

**PROYECTO DE GESTIÓN DEL DATO PARA LA  
PRODUCCIÓN CIENTÍFICA DEL PERSONAL DOCENTE E  
INVESTIGADOR DE LA UNIVERSIDADE DA CORUÑA**

**PROXECTO DE XESTIÓN DO DATO PARA A PRODUCCIÓN  
CIENTÍFICA DO PERSOAL DOCENTE E INVESTIGADOR DA  
UNIVERSIDADE DA CORUÑA**

**DATA MANAGEMENT PROJECT FOR THE SCIENTIFIC  
PRODUCTION OF THE TEACHING RESEARCH STAFF OF THE  
UNIVERSITY OF CORUÑA**

Alumna: Miriam Louro Ríos

Director: Miguel Ángel Rodríguez Luaces

Grado: Gestión Digital de Información y Documentación

Año académico: 2022/2023

Convocatoria: junio 2023

## **Agradecimientos**

Me gustaría mostrar mi agradecimiento, en primer lugar, a mi tutor, por guiarme y trasmitirme todos los conocimientos necesarios durante estos cuatro años.

En segundo lugar, a mi familia, por el apoyo incondicional, la motivación y la paciencia en todas las etapas de mi vida.

Por último, a mis amigos, en especial al grupo que hemos formado en estos breves pero intensos años de carrera, ya que, sin ellos, no habría sido lo mismo.

## Resumen

La transformación digital y el avance de las nuevas tecnologías han provocado un aumento exponencial en la creación de datos digitales, convirtiéndose, los datos, en el principal activo de las organizaciones. Por ello, resulta esencial el gobierno, gestión y calidad de los datos. Así, el objetivo principal es definir un proyecto de gestión de datos para la producción científica del personal docente e investigador de la *Universidade da Coruña*. A través de una metodología de investigación-acción, en un primer ciclo de investigación, se estudia la situación actual de la Universidad, el sistema gestor de producción científica y las principales plataformas en donde el personal docente e investigador publica sus trabajos. En un segundo ciclo de investigación se identifican los datos, a través de tablas, que se han de incorporar en el proyecto, los cuales se han relacionado a través de un modelo conceptual, representado en un diagrama entidad-relación. A continuación, se analizan las herramientas tecnológicas del gobierno del dato tanto individualmente como en conjunto, junto con los beneficios que supondría su implementación. Ambos ciclos van acompañados de dos tareas previas de identificación del problema y revisión bibliográfica y una tarea final de redacción del proyecto. Así, se crean los primeros pasos para la elaboración de un plan de gestión del dato para la producción científica en la *Universidade da Coruña*.

**Palabras clave:** Gobierno del dato, Gestión del dato, Calidad del dato, Producción científica, Universidade da Coruña, personal docente e investigador, proyecto de gestión de datos.

## Resumo

A transformación dixital e o avance das novas tecnoloxías provocaron un importante aumento na creación de datos dixitais. Converténdose, os datos, no principal activo das organizacións, por iso, resulta esencial o goberno, xestión e calidade dos datos. Así o obxectivo principal é definir un proxecto de xestión de datos para a produción científica do persoal docente e investigador da Universidade da Coruña. A través dunha metodoloxía de investigación-acción estúdase a situación actual da Universidade mediante o sistema xestor da produción científica e as principais plataformas onde o persoal docente e investigador publica os seus traballos. En un segundo ciclo de investigación identifícanse os datos, por medio de táboas, que se teñen que incorporar no proxecto, os cales se relacionan a través dun modelo conceptual, representado nun diagrama entidade-relación. A continuación, analízanse as ferramentas tecnolóxicas do goberno do dato tanto individualmente como en conxunto, xunto cos beneficios que supoñería a súa posta en funcionamento. Ambos ciclos van acompañados de dúas tarefas previas de identificación do problema e revisión bibliográfica e una tarefa final de redacción do proxecto. Así, créanse os primeiros pasos para a elaboración dun plan de xestión do dato para a produción científica na Universidade da Coruña.

**Palabras clave:** Goberno do dato, Xestión do dato, Calidade do dato, Produción científica, Universidade da Coruña, persoal docente e investigador, proxecto de xestión de datos.

