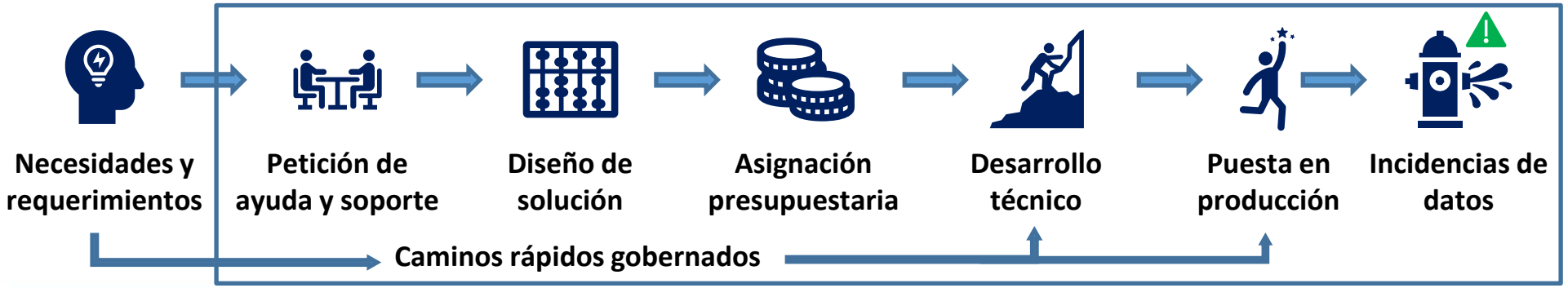
GOBIERNO DEL DATO PASIVO Y REACTIVO

- ✗ No se promueve la Cultura del Dato en la organización y los intervinientes
- ✗ Visión de Proyecto = Silos de Datos
- ✗ No hay sinergias = Ineficiencias
- ✗ No hay democratización/compartición de datos
- ✗ Errores por impactos en producción
- ✗ Incidencias de Regulación y Riesgos
- ✗ Incidencias de Calidad y Seguridad

GOBIERNO DEL DATO AL FINAL DE LA CADENA DE VALOR

GOBIERNO DEL DATO PROACTIVO Y PREVENTIVO

GOBIERNO DEL DATO AL PRINCIPIO DE LA CADENA DE VALOR Y PRESENTE A LO LARGO DE LA MISMA



- ✓ Aplicar el gobierno de los datos desde el principio lleva a tener datos de calidad
- ✓ Sinergias y reutilización = Reducción de costes
- ✓ Integración de la gestión de la demanda
- ✓ Promoción de la democratización y compartición de datos
- ✓ Gestión de impactos antes de producción
- ✓ Promoción del autoservicio gobernado
- ✓ Automatización de procesos técnicos

Ecosistema de Datos Moderno - Arquitectura

Desde una visión del gobierno de los datos, su compartición debería conseguirse gracias al establecimiento de políticas y procedimientos a la vez que se asegura su cumplimiento y custodia

Tenemos que saber quién consume qué DATOS y para qué

Queremos consumir muchos DATOS para muchos propósitos

Necesidades de **compartir datos** deben ser satisfechas de forma **ágil**

Cumplimiento de **políticas y procedimientos** de acceso a datos

Custodia, seguridad y privacidad de datos correctamente gestionada

Data Governance

- Governance model configuration
- New Generation Business Glossary
- Workflow Tools
- Centralized metadata and lineage management
- Active governance and dynamic audit
- Data portal with shopping cart
- Data Governance Dashboard
- Data contracts for the use of information

Data producers

Data Services for Ingestion

aws, Azure, CLOUDERA, kafka, nifi, cdata, ETL ELT, Events Streaming, API

Identity & Access Management (Data Privacy & Security)

LDAP, Active Directory, Sentry, Apache Ranger, SecuPi

Data Platforms (Storage, Modelling, Preparation, Processing, Publication & Sharing)

aws, Azure, hadoop HDFS, CLOUDERA, snowflake, Java JDBC, HIVE

Data Services for Consumption

aws, denodo, Azure, Data Tools, CLOUDERA, Java JDBC, kafka, nifi, API, cdata

Data consumers

ACTIVOS DE DATOS

- Almacenados en varios **repositorios**
- Soportados por **tecnologías** diferentes
- Disponibles en múltiples **formatos**

