UNIVERSITAT POLITÈCNICA DE VALÈNCIA

Escuela Técnica Superior de Ingeniería Informática

Análisis e implantación de una solución de Gobierno del Dato con la plataforma de Informatica

Trabajo Fin de Grado

Grado en Ingeniería Informática

AUTOR/A: Martínez Castellar, Olga

Tutor/a: Ramos Peinado, Enrique

CURSO ACADÉMICO: 2021/2022

Resumen

Actualmente, las compañías están siendo cada vez más conscientes de la necesidad de conocimiento sobre la información que se posee, si dicha información es de calidad, su procedencia y su alineamiento con las políticas internas. Esto hace que el gobierno de datos se haya convertido en una preocupación esencial para cualquier compañía. Es por ello que, ante el aumento de la demanda de dichos servicios, el siguiente trabajo de final de grado, desarrollado en la empresa NTT Data, se centrará en, primeramente, introducir en qué consiste un proyecto de Gobierno del Dato, como se desarrolla y las partes que lo componen y, posteriormente, hacer una descripción de la solución global de gobierno, presentando las herramientas utilizadas, concretamente, las de la plataforma de Informatica: Informatica Axon, Informatica Entrerprise Data Catalog (EDC) e Informatica Data Quality (IDQ). Finalmente, se hará una implementación en base al caso de uso presentado y las reglas asociadas a este.

Palabras clave: Gobierno del Dato, Calidad de datos, Catálogo de Datos, metodología, implementación.

Abstract

Today, companies are becoming increasingly aware of the need for knowledge about what information is held, whether that information is of good quality, where it comes from and how it aligns with internal policies. As a result, data governance has become an essential concern for any company. That is why, given the increasing demand for such services, the following final degree work, developed in the company NTT Data, will focus on, firstly, introducing what a Data Governance project consists of, how it is developed and the parts that compose it and, subsequently, a description of the global governance solution, presenting the tools used, specifically, those of the Informatica platform: Informatica Axon, Informatica Enterprise Data Catalog (EDC) and Informatica Data Quality (IDQ). Finally, an implementation will be made based on the use case presented and the rules associated with it.

Keywords: Data Governance, Data Quality, Data Catalog, methodology, implementation.

Resum

Actualment, les companyies estan sent cada vegada més conscients de la necessitat de coneixement sobre la informació que es posseïx,si es de qualitat, la seua procedència i el seu alineament amb les polítiques internes. Açò fa que el govern de dades s'haja convertit en una preocupació esencial per a qualsevol companyia. És per això que, davant de l'augment de la demanda dels dits servicis, el següent treball de final de grau, desenrotllat en l'empresa NTT Data, se centrarà en, primerament, introduir en què consistix, com es desenrotlla i les parts de què consta un projecte de Govern de la Dada i posteriorment, fer una descripció de la solució global de govern, presentant les ferramentes utilitzades, concretament, les de la plataforma d'Informàtica: Informàtica Àxon, Informàtica Entrerprise Data Catalog (EDC) i Informàtica Data Quality (IDQ). Finalment es farà la implementació en base al cas d'ús presentat i de les regles asociades a aquest.

Paraules clau: Govern de la Dada, Qualitat de dades, Catàleg de Dades, metodologia, implementació.

Índice general

Índice de figuras

Índice de tablas

1 Introducción

En este primer capítulo se exponen la motivación que han llevado a la realización de este proyecto, los objetivos que se quieren conseguir y una visión de la estructura de la memoria, que se seguirá a lo largo de todo el proyecto.

1.1 Motivación

En el panorama tecnológico actual el proceso de digitalización resulta inevitable, muestra de ello es que la mayoría de las empresas almacenan sus datos de manera digital. La globalización ha aumentado en gran medida la cantidad de información que se necesita almacenar para ofrecer un servicio, y en concreto se estima que dicha digitalización se ha acelerado hasta cinco años a causa de la pandemia COVID-19. Todos estos hechos hacen que, ante tan ingente cantidad de datos, muchas empresas no pueden hacer frente a las actividades multifuncionales básicas de sus datos, pues sus técnicas de gestión tradicional no lo permiten, este hecho las obliga a recurrir a la integración de nuevas soluciones, como son las de Gobierno del Dato (Engler, 2020).

Las herramientas de gestión de Gobierno del Dato ofrecen, entre otros, la ayuda a la toma de decisiones, la transparencia en procesos, posibilidad de construcción de procesos estándares repetibles o la reducción de costes, todas estas características son de vital importancia para afrontar el cambio tecnológico.

Es por ello por lo que, en el presente trabajo de fin de grado se pretende en primer lugar, aclarar cuáles son los pasos que se deben seguir en un proyecto de Gobierno del Dato para conseguir la implementación y transformación de los datos, aprendiendo así a gobernarlos de manera adecuada, para aprovechar su máximo potencial y consiguiendo adaptar en cada momento su estrategia a la situación del mercado (Bonet, 2021) así como presentar las diferentes soluciones de gobierno de dato existentes en el marcado. Por otra parte, se desarrollará un caso de uso a partir de un término del glosario y ya en la última parte del proyecto se elaborará una batería de reglas para la automatización del descubrimiento de Dominios de Datos.

El uso de las herramientas de Informatica Axon, con las que se tratará en este trabajo están basadas íntegramente en el cloud ofrecen la posibilidad de acelerar el proceso de transformación digital, garantizando la integración de los datos y convirtiéndolos así en un activo fundamental.

1.1 Objetivos

El objetivo principal del siguiente trabajo de final de grado consiste en la elaboración no solo de un manual que explique de manera global los aspectos más relevantes en un proyecto de gobierno del dato y su abordaje, sino también la implementación de un caso de uso en que se ponga en práctica la aplicación de reglas de calidad para el descubrimiento automático de dominios de datos y, por tanto, pueda resultar de ayuda en cualquier proyecto de este tipo.

Por otro lado, para la aplicación de reglas de calidad se realizará con una demostración práctica mediante la implementación de un caso de uso.

De igual modo, existen también una serie de objetivos secundarios entre los que se encuentran:

→ Realizar diagnóstico y definir grado de madurez

→ Definición del modelo de negocio

→ Análisis de las herramientas de gobierno del dato existentes en el mercado

→ Conocer las herramientas usadas en la creación de reglas

→ Describir la muestra y el tipo de valores que la componen

→ Determinar las reglas de calidad a crear

→ Identificar las mejoras obtenidas a partir de la creación de reglas

→ Determinar las aplicaciones de los descubrimientos en casos reales

→ Determinar las perspectivas de futuro

1.2 Estructura de la memoria

A continuación, se hará una presentación de la metodología que se seguirá a lo largo de este trabajo. Para empezar, cabe destacar que está dividido en dos grandes partes. La primera, está compuesta por el aspecto más teórico del Gobierno del Dato y la segunda, en la que se realiza una implementación de los conceptos vistos anteriormente y una demostración a través del caso de uso planteado.

En primer lugar, nos encontramos con el capítulo 2, donde se hará una presentación del concepto Gobierno del Dato (apartado 2.1): qué es, para qué sirve, En qué consiste. Dentro de ese mismo capítulo (apartado 2.2): tenemos el diagnóstico, nuevamente explicaremos qué es, medición del grado de madurez.

Para finalizar este capítulo nos encontramos con (apartado 2.3): definición del modelo de negocio, qué es y pasos previos a seguir.

En segundo lugar, tenemos el capítulo 3, en él se tratará acerca del análisis de las herramientas, concretamente tenemos la presentación del estado del arte (apartado 3.1): qué es, en qué consiste, que herramientas del panorama tecnológico actual conforman el gobierno del dato. Seguimos con las herramientas de Gobierno del Dato (apartado 3.2): se centrará en la descripción de las herramientas utilizadas, (apartado 3.3): herramienta de Gobierno del Dato de Informatica Axon. Continuamos con la herramienta de calidad de datos (apartado 3.4): cómo se integra con las anteriores, completitud, validación e indicadores de calidad. Para finalizar este capítulo tenemos la herramienta de catálogo de datos (apartado 3.5): donde tenemos el acceso y escaneo de distintos tipos de fuentes, comparación de Informatica con otros softwares del mercado.

En tercer lugar, tenemos el capítulo 4, en el que se realizará la implementación, este está dividido en varios apartados entre los que se encuentran (apartado 4.1): presentación del caso de uso, qué es, en qué consiste, (apartado 4.2): descubrimiento de automático de los dominios de datos, cómo se realiza, qué condiciones debe cumplir, (apartado 4.3): creación de reglas, herramientas en las que se pueden crear, lenguaje en el que escriben y, por último dentro de este capítulo resolución (apartado 4.4): donde se presentaran las diferentes reglas creadas y los descubrimientos realizados, así como las mejoras que estos descubrimientos han aportado a la organización.

Para finalizar, tenemos el capítulo 5 donde se presentarán las conclusiones del trabajo y los próximos pasos propuestos. Además, al final del documento también se incluirá la bibliografía con todas las fuentes consultadas durante la realización del trabajo, así como un anexo donde se realizarán aclaraciones a aspectos que así lo requieran.

2 Marco conceptual

En este primer apartado, se tratarán los principales aspectos que encontramos a la hora de enfrentarnos a un proyecto de Gobierno del Dato, como abordarlo y cómo definir una estrategia o plan de acción que se adecue a las necesidades a satisfacer.

2.1 ¿En qué consiste un proyecto de Gobierno del Dato?

No podemos responder a este apartado sin antes dar una definición de Gobierno del Dato. Algunas de las que encontramos son:

"Un marco de toda la empresa para asignar derechos y deberes relacionados con las decisiones para poder manejar adecuadamente los datos como un activo de la empresa." (Otto, 2011).

"Se define como el ejercicio de autoridad y control (planificación, monitorización y aplicación) sobre la gestión de los activos de datos." DMBoK2.

Leyendo las definiciones anteriores podemos extraer que el objetivo principal de los proyectos de Gobierno del Dato es garantizar la usabilidad, disponibilidad, seguridad e integridad de los datos, para ello, aporta un conjunto de procedimientos definidos, así como un plan de aplicación de estos.

En el inicio de este tipo de proyectos se deben establecer cuatro principales pasos de acción:

1) Establecer el responsable de los diferentes datos: en este primer paso, además de establecer el responsable de los datos se tendrá que garantizar los siguientes atributos de estos: exactitud, accesibilidad, consistencia, integridad y actualización.

2) Como se van a almacenar los datos: en este apartado además de su almacenamiento también se tendrá que platear como van a ser protegidos frente a posibles ataques.

3) Establecer un conjunto de normas para determinar su uso: en este punto se definirán los procedimientos para determinar cómo los datos van a ser utilizados por los distintos usuarios autorizados.

4) Establecer un conjunto de controles y procedimientos de auditoría: para de esta forma hay que asegurar que se cumplen las normas de gobierno.

2.2 Diagnóstico – Grado de madurez

En este apartado se tratará la medición de madurez de los datos, esto es imprescindible pues conociendo su grado de madurez nos será más sencillo identificar sus necesidades y oportunidades de mejora.

El análisis de la madurez está directamente relacionado con el análisis estratégico, para ello, se puede utilizar una herramienta como Data Compass, la cual nos permite analizar y optimizar los datos, dicha herramienta basada en 5 principios facilita el análisis de la madurez.

- **Valor de negocio:** alinea los objetivos estratégicos de la organización y de los datos para tratar de aportar un valor diferencial.

- **Gobierno del dato:** permite valorar la estrategia más adecuada para la prestación de servicios proactiva, teniendo como objetivo la eficiencia en las operaciones y los procesos, agilizando la toma de decisiones y asegurando la mejora continua.

Dentro de este apartado lo podemos dividir en dos subapartados, los cuales son:

A) Modelo operativo: define los procedimientos y políticas del dato que ayudarán a la organización al cumplimiento de sus objetivos. A su vez, también asegurará la consistencia y disponibilidad de estos.

B) Modelo organizativo: define el modelo de organización a nivel de estructura, comités, roles y responsabilidades.

- **Cultura del dato:** se debe adoptar e interiorizar en toda la organización, es esencial impulsarla a nivel interno a partir de estrategias de sensibilización. Dentro de la cultura del dato encontramos, por una parte, la culturización, que consiste en el desarrollo de las acciones de comunicación que se centran en concienciar tanto de la importancia de los datos, como de su beneficio. Por otra parte, tenemos la gestión del talento, cuya función es adaptar el modelo y os procesos de gestión de personas a los nuevos perfiles existentes.

- **Analítica avanzada:** la finalidad es dar respuesta a ciertas necesidades mediante el uso de métodos y herramientas de alto nivel, enfocadas en la proyección de tendencias y comportamientos.

- **Arquitectura:** permite organizar, asegurar, administrar, almacenar y recuperar la información que, de soporte a la estrategia de los datos de la organización, garantizando el alineamiento entre la tecnología, el modelo operativo y las personas.

Figura 1. *Principios en el análisis del grado de madurez, elaboración propia.*

Una vez hemos analizado los 5 componentes principales de Data Compass los trabajadores de la organización reciben un cuestionario en el que se les pregunta:

- El área de la organización a la que pertenecen
- Valoración de la madurez de la dimensión Valor de Negocio
- Valoración del nivel de madurez de gobierno del dato
- Valoración del nivel madurez cultura del dato y personas
- Valoración del nivel de madurez de la analítica avanzada
- Valoración del nivel de madurez de la arquitectura

Una vez respondido, se analizan los datos obtenidos para determinar el nivel de madurez y establecer un modelo de gestión y una estrategia de acción para llegar hasta él.

2.3 Definición del modelo de negocio (pasos previos)

Los datos están muy presentes en el día a día de una organización, tienen un papel fundamental, es por ello, que las organizaciones orientadas al dato son aquellas que maximizan el valor de sus datos para conocer mejor a sus usuarios, convirtiéndose así en más competitivas.

La estrategia de los datos permite establecer las bases para convertir los datos en un activo estratégico que aporte valor al negocio alineando los objetivos

estratégicos de la organización y los datos, con tal de conseguir que aporten un valor diferencial.

El objetivo principal que se persigue con este punto es entender que tareas se deben realizar y en qué orden. En primer lugar, existen dos situaciones en las que se puede encontrar la organización

1) Gobierno reactivo: cuando existen entidades que no están gobernadas y se desea comenzar a gobernarlas, entendiendo el actual alcance del modelo d gobierno.

2) Gobierno activo: si ya existe gobierno del dato, pero se necesita crear nuevas entidades o realizar alguna modificación de las existentes.

El siguiente paso consiste en elaborar una lista con las tareas a realizar en orden y clasificarlas por categoría indicando el rol o roles encargados de llevarla a cabo.

A continuación, se realizarán las tareas de despliegue entre las que se incluyen:

1) **Definición del alcance**

    a. **Selección del alcance**

    El Data Governance Manager debe definir el alcance, es decir, establecer qué entidades hay que gobernar y el significado funcional de esos datos.

    b. **Analizar las áreas impactadas**

    El Centro de Excelencia debe analizar las áreas que se ven afectadas por el nuevo alcance y la involucración de estas en el gobierno.

2) **Asignación de roles**

    a. **Identificación y asignación de Data Owners**

    Se identifica la necesidad de cubrir un rol de Data Owner y se comunica a las áreas propietarias del dato, las cuales identifican a las personas que desempeñarán dicho rol.

    b. **Identificación y asignación de Data Stewards**

    Se identifica la necesidad de cubrir un rol de Data Steward y se comunica a la dirección de sistemas, la cual identifica a la persona que desempeñará dicho rol.

3) **Comunicación**

    Se realiza la comunicación de bienvenida a la persona o personas a las que se les haya asignado el nuevo rol.

4) **Formación**
   Realizar formación en las nuevas responsabilidades en caso de que fuera necesaria.

Los tres pasos que detallaremos a continuación solo es necesario realizarlos cuando se trata de gobierno activo, puesto que, cuando se trata de gobierno reactivo se entiende que la arquitectura del dato ya está creada y, por tanto, no es necesario realizar las siguientes tareas.

5) **Definición y documentación de requerimientos funcionales**
   Los data owners reúnen y comprenden las necesidades tanto de los usuarios, como d ellos consumidores y las transforman en requerimientos funcionales para el data steward, el cual los documenta.

6) **Desarrollo técnico**
   El data steward realiza una propuesta técnica en base a los requerimientos definidos anteriormente.
   El mánager de IT debe revisar y validar la propuesta
   El data owner debe revisar la propuesta a nivel funcional.

7) **Clasificación en Tiers**
   Por último, los data owners se encargan de clasificar los datos a gobernar en tiers, los cuales se utilizan para asignar un nivel de exigencia a las funciones del gobierno del dato.

3 Análisis de las herramientas

En este tercer punto del trabajo, primeramente, se presentarán algunas de las soluciones disponibles en el mercado para, en los siguientes centrarse en explicar cada una de las herramientas de Informatica (Axon, EDC e IDQ).

Toda herramienta empleada como solución a un proyecto de Gobierno del Dato debe de cumplir una serie de requisitos como son:

1. Crear reunir y alinear las reglas
2. Resolver problemas
3. Monitorizar y hacer cumplir las reglas mientras se proporciona un apoyo constante a los interesados en los datos

3.1 Estado del arte

El crecimiento de la demanda de proyectos de gobierno del dato ha conllevado a un crecimiento en el número de softwares que se ofertan. Por ello, vamos a presentar algunas de las herramientas que podemos encontrar disponibles en el mercado.