**Abstract**

Digital transformation and the advance of new technologies have led to an exponential increase in the creation of digital data. Becoming data, the main asset of organizations, therefore, it is essential to govern, manage and quality data. Thus, the main objective is defining a data management project for the scientific production of the teaching and research staff of the University of Coruña. Through an action-research methodology, in a first research cycle, the current situation of the University, the scientific production management system and the main platforms where teaching and research staff publish their work are studied. In a second research cycle, the data to be incorporated into the project are identified through tables, which have been related through a conceptual model, represented in an entity-relationship diagram. Next, the technological tools of data governance are analysed both individually and as a whole, along with the benefits of their implementation. Both cycles are accompanied by two previous tasks of problem identification and literature review and a final task of project writing. Thus, the first steps for the development of a data management plan for scientific production at the University of Coruña are created.

**Keywords:** Data Governance, Data Management, Data Quality, Scientific production, University of Coruña, teaching research staff, data management model.



## Índice de contenido

1. Introducción.....	1
2. Objetivos.....	6
3. Metodología.....	8
4. Gobierno, Gestión y Calidad del dato .....	12
4.1. Gobierno del dato.....	12
4.2. Gestión del dato .....	13
4.3. Calidad del dato .....	14
4.4. Componentes comunes .....	17
4.5. Equipo de trabajo .....	19
4.6. Diseño y actividades de proyectos de gobierno, gestión y calidad del dato ....	20
5. Propuesta de proyecto de gestión del dato para la producción científica del personal docente e investigador de la Universidade da Coruña.....	23
5.1. Situación actual.....	23
5.2. Alcance de la propuesta .....	24
5.3. Estructuras organizativas .....	25
5.4. Identificación de plataformas de extracción de información.....	29
5.5. Identificación de datos a incorporar en el proyecto .....	33
5.6. Modelo conceptual.....	37
5.7. Requisitos de los datos.....	39
5.8. Análisis comparativo de herramientas tecnológicas.....	41
5.9. Beneficios de la implementación .....	45
6. Conclusiones.....	47
7. Referencias .....	50
8. Anexo .....	54
8.1. Glosario de términos .....	54



## Índice de tablas

Tabla 1. Tabla de principios con sus respectivas definiciones.....	18
Tabla 2. Tabla de cargos y roles del equipo de trabajo del proyecto. Elaboración propia. .....	27
Tabla 3. Relación entre puestos y roles representados en la pirámide (Ilustración 4). Elaboración propia.....	27
Tabla 4. Composición de la comisión del gobierno del dato. Elaboración propia. ....	28
Tabla 5. Composición de la comisión de gestión del dato. Elaboración propia. ....	28
Tabla 6. Composición de la comisión de calidad del dato. Elaboración propia. ....	29
Tabla 7. Tabla de recolección de datos de los identificadores de los investigadores. Elaboración propia.....	35
Tabla 8. Tabla de recolección de datos de las publicaciones. Elaboración propia. ....	35
Tabla 9. Tabla de recolección de datos de los artículos de revista. Elaboración propia.	36
Tabla 10. Tabla de recolección de datos de las revistas. Elaboración propia. ....	36
Tabla 11. Tabla de recolección de datos de los indicadores de revistas. Elaboración propia. .....	36
Tabla 12. Tabla de recolección de datos de los libros. Elaboración propia. ....	37
Tabla 13. Tabla de recolección de datos de actas de congreso. Elaboración propia.....	37
Tabla 14. Tabla de recolección de datos de capítulos de libros. Elaboración propia.....	37
Tabla 15. Tabla análisis de la herramienta Microsoft Purview. Elaboración propia. ....	42
Tabla 16. Tabla análisis de la herramienta Talend. Elaboración propia. ....	42
Tabla 17. Tabla análisis de la herramienta Apache Atlas. Elaboración propia. ....	43
Tabla 18. Tabla análisis de la herramienta Atlan. Elaboración propia.....	43
Tabla 19. Tabla análisis de la herramienta Tableau. Elaboración propia.....	44
Tabla 20. Tabla de pros de las herramientas tecnológicas analizadas anteriormente. Elaboración propia.....	45
Tabla 21. Tabla de contras de las herramientas tecnológicas analizadas anteriormente. Elaboración propia.....	45
Tabla 22. Glosario de términos.....	56