Ecosistema de Datos Moderno - Gobierno

Necesidades de **compartir datos** deben ser satisfechas de forma **ágil**

Cumplimiento de **políticas y procedimientos** de acceso a datos

Custodia, seguridad y privacidad de datos correctamente gestionada

Gobierno del Dato PROACTIVO Y PREVENTIVO

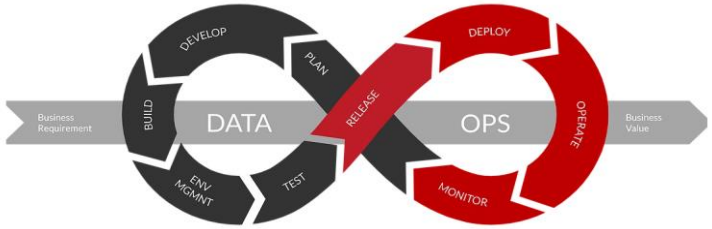
- Gobierno del Dato al principio de la **cadena de valor del dato**
- Enfoque **colaborativo** empoderando a los distintos intervinientes

- Integración con la **gestión de la demanda**
- **Integración** con otras tecnologías y piezas
- **Automatización** de procesos técnicos comunes
- Enfoque incremental e iterativo por **casos de uso**
- **Democratización** y autoservicio gobernado
- **Monitorización** para la mejora continua

- **Repositorio de metadatos** como eje central del ecosistema de datos
- Abstracción de la **gestión de datos** de tecnologías y plataformas

Automatización de procesos

- **EXTRACCIÓN DE METADATOS**
- **SAMPLEO DE DATOS**
- **GESTIÓN DE ACCESO A DATOS**
- **GESTIÓN DE ESTRUCTURAS**
- **AUDITORÍA EXTERNA DE PROCESOS**
- **OTRAS INTEGRACIONES**



DATAOPS NECESITA GOBIERNO DEL DATO PROACTIVO Y PREVENTIVO CON AUTOMATIZACIÓN DE PROCESOS

Gobierno activo

Integración sobre Plataformas de Datos para automatización de procesos técnicos

Gobierno activo es la capacidad de Anjana Data para automatizar procesos técnicos comunes de gestión de datos necesarios para compartir datos sobre Plataformas de Datos gracias a sus capacidades de integración

DATA ACCESS MANAGEMENT

- Ability to automate the granting of privileges after the corresponding approval workflow
- It is done at **object level** (dataset, process/instance path) and **user level** (nominal or service)
- It implies on the one hand the connection with identity management (user) and on the other hand with privileges management (user profile, linked group, ...)
- In addition, Anjana will need to be connected to the corresponding LDAP/DA or identity management system for roles and users management

DATA STRUCTURES MANAGEMENT

- Ability to automate the creation, modification and deletion of data structures and points of storage (i.e. tables, views or paths)
- It is more thought for **NoSQL, Big Data and Cloud environments** where there is a culture of data exploitation with dynamic storage
- It is necessary to know how the metadata and the schemas are managed in the system and which actions would have to be launched from Anjana to manage the structures or storage systems

