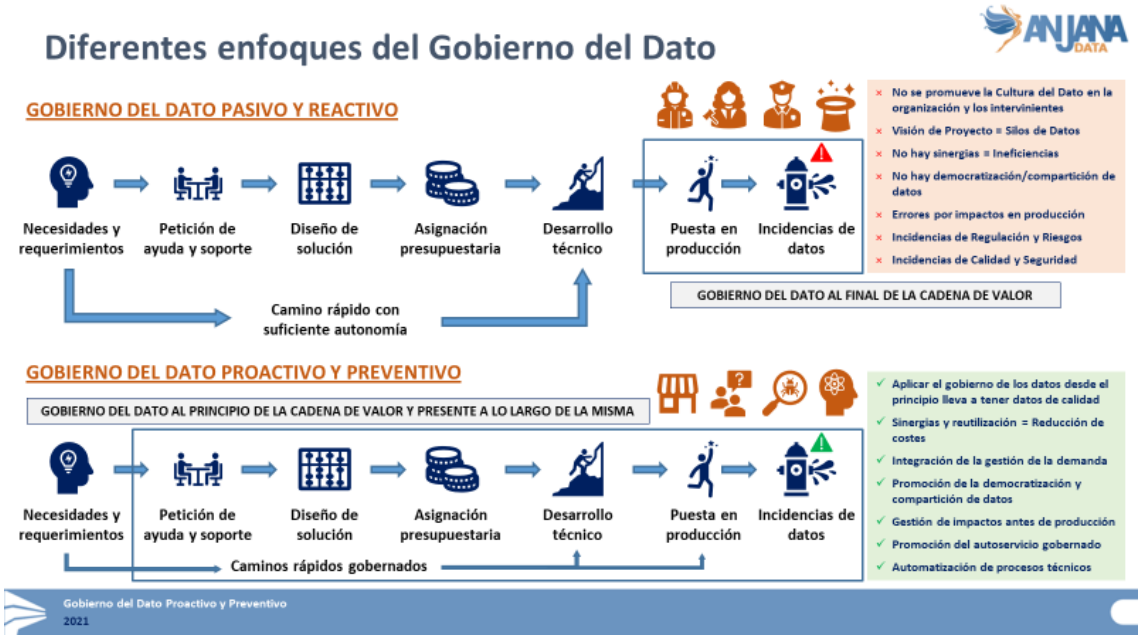
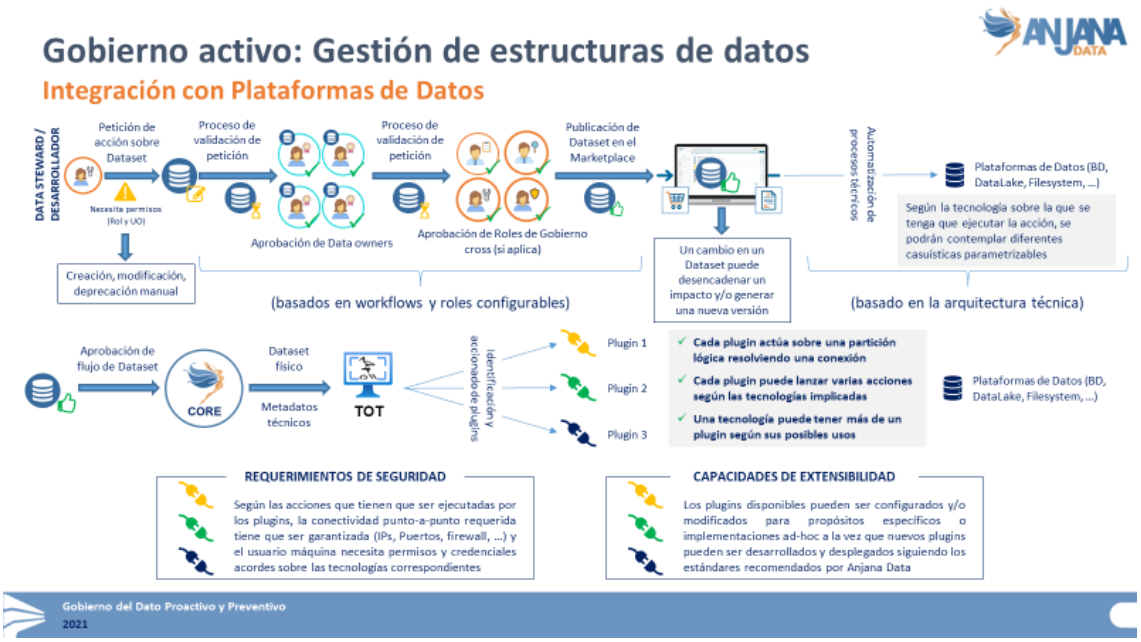
Actualización automática de metadatos en Anjana Data