## Índice de ilustraciones

Ilustración 1. Imagen extraída de la Especificación UNE0078:2023.....	2
Ilustración 2. Metodología investigación-acción. Elaboración propia.....	9
Ilustración 3. Ilustración extraída de la Especificación UNE 077:2023 2023, 12.....	17
Ilustración 4. Pirámide de roles en proyectos de Gobernanza del dato extraída de la página web de Keyrus.....	19
Ilustración 5. Ilustración extraída de la Especificación UNE 0077:2023 sobre la implementación de un sistema de gobierno del dato.....	21
Ilustración 6. Gráfica representativa de las publicaciones por áreas temáticas de los afiliados de la UDC a Scopus. Imagen extraída de la página web de Scopus.....	30
Ilustración 7. Gráfica del número de afiliados a Dialnet de la UDC. Imagen extraída de Dialnet métricas.....	31
Ilustración 8. Gráfico correspondiente al número de autores afiliados a la UDC por materia. Gráfica de elaboración propia en base a la representada en Dialnet Métricas.	31
Ilustración 9. Gráfica representativa del número de publicaciones de los afiliados a la UDC por áreas temáticas en la Web Of Science. Gráfica de elaboración propia en base a la extraída de la Web Of Science.....	32
Ilustración 10. Modelo conceptual del proyecto. Elaboración propia.....	38



## 1. Introducción

Actualmente, vivimos en la denominada “Era digital”, en una sociedad conectada, en la que las nuevas tecnologías, las TIC (Tecnologías de Información y Comunicación), han experimentado un importante avance en todos los ámbitos y en el día a día tanto de las personas como de las organizaciones y empresas. Por ello, cada vez hay más información a disposición de cualquier persona, lo cual implica, por un lado, ciertos problemas como la *infoxicación*<sup>1</sup>, una mayor facilidad de pérdida de datos, una menor capacidad para reconocer datos reales, etc. Por otro lado, también conlleva una gran cantidad de ventajas, como el acceso más fácil a la información, la comunicación inmediata, mejor aprovechamiento de los recursos, el aumento en la facilidad de la colaboración entre personas, etc.

No obstante, el avance de las nuevas tecnologías, las redes sociales, la nube o Internet de las Cosas (IoT)<sup>2</sup> es decir, la denominada transformación digital, han provocado un aumento exponencial en la creación de datos, acumulándose así a una velocidad sin precedentes en todos los sectores (Caballero, Rodríguez y Fernández, 2017, 36). En consecuencia, hoy en día, resulta habitual escuchar la afirmación de que el mundo se está *datificando* (Rodríguez, Gómez y Delgado, 2018, 28). Así como también se han popularizado otras expresiones como: “los datos son la nueva moneda”, “los datos son el nuevo petróleo” o “los datos son la mina oculta” (Piattini, Rodríguez, y Caballero y Fernández, 2018, 5). En consecuencia, la capacidad de recopilar, almacenar y utilizar datos de manera efectiva se ha vuelto crucial para mejorar, sobre todo, en la toma de decisiones de las organizaciones en todos los ámbitos, ya sea en el sector público o privado.

Ante este panorama, los datos, en los últimos 10 años, han experimentado un importante auge, lo cual ha supuesto una acumulación de estos, convirtiéndose, de este modo, en un problema a tratar. En consecuencia, surge el *Big Data*, entendiéndolo como “combinaciones de conjuntos de datos cuyo tamaño (volumen), complejidad (variabilidad) y velocidad de crecimiento (velocidad) dificultan su captura, gestión,

---

<sup>1</sup> Denominada en inglés *overload information*. Hace referencia a la sobrecarga de información y sus efectos negativos que tiene lugar en un entorno digital (Reyero y Gil 2021, 104).

<sup>2</sup> Es la red de objetos físicos que llevan incorporados sensores, software y otras tecnologías, con el objetivo de conectarse e intercambiar datos con otros dispositivos y sistemas a través de Internet (Oracle, 2023).

procesamiento o análisis mediante tecnologías y herramientas convencionales, tales como bases de datos relacionales y estadísticas” (PowerData, 2023). Por ende, los datos se han convertido en un activo fundamental y principal, sobre todo para las organizaciones, quiénes, hoy en día, se están involucrando en proyectos encargados de poner los datos en el eje central de los mismos. Al convertirse los datos en el activo digital más valiosos en la actualidad, el gobierno, la gestión y la calidad de los datos, en consecuencia, han adquirido una elevada relevancia, funcionando las tres como engranajes complementarios entre sí (Ilustración 1) con el objetivo de extraer el máximo beneficio a la hora de explotar los datos (Especificación UNE0078:2023, 2023, 6).