Gobierno activo: Gestión de acceso a datos

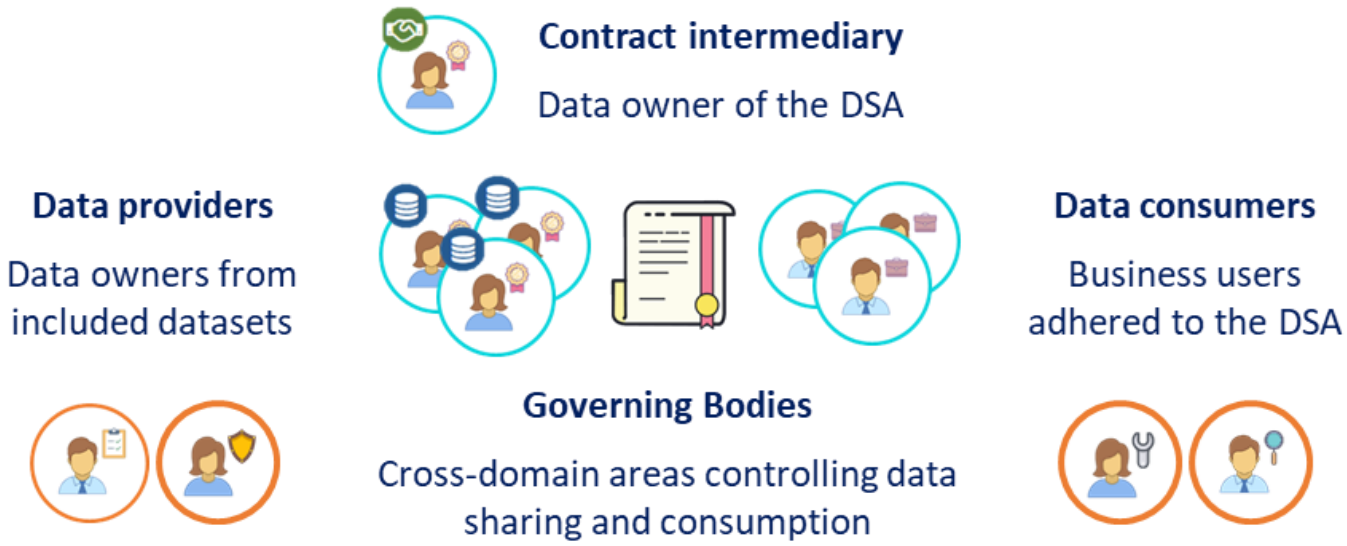
Portal de Datos y Marketplace. Operativizando los DSAs

Data Sharing Agreements

- Agrupaciones **lógicas**, flexibles y dinámicas de **Datasets**
- Tienen sus propios atributos de **metadatos** flexibles y configurables
- Pueden ser **relacionados** con cualquier tipo de entidad del Glosario de Negocio
- Entidad clave del **Marketplace** para la gestión de permisos de acceso a datos

Gobierno activo: Gestión de acceso a datos

Gobierno de DSAs: Políticas, Procedimientos y Contratos de Datos



Establecen un modelo común y estándar para la **gestión del acceso a los datos** mediante la implementación de **Contratos de Datos multiparticipados** entre productores y consumidores

Como todo **activo de datos**, tienen que tener sus propias **políticas** y **procedimientos** dentro del modelo de gobierno del dato definido y según el **dominio de datos** al que pertenezcan

Pueden tener varios **intervinientes**, definirse distintas **categorías**, identificar varios niveles de **sensibilidad** y **confidencialidad**, soportan **versionado** y tienen fecha de **expiración**

Gobierno activo: Gestión de acceso a datos

Categorías de DSAs

ROADMAP V5.0

CATEGORÍA I

- Únicamente otorgan **permisos de lectura** a los Datasets incluidos en el DSA
- Los usuarios consumidores se adhieren a estos DSAs para poder consumir los datos de los Datasets incluidos en el DSA con su **usuario nominal**
- **Ejemplo: usuario consumidor de varias tablas**

CATEGORÍA II

- Únicamente otorgan **permisos de lectura** a los Datasets incluidos en el DSA
- Los usuarios consumidores se adhieren a estos DSAs para poder consumir los datos de los Datasets incluidos en el DSA con un **usuario de servicio**
- **Ejemplo: Informe, aplicación, caso digital, …**

CATEGORÍA III

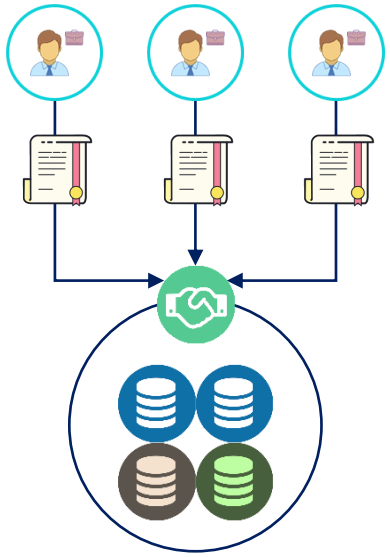
- Pueden otorgar tanto **permisos de lectura como de escritura** a los Datasets incluidos en el DSA
- Los usuarios consumidores se adhieren a estos DSAs para poder adquirir los permisos sobre los Datasets incluidos en el DSA con su **usuario nominal**
- **Ejemplo: DataLab/SandBox con usuario nominal**