1) OVALEDGE

En primer lugar, tenemos la herramienta OvalEdge la cual es un software diseñado para el gobierno del dato, que ayuda a las empresas a gestionar el análisis de datos, o el cumplimiento normativo, entre otros. Con el software se pueden supervisar las bases de datos, las herramientas de extracción, transformación y carga (ETL), las plataformas de inteligencia empresarial y los data lakes para crear una especie de inventario de los datos.

OvalEdge se caracteriza por ser una herramienta intuitiva y de sencillo uso, como podemos observar su interfaz es clara. Sin embargo, su parte de data quality no está muy desarrollada y se han observado problemas tanto respecto a errores visuales de interfaz como de programación.

Respecto a su precio, si bien es cierto que la herramienta dispone de una prueba gratuita su precio de implantación se inicia en los 100 US$/mes.

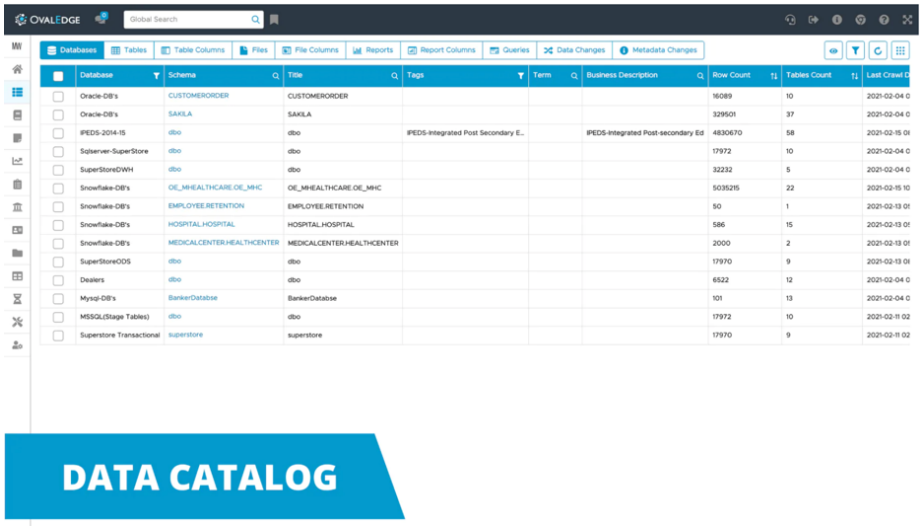

Figura 2. *Interfaz OvalEdge*

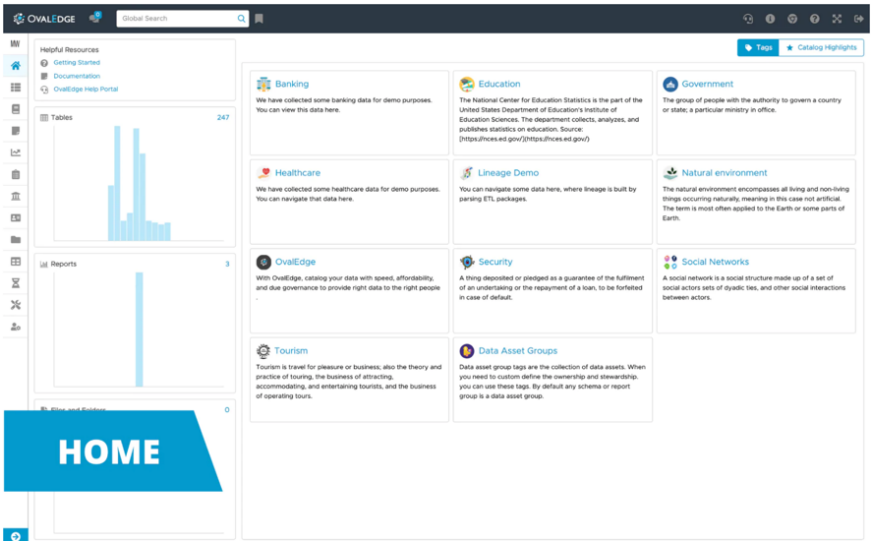

Figura 3. *Menú de inicio en interfaz OvalEdge*

2) XPLENTY

Se trata de una herramienta de software basada en la nube y que se utiliza para dirigir datos y obtener visualizaciones de ellos. Proporciona funcionalidades que permiten integrar datos ETL y ELT, preparar y procesar datos para su posterior análisis en la nube. De esta forma, su objetivo es ayudar a las empresas a la mejora de sus procesos de ventas, marketing, etc.

Al igual que la herramienta anteriormente descrita también ofrece una prueba gratuita de 7 días.

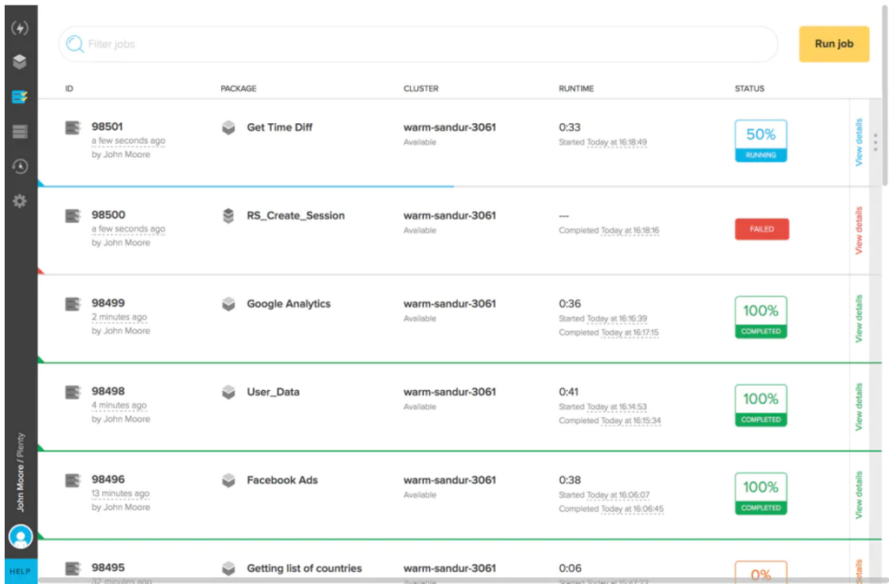

Figura 4. *Interfaz Xplenty*

3) TALEND

Software de código abierto basado en la integración de datos ETL. La ventaja de Talend es que, a diferencia de otras, ofrece todas sus funcionalidades en una única herramienta y, por tanto, al ofrecer solo un entorno también facilita la adaptación y el mayor conocimiento del usuario.

Esta potente herramienta, engloba tareas como la integración de aplicaciones, gobierno de datos, API o análisis e integración. Además, ayuda en la implementación de los procesos digitales y el análisis avanzado, contribuyendo así a facilitar la toma de decisiones. Todo ello conlleva a que sea una de las plataformas más utilizadas del mercado.

Dado que es un software de código abierto una de las ventajas es que se puede descargar y utilizar en su versión gratuita por un periodo de tiempo ilimitado y posteriormente decidir si se compra para su uso empresarial cuyo precio es alrededor de 1000$ / usuario / mes. Sin embargo, algunos de las desventajas que

presenta son su lentitud a la hora de trabajar con proyectos de gran tamaño, lo que suele ser habitual en este tipo de contextos.

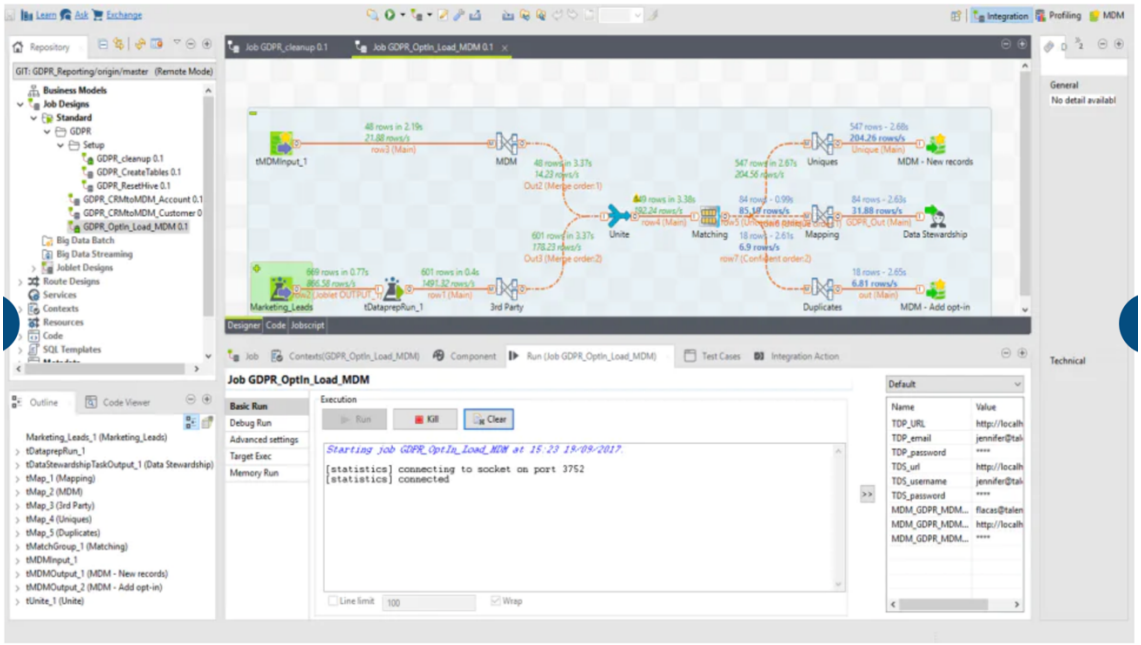

Figura 5. *Interfaz Talend, Talend 2022*

4) ANJANA

Es el único software dedicado al gobierno del dato proactivo y preventivo, esto quiere decir que la aplicación del gobierno del dato desde el primer momento conlleva a tener datos de calidad, de los cuales se pueden aprovechar las sinergias que surgen en los distintos niveles del proceso, desde la compartición de datos hasta la ayuda a la automatización de procesos técnicos.

Figura 6. *Modelo de Gobierno del Dato en Anjana, Anjanadata 2022*

Además, presenta un diseño UX/UI avanzado para facilitar el uso a los usuarios no técnicos. Entre otras funcionalidades que incluye podemos encontrar la definición de metadatos de referencia, la creación intuitiva de activos de datos, configuración de atributos mediante formularios dinámicos.

A diferencia de alguna de las herramientas con anterioridad Anjana no ofrece versión de prueba gratuita.

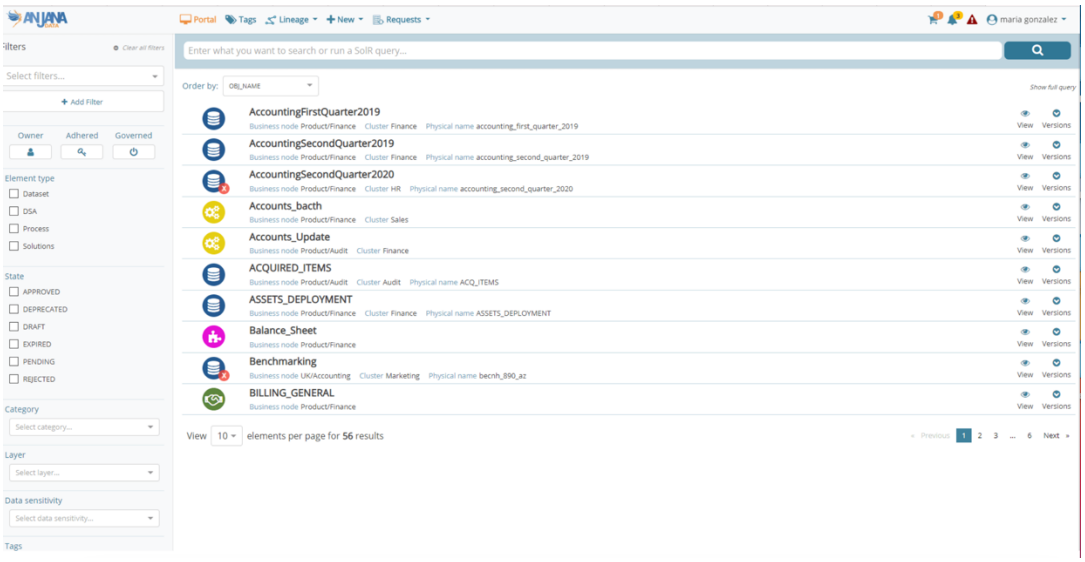

Figura 7. *Interfaz Anjana, Anjanadata 2022*

Como hemos mencionado anteriormente, una vez hecha esta introducción a las principales herramientas existentes en el mercado, en los siguientes apartados nos centraremos en las que componen la Suite de Informatica, pues es la herramienta que hemos utilizado para el desarrollo de nuestro proyecto.

3.2 Tecnologías utilizadas

Entre las tecnologías con las que hemos contado para el desarrollo del proyecto se encuentran, además de las herramientas de informática, las cuales serás descritas en el siguiente apartado, se han utilizado otras que se detallan a continuación:

En primer lugar, se han creado 6 máquinas, dos por cada entorno, una para PRE y una para PRO (2 para Axon, 2 para catálogo de datos y 2 para calidad de datos), estas han sido creadas en la plataforma cloud (en la nube) de Amazon Web Services.

Además, el sistema operativo empleado por los servidores es Red Hat Enterpirse Lniux release 8.5.

Por otra parte, la herramienta empleada para realizar la conexión al cliente ha sido MobaXterm, con ella se ha accedido a la maquina Linux en remoto. Dicho software

permite realizar la gestión y el mantenimiento de las herramientas (iniciar, reiniciar, ver logs, etc.).

Para subir los archivos que más adelante se detallarán (necesarios para realizar las pruebas de calidad) y realizar su procesamiento, se ha empleado la herramienta WinSCP[1]. Por último, comentar uqe también se han realizado escaneos de bases de datos Oracle, PL/SQL, SQL Server, entre otros.

3.3 Herramienta Gobierno del Dato

La herramienta propuesta por Informatica para el gobierno del dato, también conocida como Axon es una potente herramienta para describir cómo, porque y donde se utilizan los datos en la empresa, aportándole mucho más contexto a esta actividad. Axon facilita la colaboración entre diferentes departamentos, los cuales a veces son conscientes de la fuerte dependencia entre ellos o respecto a otros recursos.

Tiene como objetivo ayudar a las empresas con su transformación digital, para ello, cuenta con diversas funciones como, facilitar a la capa de negocio un glosario, plasmado en la herramienta mediante facets, ayuda a la creación de conocimiento, pues sincroniza los metadatos técnicos con términos de negocio para de esta forma obtener conocimiento de elementos críticos. Además, se asegura del cumplimiento de las diversas normativas que atañen a la gestión de los datos.

Una de las ventajas de utilizar la Suite de Informatica es que todas sus herramientas se encuentran completamente integradas, en este caso Axon con las dos herramientas de las que se hablará a continuación (Data Catalog y Data Quality). Esto permite la colaboración de tal modo que los usuarios de Axon pueden consultar en tiempo real la calidad de los datos, detectarlos y definir los elementos clave.

Una vez se ha hecho la introducción a la herramienta se van a enumerar de manera detallada las funciones principales con las que cuenta Axon, encontramos:

→ Recopilar el contexto que rodea a los datos

Para ello, en Axon la empresa es descrita como un conjunto de facets, con los que se crea rápidamente un inventario de todos los elementos básicos de la organización. La familiarización de los usarios con estos facets ayuda también a la compresión de los datos, a partir de lo cual se pueden empezar a entender, ver quien los utiliza, para quién son útiles, etc.

→ Conectar a la comunidad de gobierno de datos

[1] WinSCP es un cliente SFTP gráfico para Windows que emplea SSH

Con esto se pretende impulsar la colaboración, captar las conexiones y dependencias. Conectar la organización a todos los niveles, para de esta forma aprovechar las sinergias que ofrece MIX de conocimientos que aportan los distintos equipos. Esto ayuda a fomentar una visión global en la que todas las áreas de negocio están vinculadas.

➔ Medir y analizar los datos

Antes de utilizar la información extraída de los para la toma de decisiones, se ha de verificar que esta es de calidad, por tanto, integrándose con Data Quality se garantiza que los datos son correctos y fiables. Gracias a esta visión compartida del negocio se puede analizar el estado de la compañía de manera rápida, rentable y precisa. analizando también los proyectos para detectar oportunidades de colaboración, carencias o dependencias.

➔ Visualizar el esfuerzo

Se pueden observar las diferentes conexiones que existen entre los datos y como es el flujo de información, mediante el linaje de los datos. Mediante los cuadros de mando se puede visualizar las deficiencias, costes, la unión, etc. Además, se pueden visualizar aquellos procesos considerados importantes, como pueden ser los vinculados a algún tipo de normativa a seguir por la organización.

A continuación, se muestra una imagen de la interfaz de la herramienta, en la que se puede observar cómo está organizada.