Ilustración 1. Imagen extraída de la Especificación UNE0078:2023.

El gobierno del dato, también denominado en inglés *data governance*, es el conjunto de políticas, procesos y estructuras que se implementan en una organización para garantizar la disponibilidad, integridad y confidencialidad de los datos, es decir, para poder desarrollar las tareas de gestión y calidad del dato. Por otro lado, la gestión del dato, denominado en inglés como *data management*, tiene como finalidad implementar mecanismos de monitorización del nivel de calidad del dato, reportar al gobierno del dato los valores correspondientes y proporcionar las herramientas necesarias con el objetivo de satisfacer los requisitos planteados en la calidad del dato. (Rodríguez, Gómez y Delgado, 2018, 28-29). Por último, ligado a los anteriores, se encuentra la calidad del dato, denominada en inglés *data quality*, la cual se define como “el grado en el que las

características de los datos satisfacen las necesidades declaradas e implícitas cuando se utilizan en condiciones específicas” (Especificación UNE 0079:2023, 2023, 37).

En resumen, el gobierno, la gestión y la calidad del dato son elementos fundamentales en la era de la información, de modo que, su implementación resulta esencial para garantizar el éxito y el uso tanto eficaz como eficiente de los datos dentro de las organizaciones. No obstante, la falta de estos puede suponer una serie de problemas que van desde la desorganización y falta de fiabilidad de los datos, hasta la aparición de riesgos con relación a su seguridad.

Un ámbito en donde la gobernanza de datos adquiere un rol muy relevante es en el universitario, más concretamente en el relacionado con la producción científica del personal docente e investigador<sup>3</sup>, entendiendo la producción científica como “el conocimiento generado por un grupo de personas con la finalidad de dar a conocer información general de un tema de estudio” (Flores, Meléndez y Mendoza, 2019, 65). En consecuencia, el objetivo de la producción científica es transmitir conocimiento a los usuarios, específicamente dirigida hacia la comunidad universitaria o expertos en la materia, mediante el uso de jergas y lenguajes precisos. No obstante, es importante resaltar que sus destinatarios pueden ser cualquier tipo de público.

La producción científica generada por los docentes, además de servir como fuente de conocimiento, también se utiliza, para medir la calidad investigadora de las universidades, ya que, a partir de la producción científica, se realizan estudios aritméticos según ciertos indicadores estipulados dentro de la bibliometría<sup>4</sup> (Cortés, 2007, 44), que se pueden dividir en dos, por un lado, los que describen a los autores (Índice H, Índice de productividad, Índice G, Índice contemporáneo, Índice de impacto, etc.) y, por otro lado, los que describen las revistas científicas (Factor de impacto, Cuartil, *Scimago Journal Rank*, etc.). A partir de los datos que se obtienen, la producción de los investigadores de una universidad se posiciona en un ranking. Así, cuánto más arriba se sitúe, la puntuación será más alta, lo cual implica un aumento en la reputación para la universidad en cuestión, mientras que, cuánto más abajo sea la posición en el ranking, peor puntuación va a tener el investigador y, por tanto, la universidad también se va a ver afectada.

---

<sup>3</sup> PDI.

<sup>4</sup> Su objetivo es la medición de los elementos relativos a la productividad o cantidad de documentos generados por los investigadores universitarios.

Sin embargo, la gestión de la producción científica de las universidades es un problema que no siempre está correctamente resuelto debido a una serie de dificultades. La primera de ellas es que el PDI publica sus investigaciones en diferentes plataformas, tales como *Google Scholar*<sup>5</sup>, *Scopus*<sup>6</sup>, *Web of Science*<sup>7</sup>, *Dialnet*<sup>8</sup> e incluso en los sitios web creados por las propias universidades. Por lo que, muchas veces, la recuperación de toda la información puede ser engorrosa, inexacta y costosa cuando se habla de tiempo, debido al análisis que se ha de realizar sobre los resultados de cada una de las plataformas anteriormente mencionadas. Puesto que, para obtener los datos completos y actualizados, se ha de comprobar que no haya ninguna publicación repetida, así como se ha de atender a si dichas publicaciones presentan algún tipo de error, por ejemplo, en el año o número de páginas que puede ser diferente dependiendo del sitio en donde esté publicado.