CATEGORÍA IV

- Pueden otorgar tanto **permisos de lectura como de escritura** a los Datasets incluidos en el DSA
- Los usuarios consumidores se adhieren a estos DSAs para poder adquirir los permisos sobre los Datasets incluidos en el DSA con un **usuario de servicio**
- **Ejemplo: DataLab/SandBox con usuario de servicio**

Gobierno activo: Gestión de acceso a datos

Integración con Sistemas de Gestión de Identidades y Acceso a Datos



- Identificación unívoca de los **productores** y **consumidores** de los datos
- **Políticas de acceso a datos** para casos de uso específicos (cifrado, encriptado, enmascaramiento, anonimización, filtrado, …)
- **Requerimientos para los productores** para el correcto consumo de la información contenida en los Datasets (Calidad, Disponibilidad, SLAs, …)
- **Términos de licencias y condiciones de uso para los consumidores** (costes asociados, derechos y responsabilidades cedidos, cláusulas adicionales, …)

DSA Físico

- Un grupo en un Sistema de Gestión de Identidades (p.e. LDAP, DA, IAM) que contiene usuarios (nominales o máquina)
- Un rol en en un Sistema de Privacidad y Seguridad (p.e. Sentry, Ranger, DA, IAM) que contiene privilegios (CRUD, RWX, políticas) sobre Datasets (tablas, ficheros, …)

Automatización de procesos técnicos

Integración con los Sistemas y Tecnologías de Gestión de Identidades

- Cada DSA es un grupo en el SGI
- Cada vez que un usuario se adhiere a un DSA, será incluido en el grupo del SGI

Integración con los Sistemas y Tecnologías de Seguridad y Privacidad

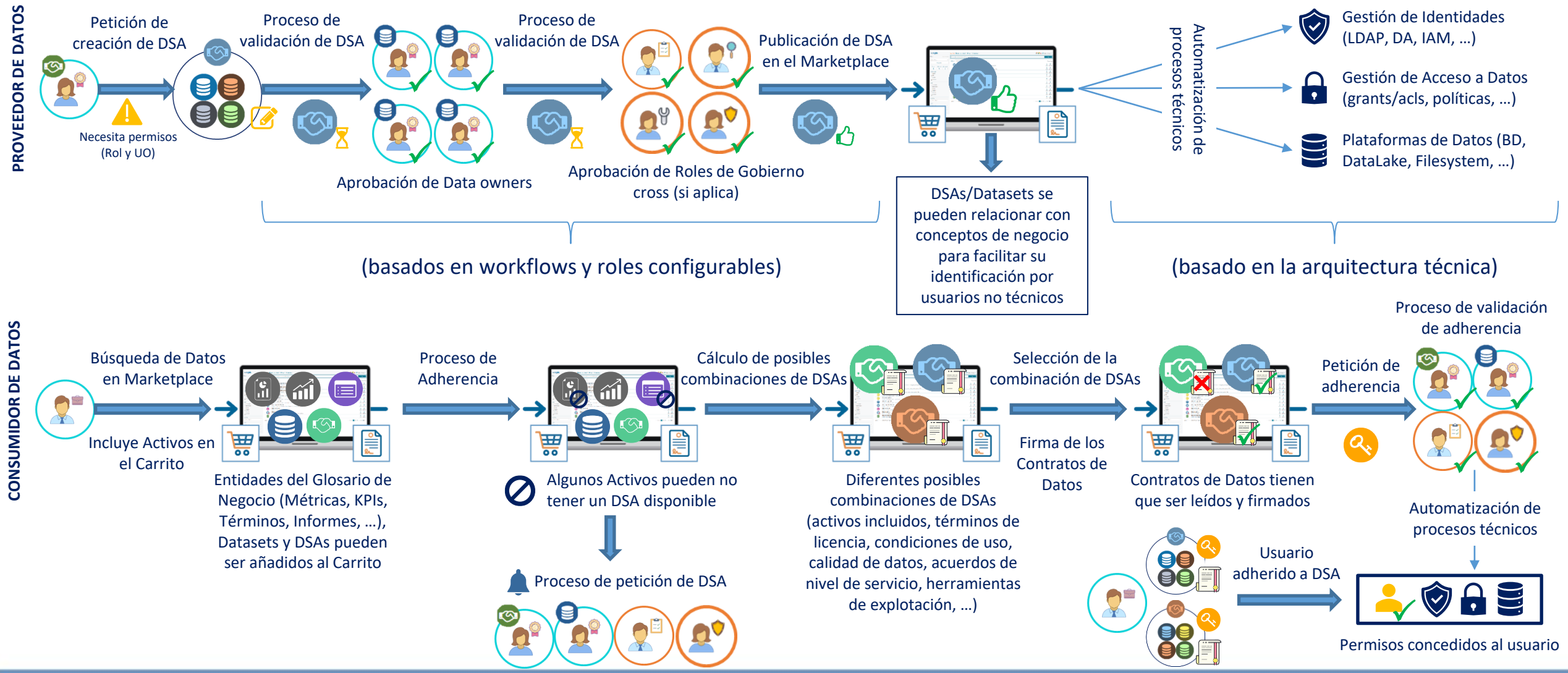
- Cada DSA puede ser un rol/grupo/perfil en el SGS correspondiente
- Cada rol/grupo/perfil tendrá los grants/acls/políticas correspondientes sobre los datos