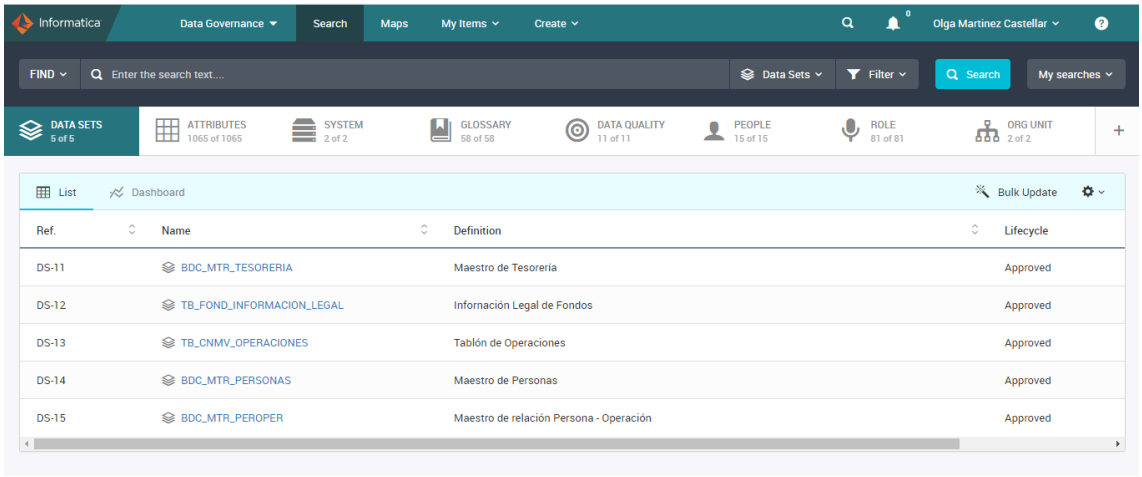

Figura 8. *Facet Data sets de la Interfaz Axon*

En Axon, existen diferentes facets que ayudan a conocer cuál es la estructura de la organización y de la información de la que dispone. Como se puede observar en la imagen anterior, figura 8. Además, cada uno de los apartados aporta un valor distinto a la organización, en el apartado de data quality se puede ver todas las reglas de calidad que se están aplicando, mientras que en glossary se pueden ver todos los términos y los dominios de datos de los que se dispone, por otra parte, están los roles que ocupan cada uno de los usuarios, entre otras funcionalidades. Estos se pueden observar en las imágenes que a continuación se presenta.

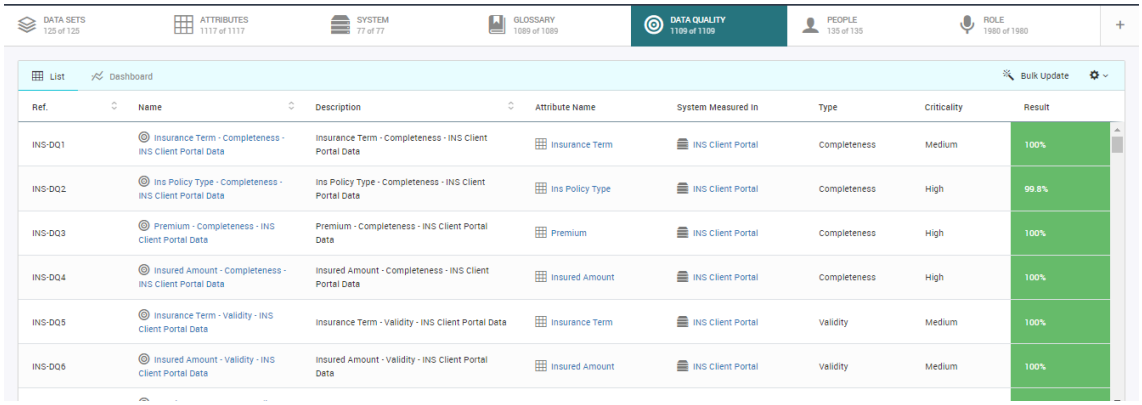

***Figura 9**. Reglas de calidad asociadas a los términos*

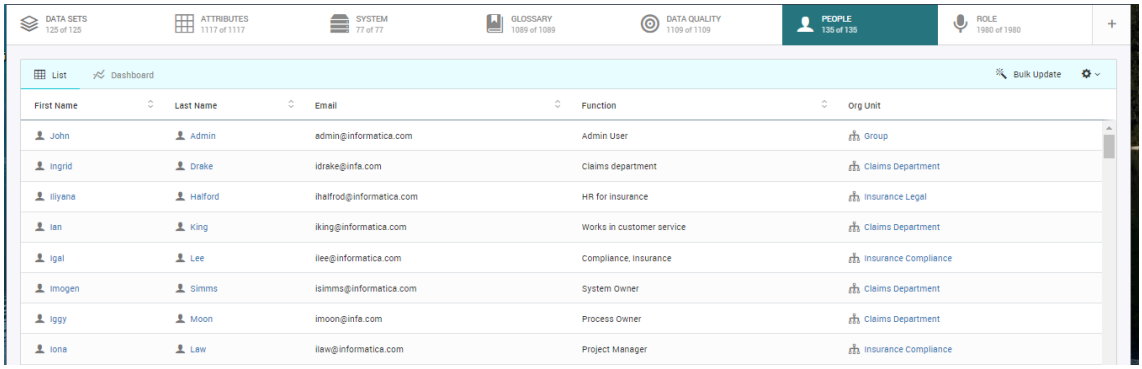

***Figura 10.** Personas y sus roles*

Además, dentro de un determinado término muestra las relaciones con otros términos, los grupos de interés que lo componen, el impacto, las reglas de calidad que se le están aplicando o los términos del catálogo de datos que lo componen. El cual si vamos a la columna “parent” y clicamos en la columna nos redirigirá a dicha herramienta.

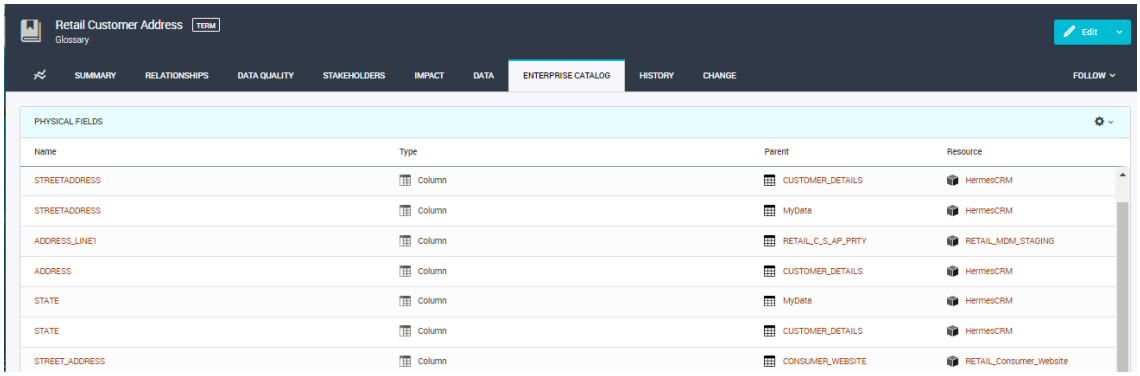

***Figura 11.** Relaciones de un término con la herramienta de catálogo*

Por último, hay que añadir que la potencia que tiene un wokflow es que permite llevar un control, es decir, un usuario puede realizar cambios según el tipo de permisos con los que cuente, en caso de querer realizar uno dependerá del Data Owner, al que se lo tendrá que comunicar y este a su vez se lo dirá al Data Steward que es quien realizará el cambio.

3.4 Herramienta Calidad de Datos

La herramienta de calidad de datos de Informatica, conocida como IDQ por sus siglas en inglés, ofrece multitud de funciones que resultan esenciales como son, entre otras, garantizar la entrega de información de calidad, evitando duplicidades y favoreciendo la consolidación de datos, otra funcionalidad sería el aumento de la productividad y eficiencia a través de la automatización de la información mediante el uso de inteligencia artificial o la generación y descubrimiento de reglas de calidad y de negocio que permitan actuar como aceleradores en los proyectos y ayuden a ahorrar tiempo y recursos. En esto concretamente nos centraremos en apartados posteriores del trabajo.

Por ello, entre las funcionalidades a las que se puede acceder a través de la herramienta de calidad de datos se encuentran:

→ Creación de los perfiles de datos

Este es un paso clave en los proyectos de gobierno del dato pues en él se forma la estructura de los datos y se define el plan de proyecto.

→ Estandarizar valores de datos

Se hace para eliminar los errores o inconsistencias presentes en los datos detectados en la ejecución de los perfiles.

→ Crear cuadros de mando

Utilizados para poder observar de manera gráfica las mediciones de calidad de un perfil.

→ Buscar registros duplicados

Se comparan dos conjuntos de registros con tal de encontrar los valores coincidentes de las columnas de datos. Para ello, se establece el nivel de similitud de una buena coincidencia entre los valores de los campos.

→ Crear y ejecutar reglas de calidad

Ejecutar reglas de calidad proporcionadas por Informatica o crear nuevas adaptadas a los objetivos propios de un proyecto dado.

A continuación, se muestran las funcionalidades de la herramienta de calidad de datos que se han descrito, en esta primera imagen vemos como de intuitiva resulta la creación de nuevos cuadros de mando (objeto de datos de tabla) o perfilados.

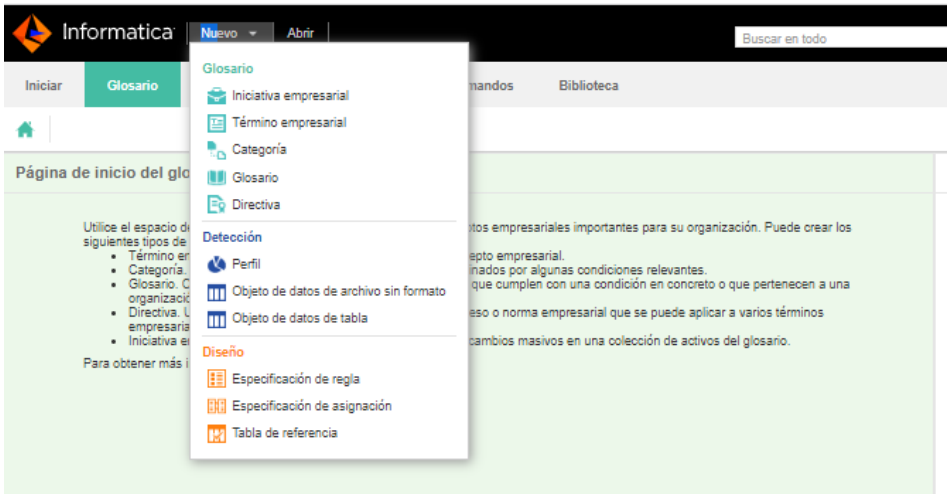

***Figura 12**. Nuevas creaciones en IDQ*

Mientras que en esta segunda imagen podemos ver como quedaría la lista de los perfilados una vez que estos han sido creados.

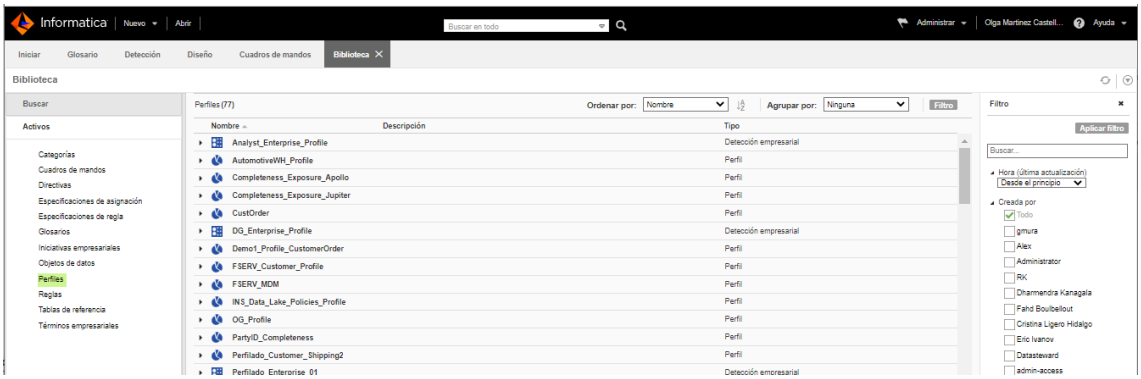

***Figura 13.** Ejemplo de perfiles creados en IDQ*

Por último, se puede observar la interfaz de diseño para la creación de reglas de calidad.

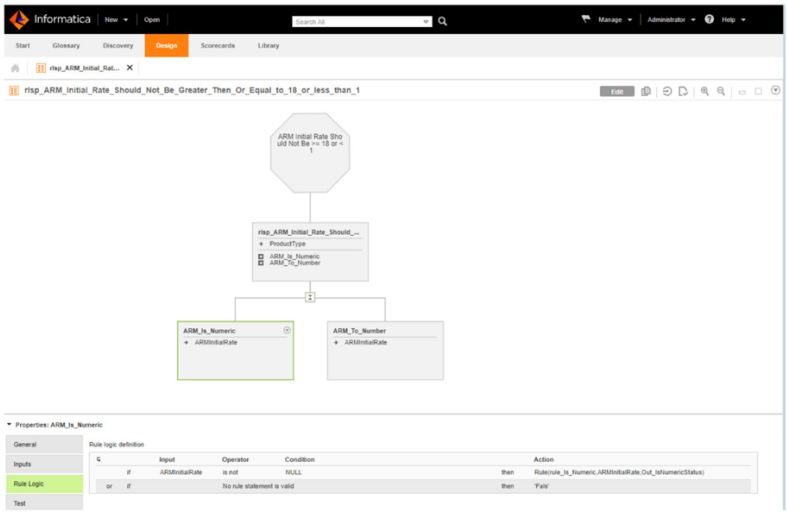

Figura 14. Interfaz de diseño en IDQ

Todas las funcionalidades que incluye esta herramienta, desde la facilidad de uso hasta la completa visión que aporta la han llevado a situarse como la más reconocida del mercado según el cuadrante mágico de Gartner. Dicho cuadrante posiciona herramientas tecnológicas que comparten características dentro de un mercado específico, para ello divide el cuadrante en líderes (funciona bien actualmente y está bien posicionado para el mañana); retadores (similar al líder pero sin una visión o plan acerca de cómo va a evolucionar el mercado); jugadores de nicho (se centra y es exitoso en un pequeño segmento, pero no innova fuera de ese nicho) y, por último, visionarios (entiende hacia donde se va a dirigir el mercado y tiene una visión definida, sin embargo, no ejecuta bien las ideas o lo hace de manera inconsistente) (Chien et al. 2021).

Los puntos fuertes que han llevado a Informatica a ocupar esta posición son la comprensión que ofrece de la calidad los datos y la capacidad de adaptación al mercado, conociendo su evolución y necesidades. Además, ofrece soluciones innovadoras con amplias carteras de productos y variados servicios de validación y estandarización de datos, es decir, ofrece los datos como servicio.

Figura 15. *Cuadrante para herramientas de calidad del dato, Gartner 2021*

3.5 Herramienta Catálogo de Datos

Como hemos visto anteriormente los datos pueden llegar a conllevar muchos beneficios para la organización y, por tanto, han de ser tratados como un activo más de la empresa. Por tanto, su gestión y transformación son completamente necesarios, y es por este motivo por el que Informatica dispone de una herramienta, conocida como Enterprise Data Catalog (EDC) que se encarga de analizar y catalogar los activos de toda la empresa.

Respecto a la gestión, uno de los pasos más complicados resulta hacer inventario de todos los datos de los que dispone la organización pues estos están repartidos entre os diferentes departamentos, entornos como el local y el cloud.

Por otra parte, respecto a la transformación EDC es un catálogo de datos que se basa en inteligencia artificial, por tanto, es un motor de búsqueda basado en el aprendizaje automático.

Se basa en el motor CLAIRE, el cual aprovecha los metadatos para proporcionar sugerencias o tareas de automatización. Esto ayudara a catalogar y analizar todos los datos de la empresa y, por tanto, hacer que la productividad de los usuarios de TI aumente a la vez que hace partícipe a los usuarios de negocio de la gestión, uso y situación de los datos.

Entre las funcionalidades que se incluyen dentro del catálogo de datos podemos encontrar:

→ Búsqueda semántica avanzada

Esto ayuda a buscar y detectar los conjuntos de datos de mayor interés para un determinado análisis, aplicando dominios de datos inferidos para que no queden datos sin detectar.

→ Linaje de datos

Permite visualizar el origen de los datos a cualquier nivel, incluyendo todos los detalles intermedios. La visualización del linaje se puede desglosar hasta mostrarlo detallado a nivel de columnas o métricas.

→ Asociación automática de términos del glosario

EDC permite importar de manera sencilla os términos del glosario de negocio desde Axon, esto permite añadir el contexto empresarial a los datos puramente técnicos, lo cual ayuda a facilitar la comprensión para ambas partes y, por tanto, contribuir a la colaboración.

→ Clasificaciones automatizadas

Clasificación automática de los dominios y entidades, esto ayuda a catalogar los datos, gobernarlos y extraer valor de ellos, así como a mejorar el filtrado en las búsquedas y las recomendaciones obtenidas.

→ Calidad de datos integrada

Esta funcionalidad permite ver las estadísticas de las reglas de calidad de datos, scorecards y grupos de métricas, de esta manera se conoce la calidad de os datos antes de que sean utilizados para algún estudio.