Para incorporar todos los activos de datos y sus metadatos, Anjana Data dispone de diferentes mecanismos de carga de metadatos según el origen de los mismos: los activos de datos y sus metadatos pueden ser declarados de forma manual a través de la interfaz, importados vía plantillas de Excel generadas de forma dinámica, incorporados mediante el uso de las APIs disponibles y a través de una serie de plugins de integración con otras tecnologías para el descubrimiento y extracción automática de metadatos desde las fuentes. Para esta última, Anjana Data cuenta con una cobertura de tecnologías y un roadmap de evolución de plugins para dar una mayor cobertura, así como de partners que ofrecen conectores específicos.

Adicionalmente, Anjana Data proporciona al cliente un kit de desarrollo para que el cliente se construya sus propios plugins si así lo desea. Además, las APIs de Anjana Data permiten la catalogación de nuevos activos, así como la incorporación de nuevos metadatos a activos ya catalogados, de tal forma que siempre existe una alternativa para la carga automática de metadatos desde cualquier origen.

Los plugins de Anjana Data están conceptualizados para que sea el usuario el que lance el proceso de descubrimiento e importación bajo demanda desde la UI aunque también se pueden accionar a través de API. Dicho proceso de descubrimiento e importación soporta de forma nativa la creación de nuevos activos pero no la actualización de activos ya existentes, es por eso que Anjana Data aboga por una visión de gobierno del dato proactivo y preventivo con capacidades para la automatización de procesos técnicos sobre las tecnologías de datos, de tal forma que se permite la catalogación inicial de los activos bajo demanda (gobierno pasivo) y después se migra a un modelo de gobierno activo en el que primero se declaren los cambios en la solución de gobierno del dato (Anjana Data) y una vez aprobados se ejecuten las acciones necesarias sobre las plataformas de datos gracias a las capacidades de gobierno activo de Anjana Data.





De todas formas, Anjana Data también soporta el gobierno del dato pasivo en el que la creación de nuevos activos de datos y la ingesta de metadatos en Anjana Data también se puede automatizar y orquestar en procesos batch y/o real-time, de tal forma que se pueden actualizar los metadatos de los activos de datos en base a la detección de cambios en las estructuras de datos. En este caso, se puede desarrollar un script que invoque a la API bien en modo batch programado o bien a partir de un trigger accionado por el origen o por un agente que monitorice cambios y después ejecutar la lógica de actualización de los objetos en Anjana Data haciendo el uso correcto de las APIs.

En futuras versiones de la solución se incluye la evolución de las capacidades de los plugins de tal forma que estos mismos ofrecerán la posibilidad de ser configurados para funcionar tanto con gobierno activo (accionados en Anjana Data) como con gobierno pasivo (accionados en origen) de forma nativa sin necesidad de hacer uso de las APIs ni construir la lógica por fuera. Adicionalmente, se podrá configurar sobre los propios plugins el modo de funcionamiento permitiendo incluso el escaneo de las fuentes bajo demanda, en batch y en real-time en base a reglas lógicas complejas para seleccionar aquellas zonas donde se desean descubrir nuevos activos o cambios en los actuales de forma automática. Por último, también se podrán configurar desde los plugins las acciones a realizar para los activos de esas zonas (p.e. sólo generar alertas, crear nuevos activos y/o actualizar los metadatos de activos existentes directamente quedando en estado aprobado, lanzar workflows de validación, ...).

Todo esto se complementa siempre con las capacidades nativas de catalogación de las tecnologías de datos. Por ejemplo, en el caso de tecnologías nativas de cloud se pueden aprovechar las capacidades de los catálogos nativos de los proveedores (AWS Glue Catalog, Azure Purview, GCP Data Catalog, ...) ya que Anjana Data dispone de integraciones nativas con estas tecnologías las cuales van siendo evolucionadas para ofrecer más capacidades según evolucionan las capacidades de los catálogos nativos de los Cloud.

Por último, la definición de un framework y una metodología de desarrollo estándar que incluya el uso de las APIs de Anjana Data para la catalogación de nuevos activos y la notificación de cambios también es una buena práctica recomendada que muchas organizaciones están implementando, sobre todo en entornos Cloud.