Integración con las Plataformas de Datos y los Servicios de Explotación

- Cada DSA definirá las políticas de acceso a datos correspondientes para los Datasets
- Cada vez que un Dataset es incluido en un DSA, las políticas definidas serán aplicadas

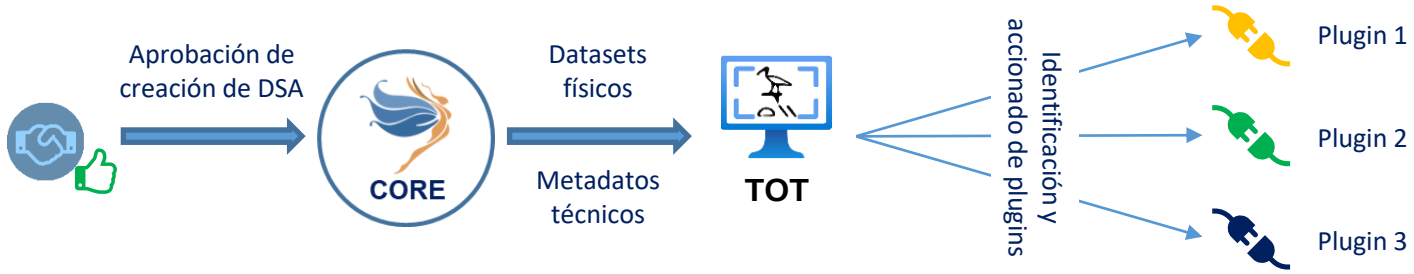
Gobierno activo: Gestión de acceso a datos

Integración con Sistemas de Gestión de Identidades y Acceso a Datos



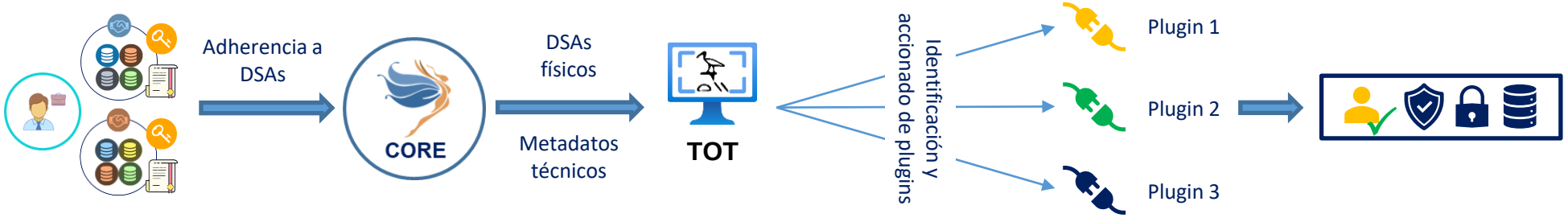
Gobierno activo: Gestión de acceso a datos

Integración con Sistemas de Gestión de Identidades y Acceso a Datos



- ✓ **Cada plugin actúa sobre una partición lógica resolviendo una conexión**
- ✓ **Cada plugin puede lanzar varias acciones según las tecnologías implicadas**
- ✓ **Una tecnología puede tener más de un plugin según sus posibles usos**