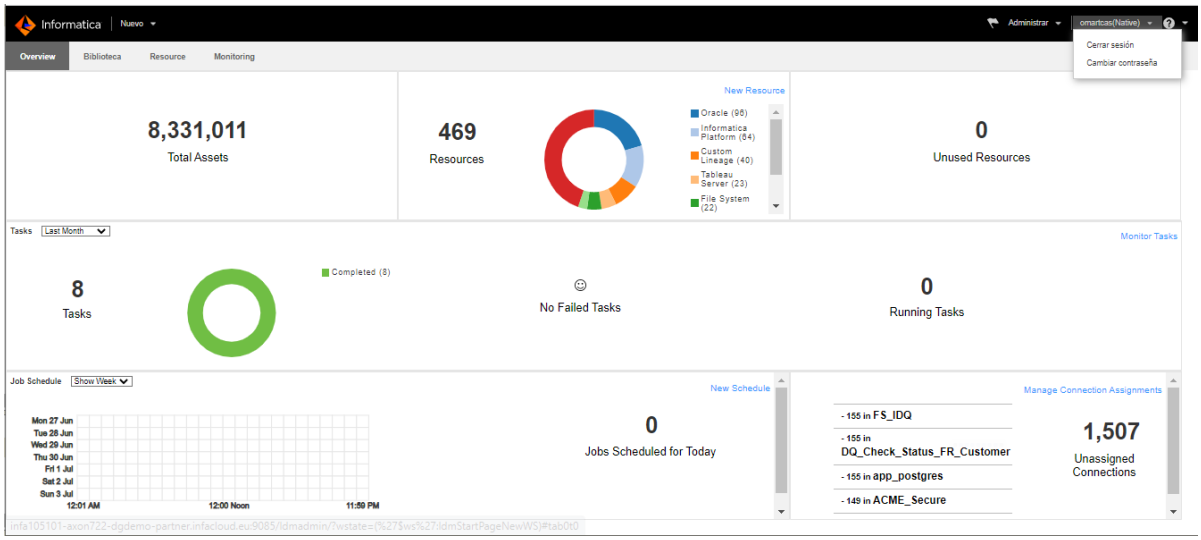

Figura 16. *Pantalla administrador de catálogo de datos*

En la siguiente imagen vemos algunos de los recursos de los que se dispone y sus tipos de datos.

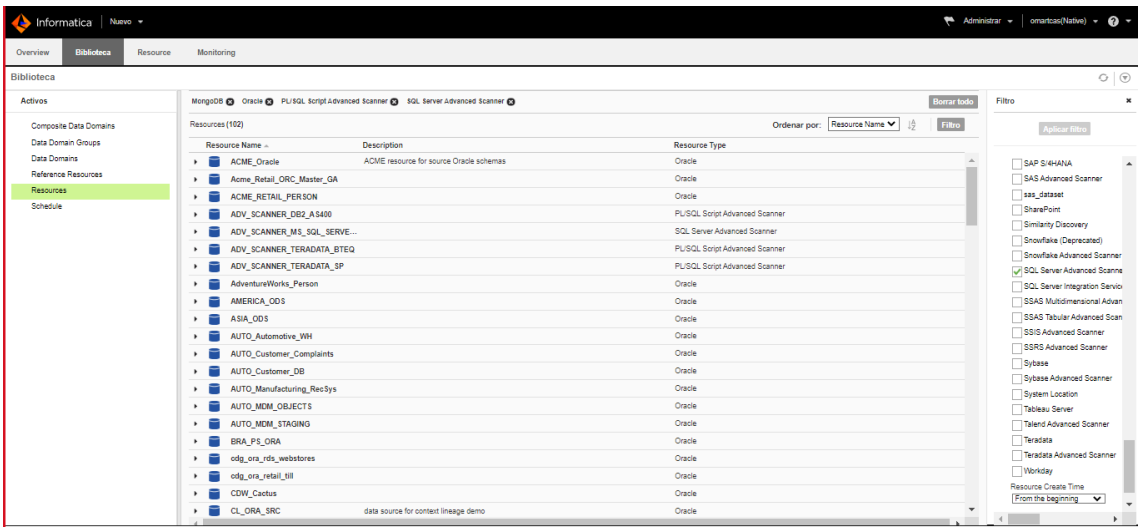

Figura 17. *Ejemplo de recursos creados en EDC*

Tomaremos como ejemplo el segundo recurso que se observa, *ACME_RETAIL_PERSON,* el cual es una base de datos Oracle, lo que implica que para acceder a ella hay que pasarle el nombre de servidor, el puerto, el usuario y contraseña que tengan permisos de lectura sobre los metadatos.

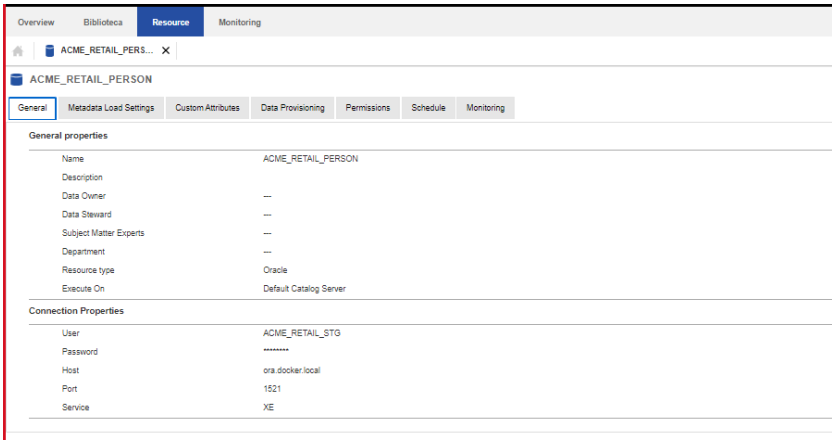

Figura 18. *Configuración de un recurso en EDC*

A la hora de crear un nuevo recurso existen diversas opciones, entre las que se encuentran, tipo de datos en la nube, No SQL, ingeniería de datos, integración de datos, glosario de términos de negocio, redes de transmisión, modelado de datos, gestión de archivos, inteligencia de negocio, manejador de hadoop[2].

[2] Estructura software de código abierto.

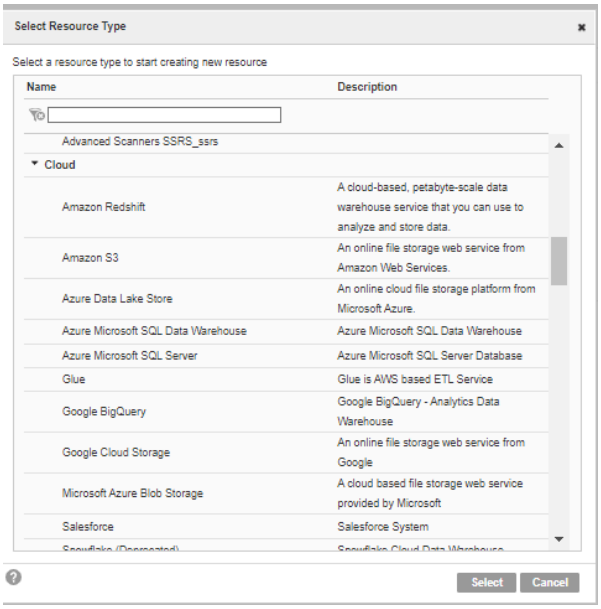

Figura 19. *Tipos de recursos en EDC*

Por último, antes de finalizar este capítulo, a modo de resumen se ha elaborado el siguiente cuadro incluyendo las funcionalidades principales de las tres herramientas que componen la Suite de Informatica.

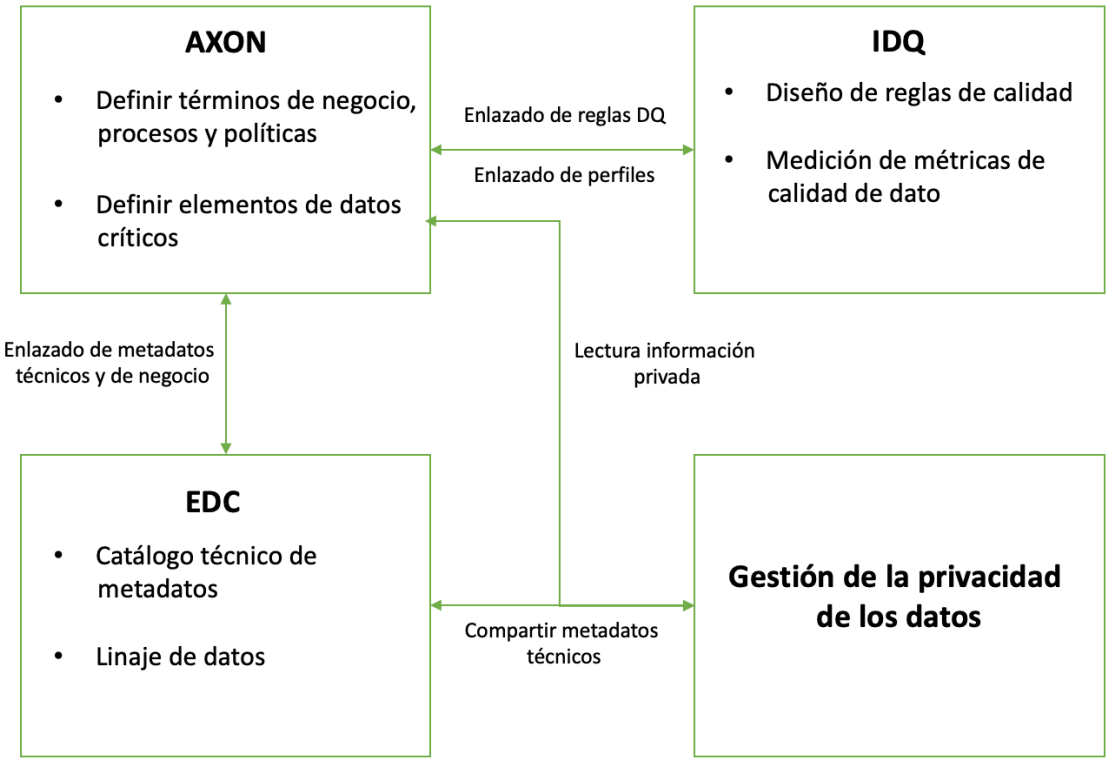

Figura 20. *Resumen herramientas Informatica, elaboración propia*

4 Implementación

4.1 Resolución de caso de uso

La finalidad principal de un caso de uso es explotar los datos con el objetivo de mejorar el rendimiento de los servicios y proporcionar así una visión de los potenciales casos de uso concretados al día a día de los usuarios. Es por ello, que este cuarto apartado del trabajo se van a presentar algunos casos de uso que se han detectado, entre ellos se encuentran el término *FS Customer ID* o los términos relativos a los clientes pertenecientes al dominio Retail.

En primer lugar, el término FS Customer ID se encuentra definido en el glosario de datos, en el cual existen distintos niveles. Primeramente, tenemos el dominio de servicios financieros, este engloba todos aquellos términos y definiciones que están relacionados con los servicios financieros. Dentro de los dominios podemos identificar las entidades, que son objetos del dominio. Aquí encontramos el dominio grupo de datos, el cual las partes son entidades que tienen una relación económica de algún tipo con la empresa con excepción de la laboral. A su vez, dentro de este encontramos el dominio de descriptores de partidos, que son elementos de datos que describen las partes. A continuación, dentro de dicha entidad pasamos a los datos personales de clientes de FS datos relacionados con personas que tienen tratos con la organización. En el siguiente paso ya nos situamos e la entidad de cliente de servicios financieros.

El término *FS Customer ID* se encuentra definido en el glosario y es el identificador relacionado con el cliente de servicios financieros. Se establece para realizar la configuración del cliente cuando este tiene que comerciar con la organización por primera vez.

A continuación, se muestra un árbol con los diferentes niveles que acabamos de mencionar.

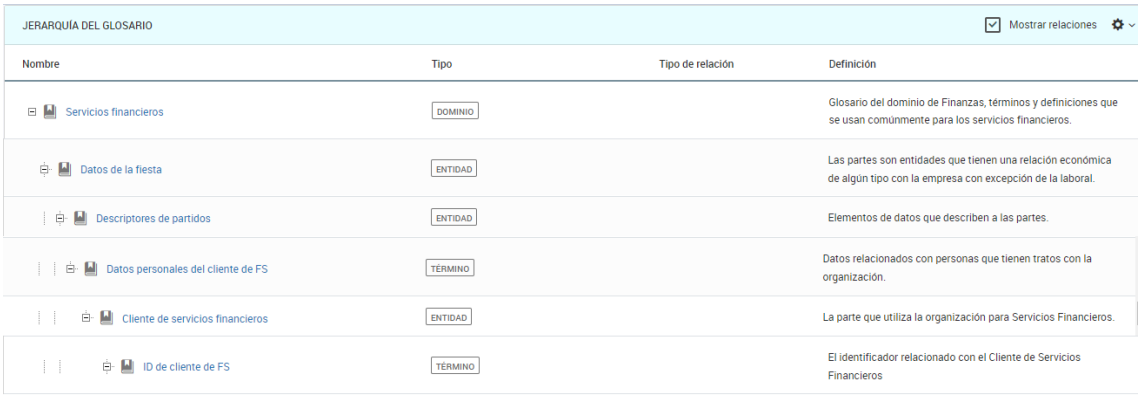

Figura 21. *Jerarquía de glosario del término CustomerID*

A continuación, pasamos a las personas de interés, en este caso al tratarse de un término relacionado con el dominio de servicios financieros las personas de

interés relacionadas directamente son aquellas que pertenecen a la unidad organizativa de servicios financieros – finanzas y entre los roles de dichas personas de interés tenemos al administrador del glosario, mayordomo Data Quality y al propietario de las definiciones de glosario.

PARTES INTERESADAS DIRECTAS

Role	Nombre	Unidad organizativa
Administrador del glosario	franco luna	Servicios financieros - Finanzas
Glosario Definición Propietario	Ley de Flynn	Servicios financieros - Finanzas
Mayordomo DQ	freya james	Servicios financieros - Finanzas
Administrador del glosario	freddy drake	Servicios financieros

Figura 22. *Grupos de interés término CustomerID*

Una vez hemos definido e ilustrado este primer término pasaremos a hacer lo mismo con los relacionados con el Retail que se ha comentado al inicio del apartado.

GLOSSARY HIERARCHY — Show Relationships

Name	Type	Relationship Type	Definition
Retail	DOMAIN		Glossary of retail terms and definitions used in everyday business between wholesalers and retailer.
Retail Customer	DOMAIN		The party that holds the Retail account online or shops in person in store.
Retail Customer Identifiers	ENTITY		Identifiers, such as a unique personal identifier (a defined term) and online identifier Internet Protocol address;
Retail Customer ID	ENTITY		The identifier relating to the Retail Customer
Generated Customer ID (NEW)	TERM		A customer ID that has been generated at point of transaciton as customer has no existing account / loyalty card.
Loyal Customer ID (VIP)	TERM		A specific type of VIP customer identification number
Retail Customer Phone Number	TERM		The current mobile phone number of the Retail Customer
Retail Customer Internet-related Information	ENTITY		Internet or other electronic network activity information, such as browsing history or interaction Ã¢ with an advertisement;

Figura 23. *Jerarquía de glosario del dominio Retail*

4.2 Descubrimiento automático de dominios de datos

En este apartado del trabajo realizaremos una batería con reglas para automatizar el descubrimiento de Data Domains. Por ello, en primer lugar, se va a realizar una explicación de qué es y en qué consiste un Data Domain.

Un data domain es un activo predefinido o usuario-definido basado en la semántica de una columna o un campo. (Ex: nuss, dirección IP, estado de una cuenta). Existen dos categorías de Dominio de Datos:

→ Basado en reglas

Si el significado semántico de una columna se puede descubrir con una regla. Estas reglas se pueden enviar de fábrica con Informatica o se pueden crear de forma personalizada según el requisito. Hay 3 categorías generales:

- Basada en Regex: para determinar si los metadatos siguen un patrón.
- Basada en tablas de referencia: para conjuntos finitos de datos que no se superponen.
- Reglas de mapplet: se crean con la herramienta Developer de Informatica y son una combinación de regex, tablas de referencia, búsquedas y otras expresiones (Ej: ver si un valor se encuentra dentro de un rango).

Se recomienda el uso de dominios de datos personalizados para reducir los falsos positivos durante el descubrimiento de dominios de datos. Los dominios de datos personalizados se pueden crear en las herramientas Informatica Analyst, Informatica Developer o Catalog Administrator.

→ Dominios de datos inteligentes (o basado en ejemplos)

No tiene ninguna regla. Es más, como un usuario que etiqueta una columna al observar varios factores, como el nombre de la columna y las estadísticas de creación de perfiles.

Este dominio se puede crear sobre la marcha y los usuarios pueden etiquetar la columna con un dominio de datos.

→ Composite Data Domain

Colección de dominios de datos u otros dominios de datos compuestos linkados utilizando reglas. Resulta fundamental la identificación de columnas similares para que de esta manera se produzca la asignación automática de etiquetas.

En EDC, la similitud de columnas se basa en la agrupación no supervisada y se calcula en función de cuatro factores:

1. Similitud de nombre
2. Similitud de patrón
3. Similitud de frecuencia de valores (también conocida como similitud de datos)
4. Similitud de valor único

4.3 Creación de reglas

Como último paso previo a la generación de reglas vamos a ver las diferentes opciones que existen para la creación de reglas. También veremos que son las expresiones regulares y cuál es el papel que desempeñan en la creación de una nueva regla.

A) Informatica Analyst

Su función consiste en convertir los requisitos empresariales que debe tener una regla en lógica de transformación. Los requisitos empresariales se guardan en una especificación de regla y una vez son compilados la herramienta analyst crea las transformaciones. Dichas transformaciones se almacenan en mapplets (objetos reutilizables).

Las especificaciones de reglas están compuestas por instrucciones (IF-THEN) que utilizan operadores lógicos para determinar si se satisfacen las condiciones especificadas.

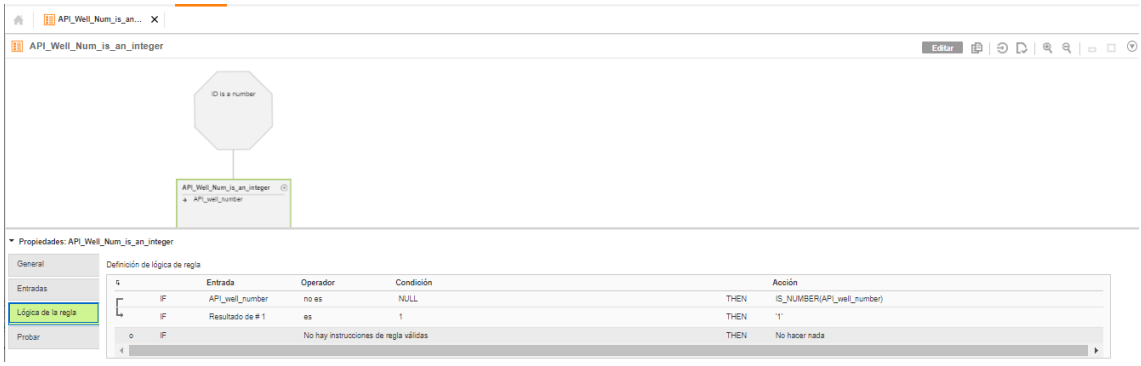

Figura 24. *Instrucción IF-THEN en Analyst*

Concretamente en este ejemplo (Figura 24) se observa la lógica que se ha seguido para la realización de una regla que comprueba que dado un ID introducido este es de tipo numérico.

B) Developer Tool

Esta herramienta permite a los usuarios para diseñar y ejecutar perfiles que analizan el contenido de los datos empresariales y ver si se ajustan a las reglas definidas. Los pasos realizados en estas herramientas son:

En primer lugar, se realizará la importación de los objetos de datos físicos, en este paso se realizará la definición de los procesos de calidad para los datos

asociados a dichos objetos. Seguidamente, se realizará la elaboración de los perfiles de datos este pase revela el contenido y la estructura de los datos. Además, en el perfilado se incluye el análisis de unión, esta comprueba si es posible una unión de datos entre dos columnas dadas. A continuación, se realizará el análisis de los datos para encontrar información de utilidad en los datos y mejorar la estructura de estos. Continuaremos con la estandarización de los datos, paso en el que se eliminarán las inconsistencias y errores encontradas durante la fase de perfilado. El paso final consiste en la validación en la que se evalúa la precisión y se hace corrección de errores en el caso de haberlos encontrado.

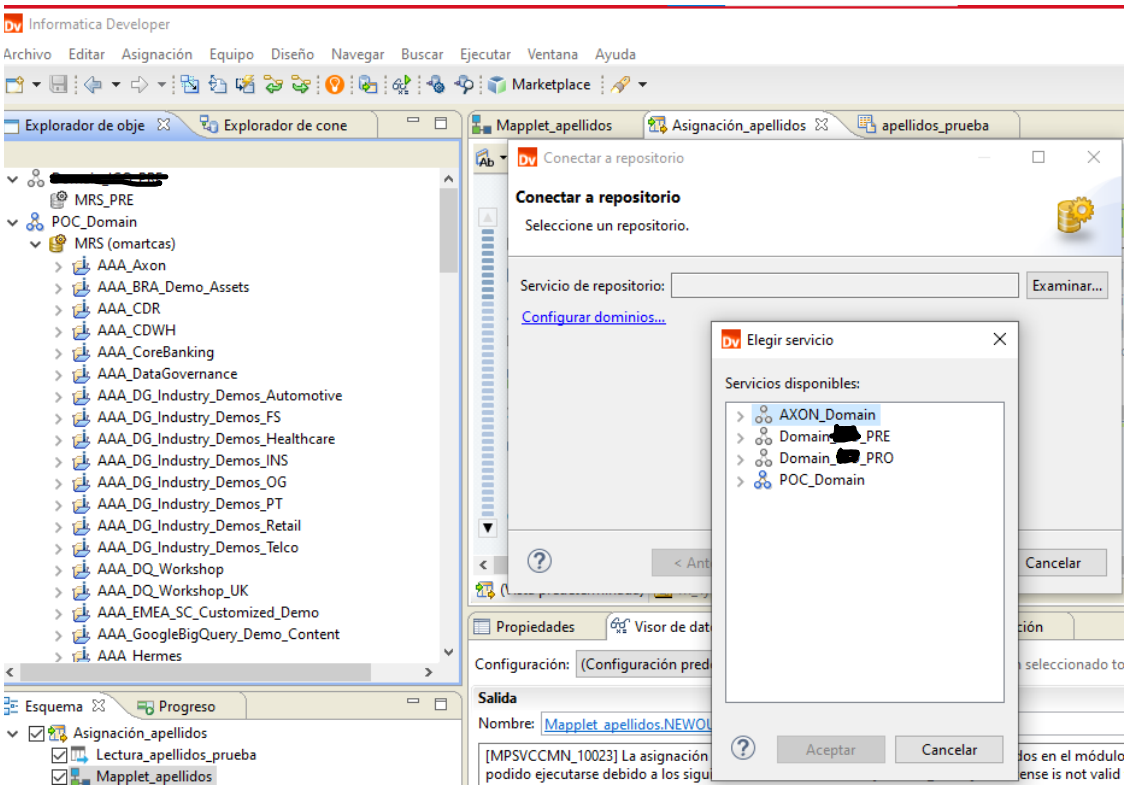

Figura 25. *Conexiones a un repositorio en Developer*

Dentro del developer tool encontramos las expresiones regulares, aunque estas no son una herramienta en las que crear reglas, son la sintaxis con las que se definen estas reglas. Por tanto, se va a explicar que son.

Las expresiones regulares son un conjunto de caracteres cuya función es filtrar textos para encontrar coincidencias como, por ejemplo, identificar direcciones de correo electrónico o documentos de identidad, entre otros.

El principal problema de las expresiones regulares es que son difíciles de identificar a simple vista. Por ello, a continuación, se explicarán las expresiones regulares creadas para definir las reglas en el anexo se ha adjuntado una tabla con los elementos básicos para crear una expresión regular.

En primer lugar, antes de pasar a explicar las expresiones que se han creado se ha cogido una expresión que ya estaba creada para entender su nomenclatura.

La expresión sirve para comprobar la validez de una fecha introducida del tipo DD-MM-YYYY. Se han subrayado sus partes de diferentes colores para facilitar su comprensión.

^ (0[1-9]|[12][0-9]|3[01])[./-](0[1-9]|1[012])[./-](19|20)[0-9]{2}$

Como se puede observar, primeramente, tenemos la parte de color amarillo en la que existen tres supuestos:

- El día empieza por 0 seguido de un número del 1 al 9
- El día empieza por 1 o 2 seguido de un número del 0 al 9
- El día empieza por 3 seguido de un 0 o un 1

En la segunda parte, la verde, se ha utilizado la misma técnica, en este caso solo hay dos supuestos, ya que como se refiere a los meses estos únicamente pueden comenzar por 0 o 1.

Por último, se observa la parte referida al año, la azul, en la que se elige entre 19 o 20 seguido de dos cifras entre el 0 y el 9.

A raíz de esta expresión se ha pasado a crear nuevas, las cuales no estaban definidas previamente, estas han sido comprobadas mediante el uso de la web regex101 y son:

- Expresión para el código postal: "0[1-9]{4}|[1-4]{1}[0-9]{4}|5[0-2]{1}[0-9]{3}$"

La cual debe cumplir la condición de que un código postal que empieza por 0 no puede ir seguido de otro 0, y si empieza por 5 solo puede ir seguido de un 0, 1 o 2. En cualquier caso está formado por 5 dígitos.

- Expresión para número de teléfono español: “(\+34\s?)[67][0-9]{8}”

El cual lleva el prefijo 34 pues es el prefijo para números españoles seguido de un 6 o un 7 como número de inicio y seguido de 8 dígitos más.

- Expresión para un DNI español: “^\d{8}[-]([A-Z]|[a-z]){1}

El DNI español está formado por 8 dígitos seguido de una letra, se ha considerado el caso de que la letra puede ser introducida en mayúscula o minúscula.

4.4 Resolución / implementación

Hasta ahora el enlazado de términos con las columnas físicas con las que tiene relación se hacía de forma manual, revisando una a una las columnas para encontrar los patrones comunes, resultando eso muy costoso para la organización tanto en tiempo como en recursos. Es por ello, que una vez hecha la introducción a los dominios de datos pasaremos a la elaboración de una batería de reglas para automatizar este proceso.

Dada la demanda de proyectos actuales, en su mayoría relacionados con servicios financieros y dado el termino seleccionado (identificador de cliente) esta automatización estará enfocada a abarcar aquellos proyectos de índole financiera.

A continuación, se detallarán los pasos seguidos para la obtención de las reglas:

En primer lugar, desde la plataforma de Informatica (Axon) hemos accedido al glosario y hemos aplicado el filtro de dominios de datos, para proceder a realizar una selección y quedarnos con aquellos que resultan de interés para nuestro proyecto.

Hemos seleccionado el dominio Servicios Financieros, dentro de este también se verán las tablas de las entidades y los términos con los que tiene relación.

Figura 26. *Jerarquía de glosario del término servicios financieros*

En la imagen anterior (Figura 26), se pueden observarlas entidades que hay comprendidas dentro del domino de servicios financieros. En primer lugar, vamos a ahondar en la entidad de organización de datos, la cual describe la compañía y su estructura.

Por otra parte, también se ha seleccionado el dominio Retail, el cual como se puede ver contiene información acerca de los clientes, como son sus datos personales, país, comunidad autónoma, provincia, etc.

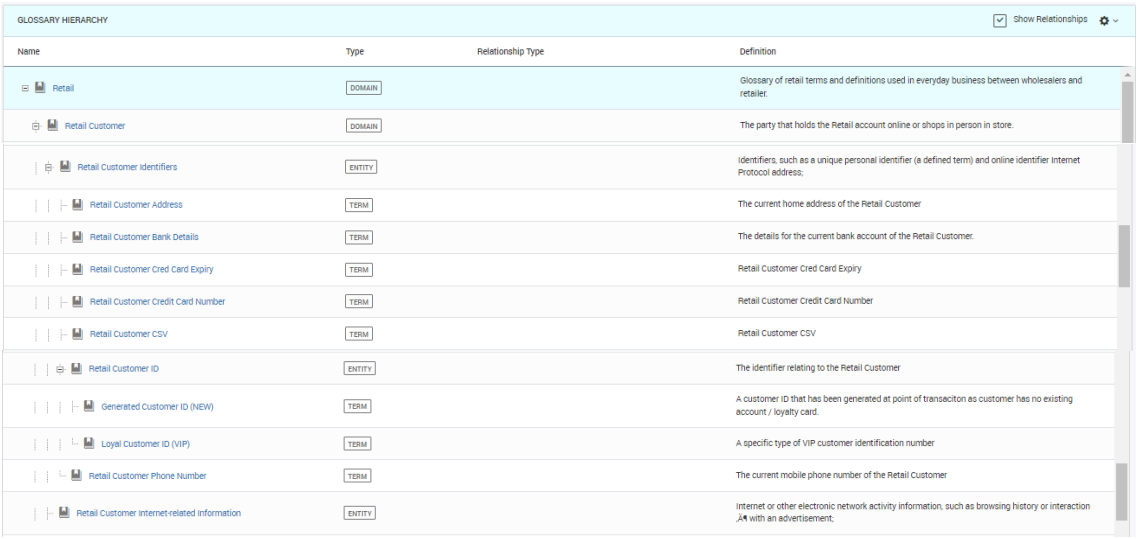

Figura 27. *Jerarquía de glosario del dominio Retail*

En concreto, cuando hemos accedido a las tablas que componen los términos que observamos en la imagen anterior hemos observados cuales son aquellos términos que carecen de reglas. Se muestra a continuación, algunos ejemplos detectados.

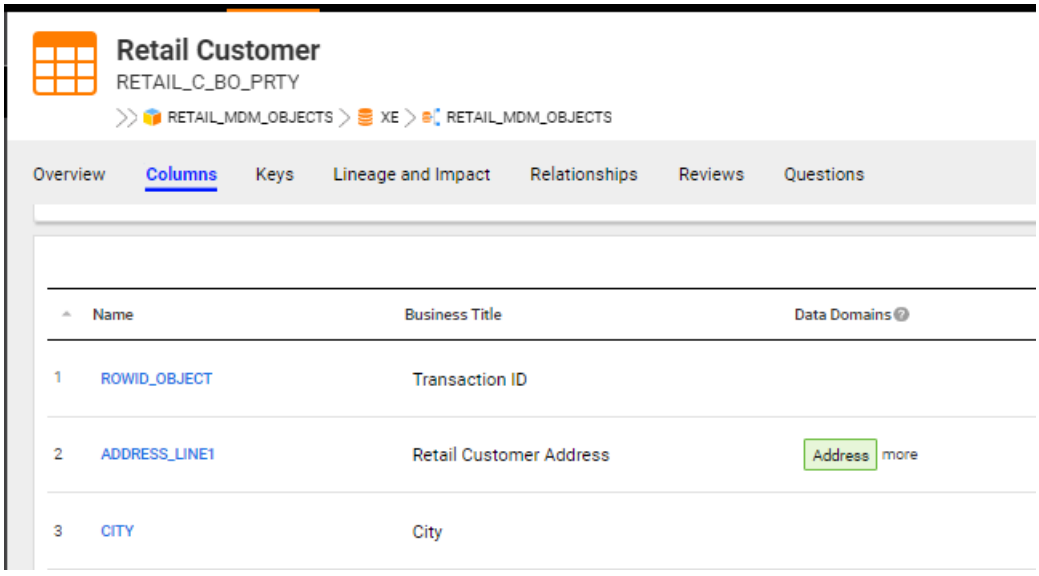

Figura 28. *Ejemplo de reglas en términos*

Como se ha comentado anteriormente la herramienta de Axon nos puede redirigir hasta la de catálogo de datos, para ello, como se ha visto en la Figura 11, una vez seleccionamos el padre al que queremos llegar la información que obtendríamos sería la siguiente:

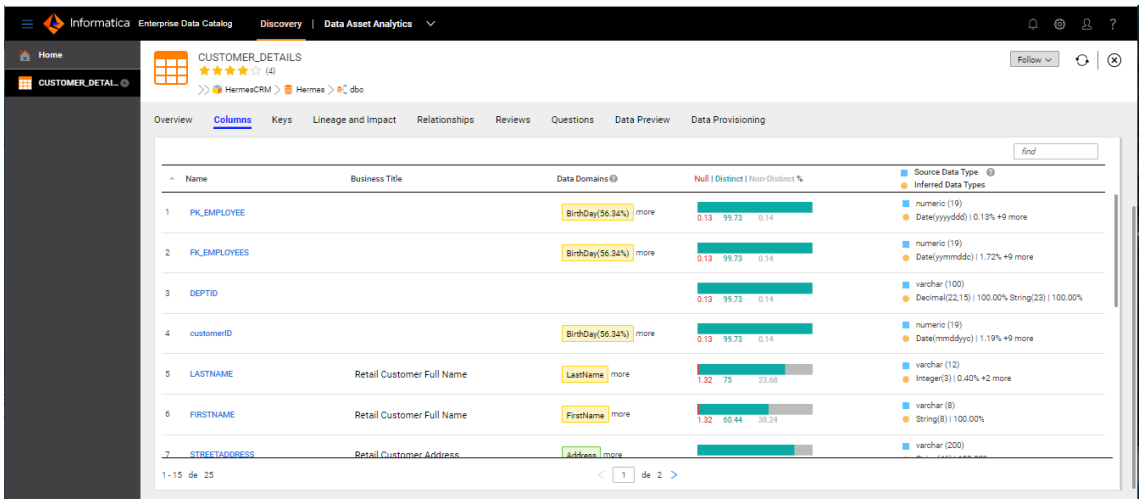

Figura 29. *Vista de un término de Axon en EDC*

Una vez estamos en el catálogo de datos tenemos acceso a todas las columnas de las que se dispone, así como el linaje o las relaciones, entre otros.

4.4.1 Obtención de los datos

Una vez detectados y seleccionados esos dominios de datos se ha decidido cuales son las reglas que pueden aportar valor a la organización y tras observar los términos que no tenían una regla definida se ha decidido abordar los siguientes casos:

1. Nombre
2. Apellidos
3. Provincia
4. Comunidad Autónoma

Dichas reglas nos ayudaran a verificar la calidad y fiabilidad de los datos introducidos. En el caso del nombre, la provincia y la comunidad autónoma se hace para comprobar principalmente que el input introducido es correcto, es decir, existe y está bien escrito. Pero, sin embargo, en el caso de los apellidos se ha detectado que ambos apellidos están escritos en la misma columna seguidos, por ejemplo “López Martínez” y que se necesita tener cada uno de los apellidos en columnas distintas.

Una vez se ha hecho esta selección era necesario obtener datos fehacientes con los que poder contratar los inputs de los usuarios, para ello, se ha recurrido a la base

de datos del Instituto Nacional de Estadística (INE), con el objetivo de aprovechar la fiabilidad que ofrece dicho organismo oficial.

Finalmente, las muestras obtenidas han estado formadas por 54.916 registros en el caso de los nombres y 78.355 en el caso de los apellidos, siendo estos los nombres y apellidos registrados en España con una frecuencia superior a 20, esta muestra se presupone suficiente para validar la calidad de los datos. En el caso de las comunidades y las ciudades autónomas se han incluido las 19 y en el caso de las provincias las 54.