- Gestión de Identidades (LDAP, DA, IAM, …)
- Gestión de Acceso a Datos (grants/acls, políticas, …)
- Plataformas de Datos (BD, DataLake, Filesystem, …)



- ✓ **Los usuarios tendrán permisos de acceso a los datos físicos donde estén almacenados o sean servidos**
- ✓ **Los datos no son accedidos ni servidos a través de Anjana Data**
- ✓ **La complejidad de las políticas que pueden ser aplicadas dependen de las tecnologías subyacentes**

REQUERIMIENTOS DE SEGURIDAD

Según las acciones que tienen que ser ejecutadas por los plugins, la conectividad punto-a-punto requerida tiene que ser garantizada (IPs, Puertos, firewall, …) y el usuario máquina necesita permisos y credenciales acordes sobre las tecnologías correspondientes

CAPACIDADES DE EXTENSIBILIDAD

Los plugins disponibles pueden ser configurados y/o modificados para propósitos específicos o implementaciones ad-hoc a la vez que nuevos plugins pueden ser desarrollados y desplegados siguiendo los estándares recomendados por Anjana Data

Gobierno activo: Gestión de acceso a datos

Integración con Sistemas de Gestión de Identidades y Acceso a Datos

ROADMAP V5.0

La relación Dataset-DSA se establece en el proceso de creación/edición de DSA

En la relación se indican los permisos que aplican para ese Dataset por estar incluido en ese DSA

- (LDA
- Gestión de (grants/acls, po ...)
- Plataformas de Datos (BD, DataLake, Filesystem, ...)

La relación Usuario-DSA se establece en el proceso de adherencia a DSA

En la relación se indica el usuario (nominal o de servicio) que tiene que adquirir los permisos

RELACIONES DE PERMISOS

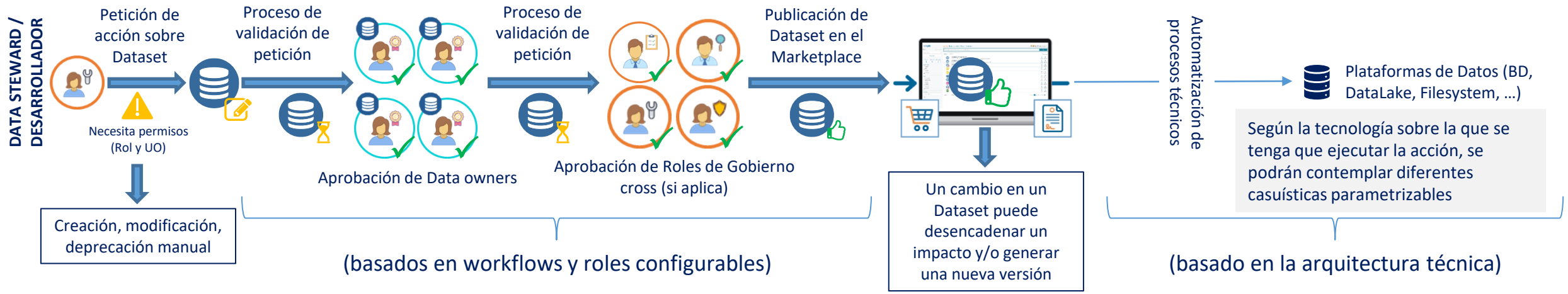
- Se pueden configurar atributos específicos que se soliciten al usuario cuando relaciona Dataset y DSA como por ejemplo una entidad del BG, un fichero adjunto, un campo de texto o una fecha.
- La complejidad de los permisos que pueden ser aplicados (columna/fila, masking, ...) dependen de las tecnologías subyacentes.