Además, en el caso de la regla utilizada para separar el campo “apellidos” en dos columnas se ha utilizado un generador de apellidos online que ha proporcionado apellidos aleatorios que se han juntado y con los que posteriormente se ha testeado la regla creada.

4.4.2 Implementación de las reglas

Ahora sí, se ha procedido a la creación de reglas mediante el uso de la herramienta developer, las reglas creadas han sido las siguientes:

En primer lugar, se va a dar una visión general acerca de cómo es el proceso de creación de una nueva regla. Primeramente, se ha creado una tabla de referencia en la cual se ha pasado el fichero con los datos obtenidos del INE. A continuación, se ha creado el mapplet de prueba (objeto para transformaciones reutilizable) en el que se han realizado las transformaciones necesarias, incluyendo estas la creación de un input, seguido de una búsqueda a la tabla de referencia, una expresión en la que comprobar si los datos coinciden y por último un output. El detalle de cada una de las reglas creadas será mostrado a continuación.

- La primera regla creada se la ha llamado *provincias,* puesto que, su objetivo que es que dado un input de tipo string se valida si esa entrada corresponde con una provincia de las aceptadas, es decir, pertenece a la lista de provincias existentes en España.

  Como se acaba de comentar en esta primera imagen se puede observar la tabla de referencia creada para las provincias, en ella se han incluido todas las provincias españolas existentes en el territorio español.

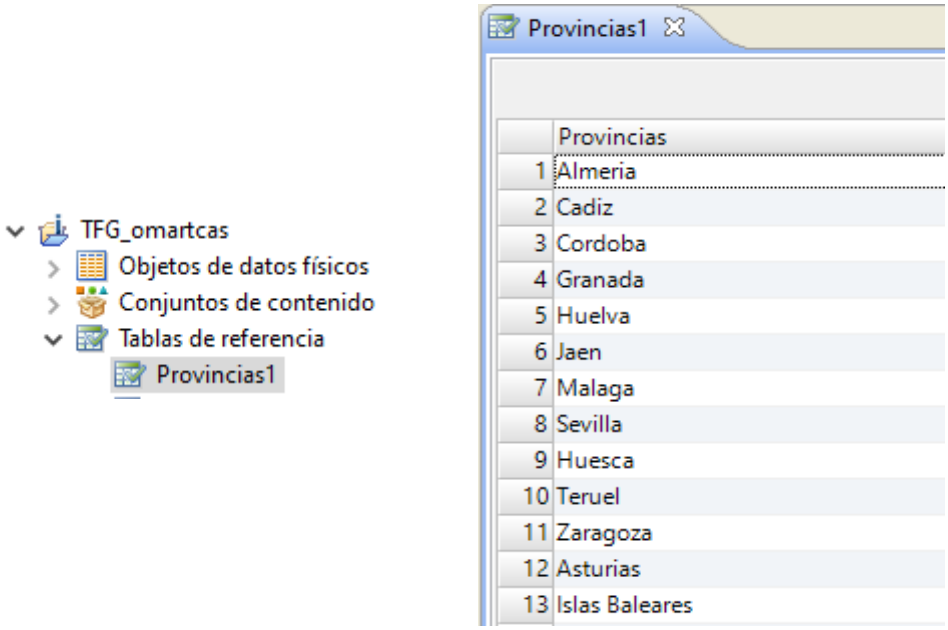

***Figura 30**. Lista de provincias introducidas*

Tras esto, se ha procedido a crear el mapplet de provincias, para ello, como vemos en la imagen se ha introducido un input, seguido de una transformación de búsqueda.

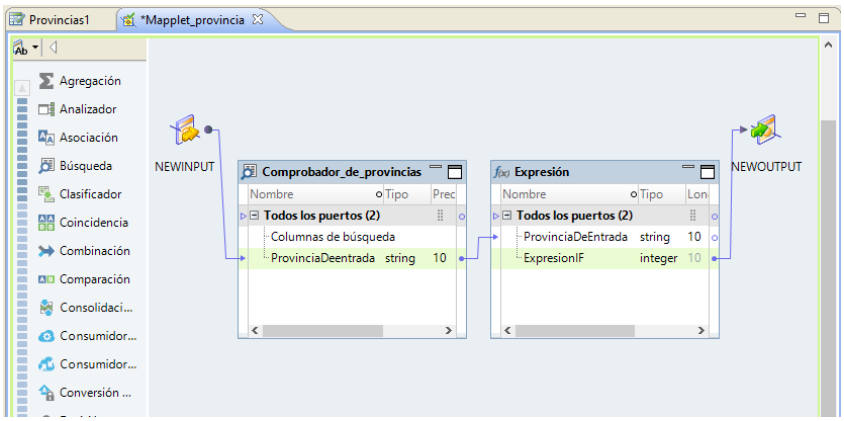

***Figura 31.** Regla de calidad para las provincias*

A continuación, vemos una transformación que contienen una expresión en la cual se ha comprobado si el input que se le ha pasado coincide con alguno de los valores de la lista de provincias. En el caso de que el input introducido sea null devuelve false, es decir, 0, en caso contrario (true), devuelve 1. Por último, añadimos el output de salida.

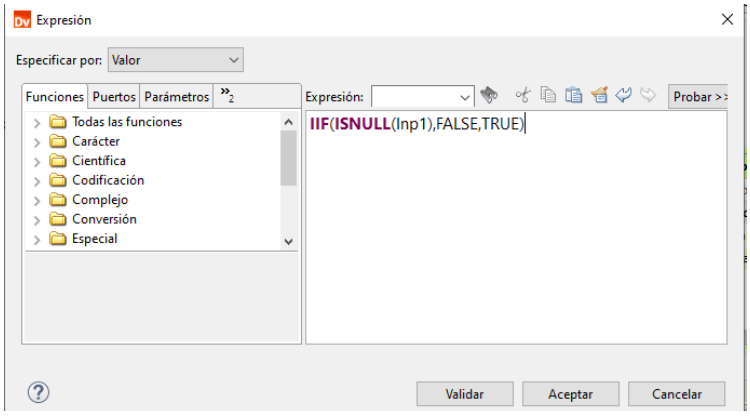

Figura 32. *Condición de comprobación para las provincias*

El siguiente paso ha consistido en crear la asignación, en la cual se ha elegido una tabla de tipo objeto de datos con los datos de origen, seguido del mapplet que se ha creado anteriormente y por último una tabla de tipo objeto de referencia, necesaria para la escritura de la salida. La asignación de prueba se ejecuta para comprobar que la regla funciona correctamente y en el visor de datos podemos observar los resultados obtenidos.

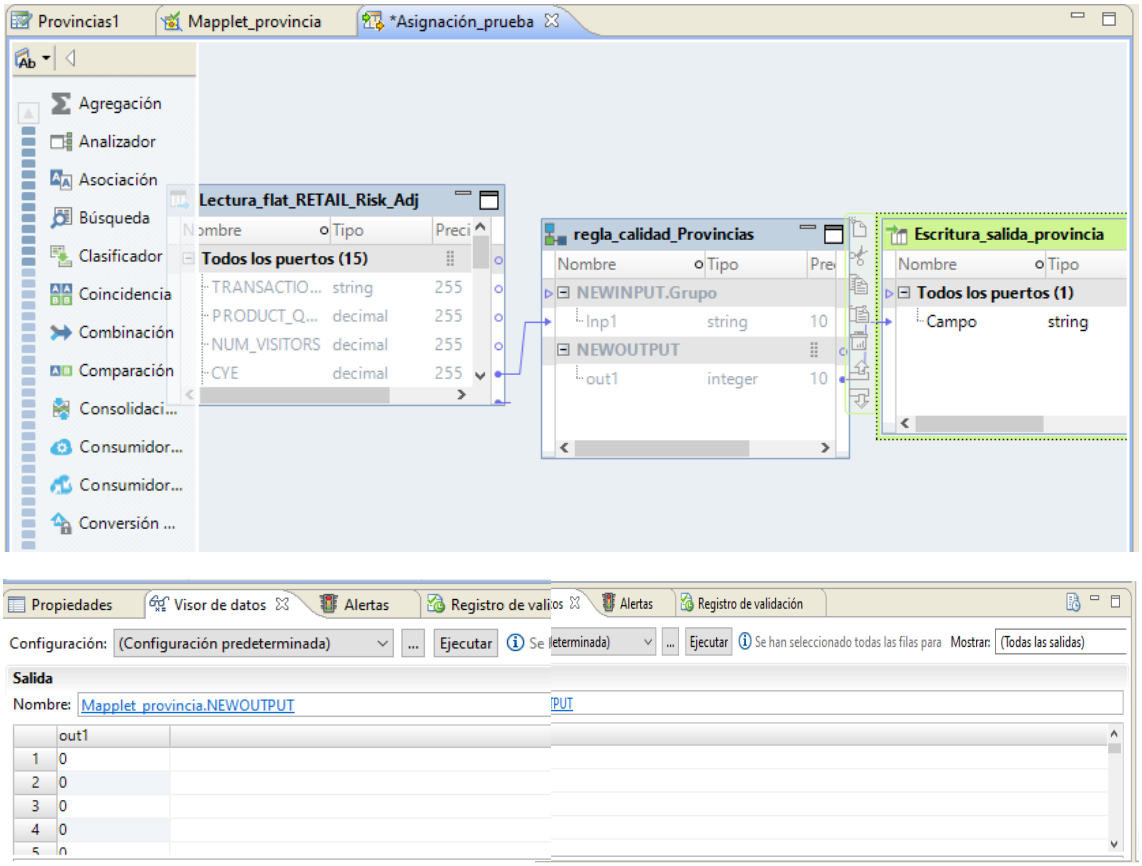

Figura 33. *Prueba de la asignación en la regla provincia*

Para las dos siguientes reglas creadas, nombre y comunidad autónoma las transformaciones que se han seguido has sido las mismas.

Aunque se conoce que la regla funciona porque se ha testeado con la expresión introducida. Se han realizado tres tipos de prueba más, para ello, se han subido los ficheros correspondientes a la máquina remota de IDQ, estos se han cagado mediante objetos de datos de archivo sin formato. Las pruebas han sido:

1) Pasar como input la misma tabla a comprobar, es decir, en el caso de las provincias pasarle la propia tabla con las provincias, en el caso de los nombres la tabla con los nombres y en el caso de las comunicades autónomas la tabla con todas las comunidades autónomas y las dos ciudades autónomas.

2) Pasar una lista como vacía, en este caso debe devolver todo a false, ya que se cumpliría la condición de la expresión que dice que si es falso devuelva 0.

3) Pasar como input una lista que contiene provincias que son correctas y, por tanto, deben devolver true pero también provincias que son incorrectas, en ese caso debe devolver false.

Al ejecutar las pruebas se ha detectado que todos los valores de las tablas obtenidas del INE estaban en mayúscula y que las tablas en las que se han pasado los valores para realizar las pruebas estaban en minúscula, para ello, se ha añadido una transformación más en la que convierte a mayúscula los valores de las tablas con los datos de la prueba.

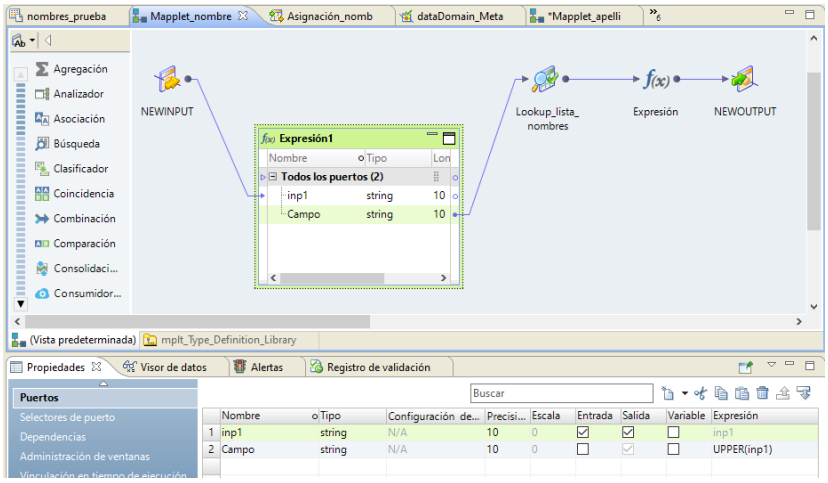

Figura 34. *Regla de calidad para los nombres*

Para realizar las pruebas que se acaban de comentar se han creado un objeto de datos con una lista de valores de pruebas por cada una de las reglas creadas. Como se ha comentado anteriormente, dichos ficheros están en los servidores

(máquinas Linux creadas en Amazon Web Services), por tanto, la herramienta deberá conectarse a dichos servidores. Para ello, una vez cargada la lista con las palabras de prueba, se cambiará el origen del archivo para que apunte a dicho servidor.

Tiempo de ejecución: lectura	
Tipo de entrada	Archivo
Tipo de origen	Directo
Nombre del archivo de origen	apellidos-prueba.csv
Directorio del archivo de origen	/home/infauser/pruebas-mapping
Partición de lectura simultánea	Optimizar el rendimiento
Tipo de conexión	Ninguna
Comando	

Figura 35. *Cambio de directorio de archivo*

Para acceder a dichos servidores en los que alojar dichos ficheros de prueba se han empleado la herramienta WinSCP, esta aplicación presenta la siguiente forma.

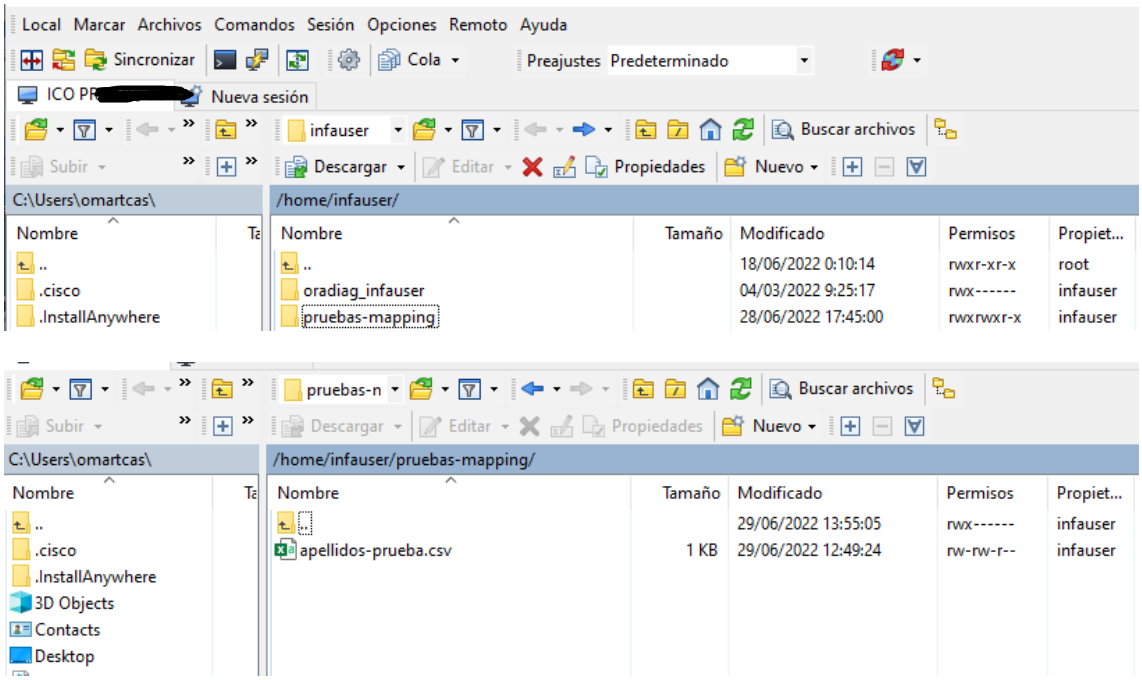

Figura 36. *Subida de fichero de pruebas a WinSCP*

La salida que obtenemos después de comprobar la regla (nombres) vemos que cumple con la condición de que se estableció, pues muestra un 1 cuando el valor sí que se encuentra en la lista, en este caso Ana, Luis y Pablo que son nombres existentes y 0 en el caso de los nombres que se han inventado.