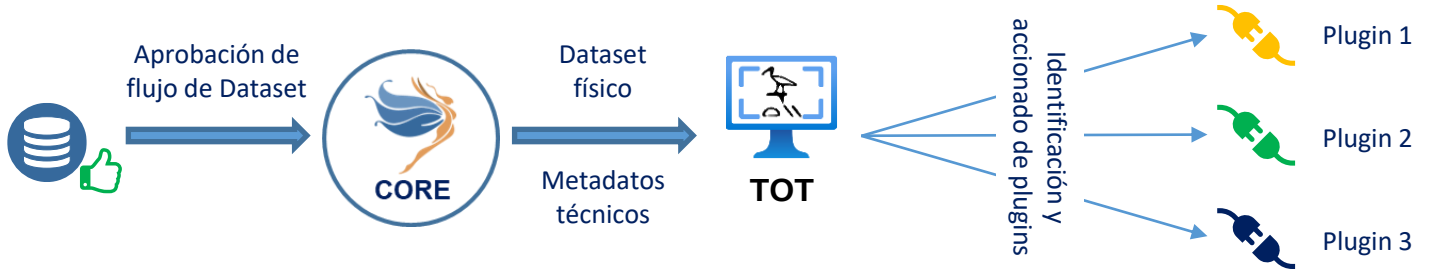
RELACIONES DE ADHERENCIA

- Se pueden configurar atributos específicos que se soliciten al usuario cuando solicita adherencia a un DSA como por ejemplo una entidad del BG, un fichero adjunto, un campo de texto o una fecha.
- Los usuarios adheridos a los DSAs (nominales o de servicio) tienen que existir previamente en el gestor de identidades.

Gobierno activo: Gestión de estructuras de datos

Integración con Plataformas de Datos





- ✓ **Cada plugin actúa sobre una partición lógica resolviendo una conexión**
- ✓ **Cada plugin puede lanzar varias acciones según las tecnologías implicadas**
- ✓ **Una tecnología puede tener más de un plugin según sus posibles usos**

Plataformas de Datos (BD, DataLake, Filesystem, …)

REQUERIMIENTOS DE SEGURIDAD

Según las acciones que tienen que ser ejecutadas por los plugins, la conectividad punto-a-punto requerida tiene que ser garantizada (IPs, Puertos, firewall, …) y el usuario máquina necesita permisos y credenciales acordes sobre las tecnologías correspondientes

CAPACIDADES DE EXTENSIBILIDAD

Los plugins disponibles pueden ser configurados y/o modificados para propósitos específicos o implementaciones ad-hoc a la vez que nuevos plugins pueden ser desarrollados y desplegados siguiendo los estándares recomendados por Anjana Data

Auditoría externa de procesos de datos

Monitorización de logs de auditoría de Plataformas y Tecnologías de Datos

- Plataformas de Datos (BD, DataLake, Filesystem, …)
- Tecnologías de Datos (ETL, Procesamiento, BI, …)
- Repositorios de logs de auditoría

Agente 1

Agente 2

Agente 3

HEIMDAL

Activos físicos gobernados

Logs de auditoría

CORE

Información incorporada como Auditoría

ACTIVOS FÍSICOS GOBERNADOS

- DATASETS
- INSTANCIAS

- Procesos ejecutados sobre los Datasets en batch / real-time
- Consultas o acciones lanzadas por los usuarios sobre los Datasets
- Modificaciones sobre los Datasets
- Eventos generados por código

- ✓ **Cada agente actúa sobre una conexión según la configuración establecida**
- ✓ **Cada agente puede capturar diferentes acciones según las tecnologías implicadas**
- ✓ **Una tecnología puede tener más de un agente según sus posibles usos**
- ✓ **Un agente puede capturar eventos ad-hoc generados por código si así se diseña**

- Los logs capturados que no se corresponden con ningún activo físico gobernado pueden ser analizados para detectar inconsistencias
- Se pueden generar alertas en base a la interpretación de los logs capturados e incorporados como Auditoría
- Se puede realizar detección de inconsistencias gracias al análisis de los logs capturados y los metadatos declarativos (p.e. procesos cuya ejecución no se corresponde con lo definido, accesos indebidos por un usuario sin adherencia a ningún DSA que contenga el Dataset, …)

REQUERIMIENTOS DE SEGURIDAD

La conectividad punto-a-punto requerida tiene que ser garantizada (IPs, Puertos, firewall, …) para que los agentes puedan capturar los logs de auditoría y el usuario máquina necesita permisos y credenciales acordes sobre las tecnologías correspondientes

CAPACIDADES DE EXTENSIBILIDAD

Los agentes disponibles pueden ser configurados y/o modificados para propósitos específicos o implementaciones ad-hoc a la vez que nuevos agentes pueden ser desarrollados y desplegados siguiendo los estándares recomendados por Anjana Data