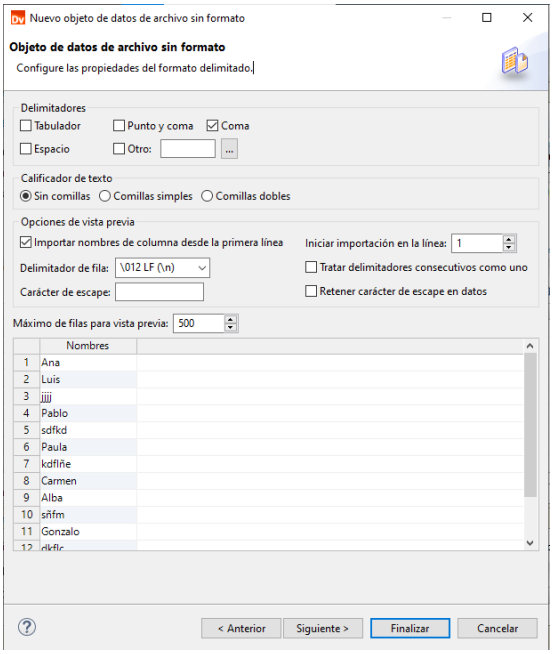

Figura 37. *Listado de nombres de prueba*

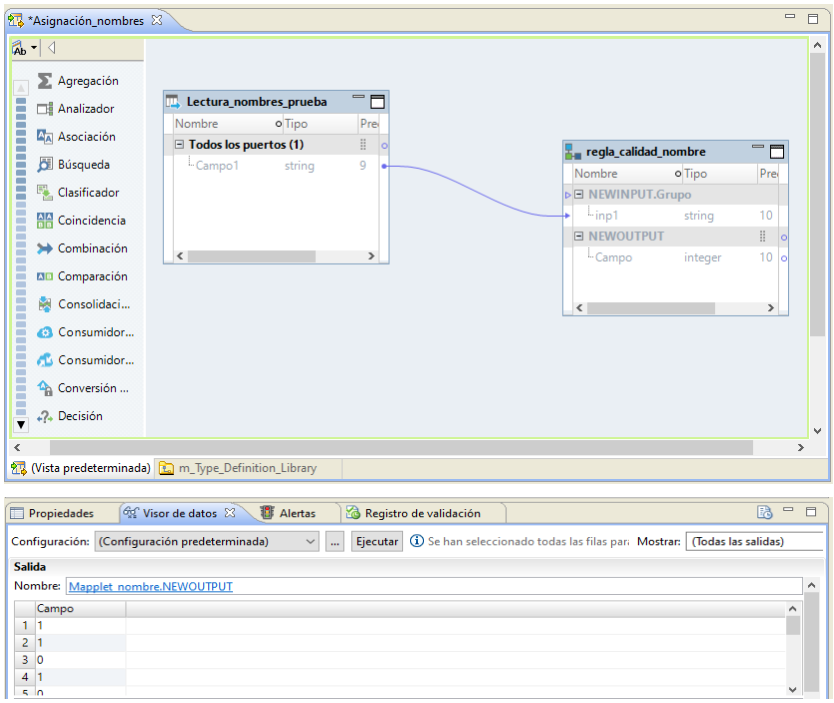

Figura 38. *Salida de la prueba de la regla de calidad*

Por lo tanto, como se observa en las imágenes anteriores para los nombres de entrada de prueba que se le han pasado la salida sería del siguiente tipo:

1-nombre existente | 0-nombre inexistente

Tabla 1. *Aclaración de la salida de la regla de calidad*

NOMBRE DE ENTRADA	SALIDA
Ana	1
Luis	1
jjjj	0
Pablo	1
sdfkd	0

La siguiente regla creada, es la utilizada para separar los apellidos de una columna y ponerlos en dos columnas distintas. Esta regla en concreto representaba un problema, ya que, tener una columna con ambos apellidos juntos no resultaba útil y dada una lista de usuarios de, por ejemplo, una entidad financiera esta se puede alargar a miles de usuarios, lo cual resulta muy costoso tanto en tiempo como en recursos, además de los errores manuales que dicha actividad llevaría asociados. Al ser esta regla un poco más compleja que las anteriores se ha precisado la realización de un proceso ETL (extracción, transformación y carga) que sea capaz de filtrar la columna y detectar dos apellidos contiguos. El proceso de creación de la regla ha sido el siguiente:

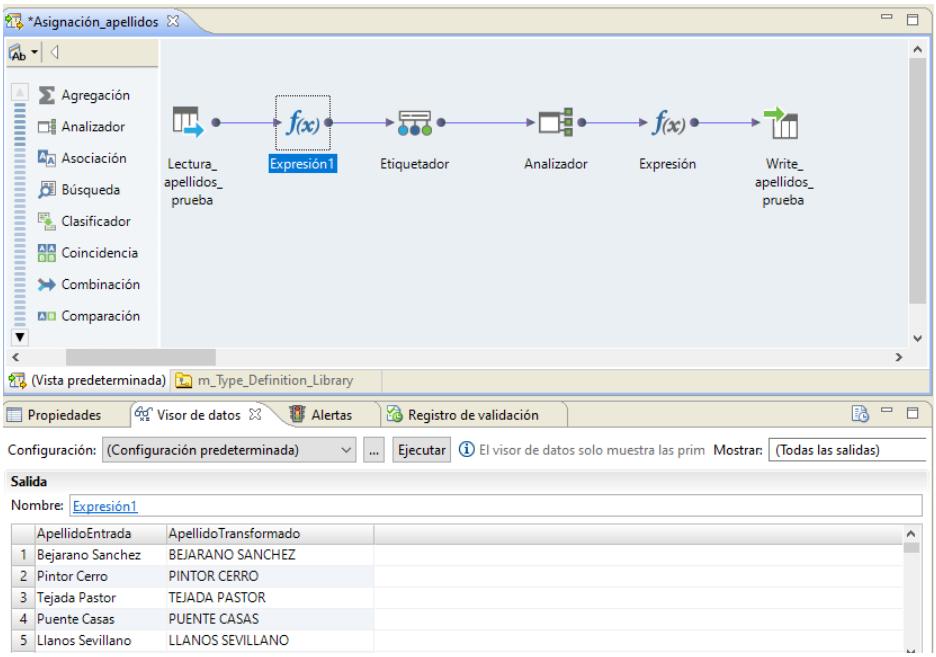

Figura 39. *Transformación a mayúsculas*

En primer lugar, se observa una primera expresión, en la cual se han transformado los apellidos de prueba a mayúscula, pues se ha observado que en la base de datos obtenida del INE estaban en mayúscula, de no cambiarlos ninguno de ellos se mostraría como coincidencia.

Respecto al lanzamiento de las pruebas para comprobar la validez de la regla se han presentado dos opciones:

1) Primeramente, una de las dos opciones consistía en añadir un objeto de transformación de tipo etiquetador (labeler). Dentro de este etiquetador se ha llamado a la tabla de referencia creada anteriormente que contiene todos los apellidos de la muestra, dicho labeler va comprobando si alguna parte del input introducido coindice con alguna palabra de las de la lista de muestra, en caso de ser así cambia esa palabra por “apellido”. Por ejemplo, en el caso de pasar como entrada “López García” tras pasar por el etiquetador la salida sería “apellido apellido” de esta forma sabríamos que en esa columna están dos apellidos seguidos. Si por el contrario dicha transformación se comprobara sobre una columna que no tiene dos apellidos juntos, por ejemplo, “Ana Pérez” la salida que obtendríamos sería “Ana apellido”, por tanto, se sabría que en dicha columna no hay dos apellidos juntos.

2) Por otra parte, la segunda opción que se ha planteado era elaborar una tabla con todas las combinaciones de los apellidos posibles, es decir hacer el producto cartesiano de los apellidos de la muestra y juntar “todos con todos”, para que de esta forma cuando se le pase un input sea capaz de detectar si se trata de dos apellidos juntos o si por el contrario es otro tipo de entrada. Aunque esta opción resulta menos costosa en que a ejecución respecta, se ha decidido hacer el proceso como se indica en la opción uno, pues para el proyecto resulta más interesante ver el procese de transformación ETL que se ha realizado para la obtención y comprobación de la regla.

Seguidamente, como acabamos de comentar se ha elegido la opción del etiquetador, al cual se le ha pasado la tabla de apellidos del INE, que contiene todos los valores como referencia y se le ha indicado que en caso de encontrar una coincidencia sustituya dicha palabra por “apellido”.

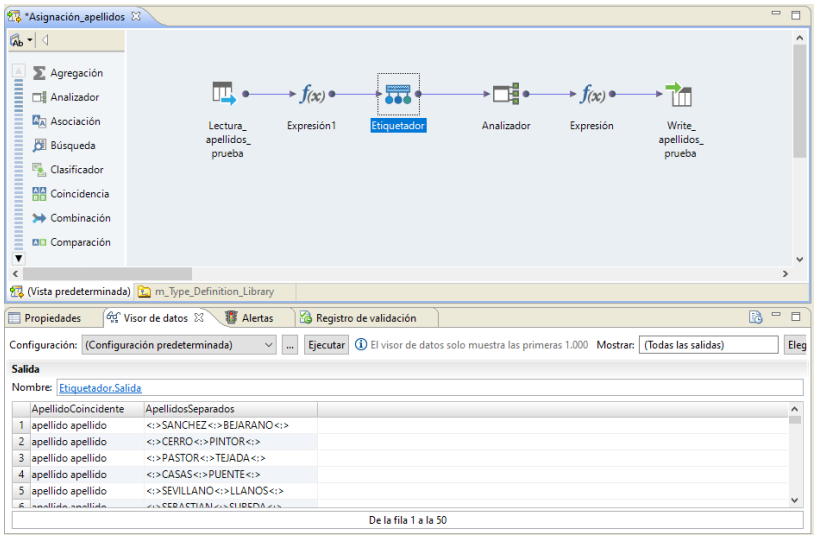

Figura 40. *Comparador de apellidos con tabla de referencia*

La siguiente transformación que observamos es el analizador, en este paso una vez se han identificado los dos apellidos correctos o no, se encarga se separarlos en dos columnas, así mismo se ha añadido una tercera columna en las que se muestran aquellos apellidos que no son correctos. Por tanto, la ejecución del analizador quedaría de la siguiente manera.

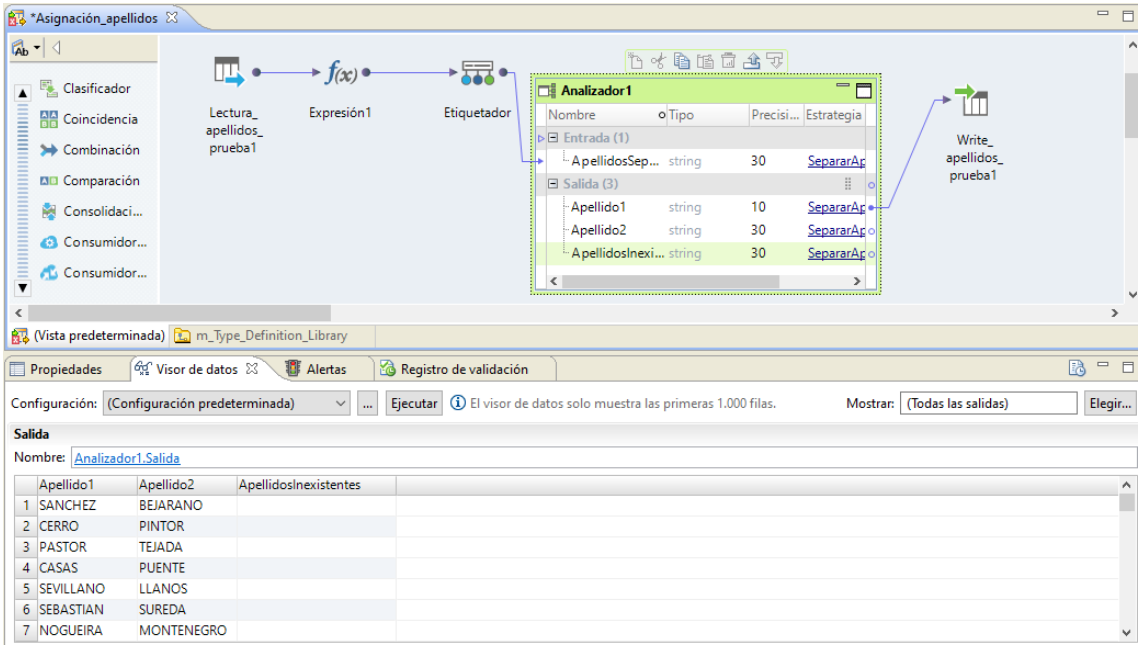

Figura 41. *Separador de apellidos en columnas distintas*

Esta primera imagen muestra el caso en el que ambos apellidos con correctos, mientas que la siguiente muestra la salida cuando uno de los dos apellidos es erróneo.

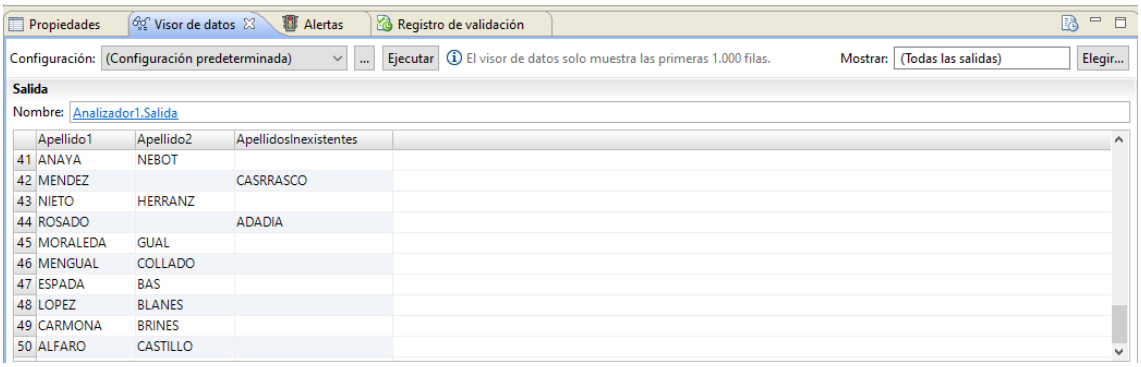

Figura 42. *Salida de la separación de apellidos erróneos*

La siguiente expresión, que se puede observar en la imagen 43 contiene la condición de comprobación y por último se encuentra el fichero de escritura de salida. Dicha condición se puede observar en la siguiente imagen:

Figura 43. *Expresión de comprobación de apellidos*

Anteriormente se había realizado la creación de la regla primeramente en un mapplet y luego se había realizado el mapeado de la misma a través de una asignación. Sin embargo, en esta ocasión se ha prescindido del mapplet y se ha elaborado directamente sobre la asignación. Esta técnica comporta algunas ventajas como son el poder ejecutar la regla creada paso a paso y ver los resultados que se van obteniendo, por el contrario, también cuenta con desventajas y es que, un cambio que se realice en la regla deberá ser modificado manualmente, es decir, al hacer uso del mapplet y "arrastrarlo" a la asignación una modificación realizada en dicha asignación o cualquiera de las que contenga el mapplet será realizada de manera automática.

4.4.3 Mejoras asociadas a la creación de reglas

En este penúltimo subapartado del trabajo y antes de pasar a exponer las conclusiones, se van a tratar los cambios y las mejoras que la implementación de las reglas ha supuesto para la organización.

Para comenzar, se va a hacer una comparativa de cómo habría sido la resolución del caso de uso sin la creación de las reglas, para luego introducir los cambios que estas han supuesto. Esto se ha dividido en tres apartados:

1. Cuanto se ha tardado en hacer las reglas

Este ha resultado quizás uno de los pasos más costosos, pues había que ir repasando las relaciones de los 1089 términos del glosario para entre los 6 dominios existentes ver que término contaban con reglas y determinar la viabilidad de crear y asociar nuevas.

2. Cuanto se ha tardado en configurar las reglas

Respecto a la creación de las reglas, en el apartado anterior (4.4.2) se ha visto que compartían las bases de la transformación, por tanto, una vez se ha realizado y validado la primera de ellas el tiempo de creación de las siguientes se ha reducido considerablemente. Sin embargo, respecto a la regla de la creación de los apellidos, aunque su definición ha resultado notablemente más costosa, a continuación, veremos los beneficios que reporta.

3. Cuanto han tardado en ejecutarse (volumetría)

Respecto a la volumetría, es decir, coste de ejecución de las reglas se prevé enormemente reducido, pues estas reglas de calidad pueden ser aplicadas de manera eficiente sobre todos los datos de la organización. Sin embargo, si no se dispusiera de este tipo de reglas el procedimiento de revisión de todas las tablas sería costoso a la par que tedioso. Pues como podemos observar en la siguiente imagen en un solo recurso ya existen 87 tablas.

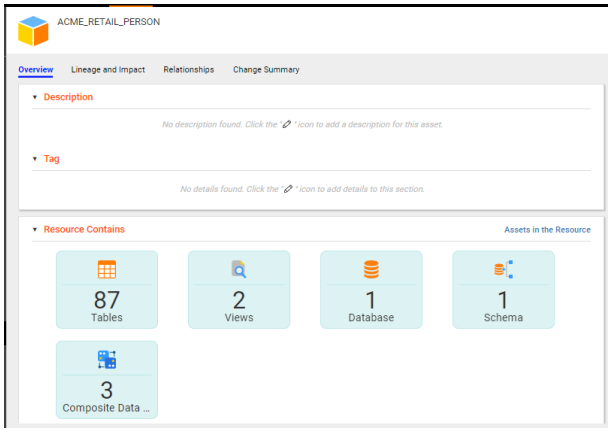

Figura 44. *Ejemplo del número de tablas de un recurso*

El planteamiento de dichas reglas no sólo ayuda al descubrimiento de nuevos dominios, sino que también es beneficioso para el descubrimiento de posibles datos que pudieran existir en la organización y que fueran desconocidos, realizando el lanzamiento de reglas sobre todos los datos de la organización, no sólo ahorro económico (recursos), sino de tiempo. Además, de reducir el número de errores humanos cometidos y conseguir nuevos valores intangibles para las compañías.

4.4.4 Aplicaciones en casos reales

Una vez hemos vistos los beneficios que se obtienen en la organización, a partir de la creación de reglas ahora se van a ver algunas de las aplicaciones de estos tipos de proyectos en casos reales.

Este tipo de proyectos que se ha descrito a lo largo de todo el trabajo resulta muy útil en organizaciones en las que se quiera aplicar el gobierno del dato. En las cuales es necesario realizar el transporte o migración del data warehouse o almacén de los datos a la nube, debido a su volumen. En este proceso intervienen nuevas herramientas con las que llevarlo a cabo, entre ellas, Azure (plataforma cloud que ayuda a crear soluciones a los problemas o dificultades existentes en la actualidad, permite crear, ejecutar y administrar aplicaciones en varias nubes); o Snowflake (plataforma de tipo cloud orientada a la carga y el tratamiento masivo de datos de las organizaciones).

Además de hacer uso de estas dos posibles herramientas que se acaban de comentar, resulta imprescindible el uso del catálogo de datos el cual permite descubrir activos de los datos registrados y desbloquear su potencial, así como facilitar la integración de estos en otras herramientas.

Por último, el uso y tratamiento de estos datos conduciría convertirse en una compañía data driven, lo cual consiste en construir herramientas, habilidades y lo más importante una cultura que haga énfasis en el dato (Anderson 2015).

Para ello, existen dos requisitos que las organizaciones deben tener en cuenta, en primer lugar, la organización debe recoger datos de calidad, relevantes, no sesgados y lo más importantes creíbles. El segundo requisito es que los datos deben ser accesibles y deben poder realizarse consultas, es decir, su acceso debe estar garantizado por la organización en cualquier momento en el que sea necesario.

5 Conclusiones

En este último capítulo, se van a reunir los principales aspectos vistos a lo largo del trabajo, destacando la importancia de los apartados más relevantes en el trabajo: el gobierno del dato y la creación de reglas de calidad para el impulso de este. Además, se incluirá una interpretación de los resultados que aplicar dichas reglas de calidad han reportado a la organización.

5.1 Síntesis

En la organización en la que se ha desarrollado el presente trabajo de final de grado se había detectado no solo el incremento de proyectos de gobierno del dato, sino la importancia de la aplicación en las empresas y los beneficios que dicho gobierno les reporta, por ello, se detectó la necesidad de cubrir los objetivos que a continuación se detallan.

El objetivo principal del trabajo consistía en explicar de manera detallada en qué consiste un proyecto de gobierno del dato y cuáles son las diferentes fases que pasa en su desarrollo. Además de la aplicación práctica de un caso de uso, para ver de manera práctica la aplicación de las reglas de calidad que se utilizan en este tipo de proyectos. Para alcanzar dicho objetivo ha sido necesaria la implicación de objetivos secundarios relacionados con las herramientas tecnológicas en la que se contextualiza el trabajo.

El primer objetivo implicaba dar una visión acerca de qué consiste un proyecto de gobierno del dato, para ello, se han definido los cuatro pasos de acción: establecer responsables, decidir el almacenamiento de los datos, establecimiento de normas para determinar su uso, establecer controles y procedimientos de auditoría. Esto se encuentra directamente ligado al siguiente objetivo que consistía en realizar un diagnóstico y determinar el grado de madurez, para ello, se puede hacer uso de herramientas adicionales que faciliten el diagnóstico como Data Compass, basada en cinco principios: valor de negocio, gobierno de madurez, cultura del dato, analítica avanzada, arquitectura. Adicionalmente, una vez analizados dichos 5 componentes se facilita un cuestionario a los empleados en el que valoren ellos mismos la madurez de los datos de la organización.

El siguiente objetivo consistía en la descripción y comprensión de cuáles son los pasos previos a realizar, para ello, en primer lugar se ha determinar las tareas a realizar y el orden en el que han de ser ejecutadas, posteriormente se pasa al plan de despliegue, el cual consta de siete pasos (alcance, asignación de roles, comunicación, formación, requerimientos funcionales, desarrollo técnico y clasificación en tiers), los cuales son fundamentales para la consecución de las tareas previamente definidas.

Por otro lado, el siguiente objetivo, era conocer las herramientas del mercado que se dedican al gobierno del dato, entre las elegidas se encontraban Ovaledge, Xplenty, Talend y Anjana, para ello, se analizaron las ventajas y funcionalidades que ofrecen dichas herramientas, competidoras, frente a las que ofrece la herramienta de Informatica.

El siguiente objetivo consistía en la definición de un caso de uso, en el que investigando dominios de datos se encontraran patrones de datos que pudieran ser utilizados como reglas, en ese caso, el caso de uso de ha definido a partir de dos términos. Por una parte, el termino *FS Customer ID* el cual se encuentra englobado dentro del dominio de Servicios Financieros. Por otra parte, dentro del dominio Retail se han identificado diversos términos que carecían de regla, los cuales son provincia, comunidad autónoma, nombre y apellidos.

Además, otro de los objetivos era conseguir una muestra suficiente y fehaciente con la que poder validar y contrastar la calidad de los datos introducidos, para ello, se recurrió a la base de datos del INE, obteniendo así una lista con todas las comunidades autónomas y provincias que forman parte del territorio español, así como los nombres y apellidos registrados con una frecuencia superior a 20.

A partir de la definición del caso de uso y la posterior obtención de la muestra de validación, se ha pasado a la creación de las reglas, las cuales se han realizado en la herramienta Informatica Developer y cuyas pruebas se han contrastado con la ayuda de la herramienta de calidad de datos de la misma Suite, así como la ayuda de otras aplicaciones que han facilitado el uso de los servidores para subir los ficheros de pruebas.

Por tanto, podemos concluir que todas las reglas que se han creado resultan de vital importancia para la organización sobre todo en lo que a ahorro de costes y reducción de errores cometidos por humanos respecta. Así como, para el descubrimiento de dominios de datos en toda la organización, en general.

La creación de reglas no solo ayuda a mejorar la trazabilidad de los datos, sino que ayuda a determinar de una manera más completa la calidad de estos y el lugar que ocupan dentro de la organización.

Además, se puede afirmar que el gobierno del dato resulta un enlace clave entre las áreas de negocio y la visión técnica, pues ayudan a clarificar el origen de los datos y, por tanto, a definir el uso y la aplicación que van a tener. Asimismo, también resultan una herramienta muy potente para asegurarse del cumplimiento de las regulaciones legales vigentes en lo que al tratamiento y protección de datos informáticos respecta (GDPR).

5.2 Perspectivas de futuro

En este último apartado del proyecto se van a tratar aquellas perspectivas o ideas de mejora que se hayan detectado durante la realización de este.

En primer lugar, el proceso de gobierno del dato se trata de un proceso de tipo iterativo, el cual debe ser probado y revisado durante varios meses antes de poder decirse que se está satisfecho con los resultados obtenidos y puede ser aplicado a los datos de la organización.

Dado que tanto este proyecto como la resolución del caso de uso que en él se ha tratado estaban orientado a proyectos de índole bancaria o financiera, se deberían aplicar unas determinadas reglas de calidad para el descubrimiento de dominios de datos. Por tanto, dado que el tema de los proyectos se encuentra relacionado conforme se va avanzando en el descubrimiento de las reglas y se tiene una mayor batería de estas, se podría elaborar un paquete que las contenga que ayude a facilitar el proyecto

Además, como otra posible idea sería el desarrollo de procesos que sean capaces de arreglar los datos y de esta forma devolverlos al origen, por tanto, se dejaría únicamente de abarcar un proyecto de tipo gobierno del dato para convertirse en un proyecto complementario de calidad de los datos.

6 Bibliografía

Anderson, C. (2015). *Creating a data-driven organization: Practical advice from the trenches*. " O'Reilly Media, Inc.".

Anjana Data. (2022, 12 mayo). *Funcionalidades | Anjana Data: solución para el Gobierno del dato*. https://anjanadata.com/solucion/funcionalidades/

Bonet, M. (2021). *El gobierno del dato ante la transformación digital.* If geek then. https://ifgeekthen.nttdata.com/es/el-gobierno-del-dato-ante-la-transformación-digital-ciclo-el-gobierno-del-dato-y-transformación

Chien, M (2021). Cuadrante mágico para soluciones de calidad de datos. Gartner. https://www.gartner.com/cuandrante-magico-para-calidad-de-datos

Data Cloud. (2022, 29 abril). Snowflake. https://www.snowflake.com/?lang=es

Engler, S. (2020). *Lack of Skills Threatens Digital Transformation*. Gartner. https://www.gartner.com/smarterwithgartner/lack-of-skills-threatens-digital-transformation

INE. Instituto Nacional de Estadística (2022). Obtenido de https://www.ine.es

Informatica. (2021). *2021 Gartner Magic Quadrant for Data Quality Solutions*. https://www.informatica.com/es/data-quality-magic-quadrant.html

Informatica. (s. f.). *Intelligent Data Management Cloud | Informatica España*. https://www.informatica.com/es/platform/data-management/magic.cuadrants

Informatica Data Quality: principales puntos fuertes de la herramienta. (2014). Power Data. https://blog.powerdata.es/data-quality-principales-puntos-fuertes-de-la-herramient

Informatica Developer Tool. (2018). Informatica. Obtenido de

https://docs.informatica.com/data-quality-and-governance/data-quality/10-2/getting-started-guide/getting-started-overview/the-tutorial-structure/informatica-developer-tool.html

Magic Quadrant Research Methodology. (s. f.). Gartner. https://www.gartner.com/en/research/methodologies/magic-quadrants-research

Miñana, A. (2020, 29 mayo). Resumen Webinar: La importancia del Gobierno del Dato en el framework de DAMA. Anjana Data.

https://anjanadata.com/resumen-la-importancia-del-gobierno-del-dato-en-el-framework-de-dama/

Otto, B. (2011). Organizing Data Governance: Findings from the Telecommunications Industry and Consequences for Large Service Providers. Communications of the Association for Information Systems, 29, pp-pp. https://doi.org/10.17705/1CAIS.02903

OvalEdge. (2021, 21 julio). SoftwareAdvice. Obtenido de https://www.softwareadvice.es/software/260635/ovaledge#about

Qué es Azure: Servicios en la nube de Microsoft. (s. f.). Microsoft Azure. https://azure.microsoft.com/es-es/overview/what-is-azure/

Random Name Generator - First & Last Names. (s. f.). FossBytes. Obtenido de https:// fossbytes.com/tools/random-name-generator

Regex. (s. f.). *regex101: build, test, and debug regex*. Regex101. https://regex101.com

Xplenty. (2019, 3 octubre). *Integrate.io*. GetApp. Obtenido de https://www.getapp.es/software/107375/xplenty

Talend. (2018, 30 julio). *Opiniones de Talend Data Fabric*. Capterra. Obtenido de https://www.capterra.es/reviews/118978/data-integration

Anexo A

Anchors	
^	Start of string, or start of line in multi-line pattern
\A	Start of string
$	End of string, or end of line in multi-line pattern
\Z	End of string
\b	Word boundary
\B	Not word boundary
\<	Start of word
\>	End of word

Character Classes	
\c	Control character
\s	White space
\S	Not white space
\d	Digit
\D	Not digit
\w	Word
\W	Not word
\x	Hexadecimal digit
\O	Octal digit

POSIX	
[:upper:]	Upper case letters
[:lower:]	Lower case letters
[:alpha:]	All letters
[:alnum:]	Digits and letters
[:digit:]	Digits
[:xdigit:]	Hexadecimal digits
[:punct:]	Punctuation
[:blank:]	Space and tab
[:space:]	Blank characters
[:cntrl:]	Control characters
[:graph:]	Printed characters
[:print:]	Printed characters and spaces
[:word:]	Digits, letters and underscore

Assertions	
?=	Lookahead assertion
?!	Negative lookahead
?<=	Lookbehind assertion
?!= or ?<!	Negative lookbehind
?>	Once-only Subexpression
?()	Condition [if then]
?()|	Condition [if then else]
?#	Comment

Quantifiers			
*	0 or more	{3}	Exactly 3
+	1 or more	{3,}	3 or more
?	0 or 1	{3,5}	3, 4 or 5

Add a ? to a quantifier to make it ungreedy.

Escape Sequences	
\	Escape following character
\Q	Begin literal sequence
\E	End literal sequence

"Escaping" is a way of treating characters which have a special meaning in regular expressions literally, rather than as special characters.

Common Metacharacters			
^	[	.	$
{	*	(	\
+	)	|	?
<	>		

The escape character is usually \

Special Characters	
\n	New line
\r	Carriage return
\t	Tab
\v	Vertical tab
\f	Form feed
\xxx	Octal character xxx
\xhh	Hex character hh

Groups and Ranges	
.	Any character except new line (\n)
(a|b)	a or b
(...)	Group
(?:...)	Passive (non-capturing) group
[abc]	Range (a or b or c)
[^abc]	Not (a or b or c)
[a-q]	Lower case letter from a to q
[A-Q]	Upper case letter from A to Q
[0-7]	Digit from 0 to 7
\x	Group/subpattern number "x"

Ranges are inclusive.

Pattern Modifiers	
g	Global match
i *	Case-insensitive
m *	Multiple lines
s *	Treat string as single line
x *	Allow comments and whitespace in pattern
e *	Evaluate replacement
U *	Ungreedy pattern

* PCRE modifier

String Replacement	
$n	nth non-passive group
$2	"xyz" in /^(abc(xyz))$/
$1	"xyz" in /^(?:abc)(xyz)$/
$`	Before matched string
$'	After matched string
$+	Last matched string
$&	Entire matched string

Some regex implementations use \ instead of $.

Anexo B - ODS

OBJETIVOS DE DESARROLLO SOSTENIBLE

Grado de relación del trabajo con los Objetivos de Desarrollo Sostenible (ODS).

Objetivos de Desarrollo Sostenibles	**Alto**	**Medio**	**Bajo**	**No Procede**
ODS 1. **Fin de la pobreza.**			X	
ODS 2. **Hambre cero.**			X	
ODS 3. **Salud y bienestar.**			X	
ODS 4. **Educación de calidad.**			X	
ODS 5. **Igualdad de género.**				
ODS 6. **Agua limpia y saneamiento.**			X	
ODS 7. **Energía asequible y no contaminante.**			X	
ODS 8. **Trabajo decente y crecimiento económico.**	X			
ODS 9. **Industria, innovación e infraestructuras.**	X			
ODS 10. **Reducción de las desigualdades**.			X	
ODS 11. **Ciudades y comunidades sostenibles**.			X	
ODS 12. **Producción y consumo responsables.**	X			
ODS 13. **Acción por el clima.**			X	
ODS 14. **Vida submarina.**			X	
ODS 15. **Vida de ecosistemas terrestres.**			X	
ODS 16. **Paz, justicia e instituciones sólidas.**			X	
ODS 17. **Alianzas para lograr objetivos.**		X		

Reflexión sobre la relación del TFG/TFM con los ODS y con el/los ODS más relacionados.

En este anexo analizaremos la implicación que tienen los objetivos de desarrollo sostenible en el tema que desarrollamos en el TFG, centrándonos más en aquellos que tienen un mayor impacto en el mismo.

Los ODS consisten en un conjunto de objetivos globales llevados a cabo a través de medidas para tratar de acabar con la pobreza, la desigualdad, proteger el planeta y asegurar prosperidad y crecimiento. En 2015, la ONU aprobó la Agenda 2030 sobre el desarrollo sostenible, mediante la cual los estados miembros se comprometen a llevar a cabo acciones para transformar el mundo a través de 17 objetivos de desarrollo sostenible.

En este mundo globalizado, las fuentes de información digitales tienen una presencia fundamental en la toma de decisiones de cualquier empresa. De este modo, una de las principales ventajas competitivas de las organizaciones es tanto poseer dicha información, como también que la misma sea fiable. Así pues, poseer un software que almacene y verifique que dicha información es completamente veraz es una herramienta básica para la toma de decisiones de cualquier empresa.

De este modo, cumpliendo con la **ODS 8**, herramientas de plataforma informática como AXON Informática, EDC o IDQ permiten que las empresas obtengan y almacenen una información más fiable y con menores errores con el objetivo de tomar decisiones de forma adecuada y así garantizar el crecimiento económico de la organización.

En segundo lugar, cabe destacar que el éxito para el crecimiento económico comentado en punto anterior para cualquier empresa vigente en un mercado donde cada vez más la competencia es mayor, tiene una de sus bases fundamentales en la innovación. Por tanto, y cumpliendo con la **ODS 9**, hay que destacar que los softwares tratados en el trabajo de fin de grado permiten que dicha innovación se haga de una forma honesta gracias al análisis de datos que se desprenden del análisis de la información obtenida.

Por otra parte, cabe destacar también la **ODS 12**, que consiste en la producción y consumo responsable. Mediante los softwares citados anteriormente, las empresas garantizan un almacenamiento de datos fiable, clave para una toma de decisiones sin desperdicio de recursos. Así pues, las herramientas de gobierno de datos permiten crear estándares de actuación y general reglas para futuras implementaciones.

Por último, hay que considerar que, en muchas ocasiones para el crecimiento económico de una empresa y el beneficio social general, son necesarias las alianzas con el fin de obtener objetivos comunes, conforme a la **ODS 17**, las herramientas de gobierno de datos permiten almacenar y filtrar la información recopilada por las organizaciones con el objetivo de general alianzas que beneficien tanto a las mismas organizaciones como toma de decisiones y alianzas que sean de beneficio social común.

En conclusión, queda patente la importancia de las herramientas de Gobierno de datos. Pues es de vital importancia para las empresas obtener información que puedan almacenar de forma digital, que puedan consultar y analizar tantas veces sea necesarias y que de igual manera garantice transparencia y fidelidad con la realidad, para un análisis y toma de decisiones correctos. Así pues, también mediante las herramientas de Gobierno de datos se permiten crear reglas de actuación que disminuyen en tiempo de actuación frente a tomas de decisiones, lo que supone un ahorro de tiempo en la puesta en marcha de acciones a tomar por parte de las empresas.