La segunda de las dificultades es la incoherencia entre las fuentes de datos, por ejemplo, con la redacción de los nombres de los investigadores, puesto que pueden ser escritos de diversas formas, debido a que, a veces, si se trata de un nombre compuesto, puede aparecer con un guion o no. En ciertas ocasiones, se omite el segundo nombre o uno de los apellidos o también se utilizan únicamente las iniciales del nombre. Además, esto también sucede con la forma de denominar las publicaciones y, en consecuencia, resulta dificultoso, además de suponer, como bien se ha indicado, una pérdida de tiempo a la hora de buscar todas las posibilidades que existen de escribir un nombre, la tipología de la publicación, un título, etc.

Adicionalmente, se podría hablar de una tercera dificultad relacionada con la inconsistencia de los propios investigadores, puesto que, ellos mismos firman sus trabajos, suben publicaciones y crean perfiles en diversas redes sociales académicas cambiando su nombre, añadiendo iniciales, guiones o poniendo el nombre completo. Además, también sucede con la nomenclatura de la afiliación, la cual escriben de diversas formas sin seguir un formato estandarizado.

Por ello, teniendo en cuenta estas tres dificultades, en este trabajo se propone el marco teórico para un sistema de gestión de datos sobre la producción científica del

---

<sup>5</sup> Página de acceso a *Google Scholar*: <https://scholar.google.es/>

<sup>6</sup> Página de acceso a *Scopus*: <https://www.scopus.com/home.uri>

<sup>7</sup> Página de acceso a *Web of Science*: <https://www.webofscience.com/wos/woscc/basic-search>

<sup>8</sup> Página de acceso a *Dialnet*: <https://dialnet.unirioja.es/>

personal docente e investigador de la *Universidade da Coruña*, con el objetivo final de modelar los primeros pasos en la implementación de un modelo de gestión del dato. Para que, en un futuro, se pueda recoger en un único lugar, coordinado por un equipo de trabajo conformado por miembros de la universidad, todos los datos correspondientes a la producción científica con la finalidad de mantener los datos actualizados, completos y seguros.

Para conseguirlo, en primer lugar, se realiza una aproximación teórica al gobierno del dato, gestión del dato y calidad del dato, junto con sus correspondientes componentes, a través de la recopilación de información derivada de la realización de una revisión bibliográfica previa. En segundo lugar, se presentan las primeras actividades a seguir para implementar un proyecto de esta índole dentro de la universidad. Finalmente, se presentan las conclusiones, acompañadas de las lecciones aprendidas extraídas de la elaboración del trabajo y cuáles serían los futuros trabajos más próximos que pueden dar continuidad a esta primera práctica.

## 2. Objetivos

El principal objetivo que se ha determinado para este trabajo es el de elaborar un proyecto de gestión del dato para la producción científica del personal docente e investigador de la *Universidade da Coruña*. Este objetivo principal se divide en los siguientes objetivos específicos:

- Analizar fuentes de datos de producción científica para identificar cuáles son los lugares en donde los investigadores suben o publican más contenido, en este caso, en entornos digitales.
- Definir un modelo de datos para almacenar la producción científica del personal docente e investigador de la *Universidade da Coruña* para tener todos los datos recogidos y ordenados en un mismo lugar.
- Elaborar un plan de gobierno, gestión y calidad del dato con la finalidad de que los datos de la producción científica del PDI de la *Universidade da Coruña* se mantengan actualizados y sean reales, confiables y útiles, para que, a la hora de obtener información sobre su actividad científica, por un lado, no existan los problemas mencionados anteriormente y, por otro lado, para proporcionar que la información sea veraz, real, actualizada y completa.

Además de los objetivos relacionados con la propia elaboración del proyecto, también se han desarrollado objetivos personales que se pretenden alcanzar tras la elaboración de este como:

- Analizar la situación actual de la *Universidade da Coruña* para conocer de qué manera y cómo se están tratando, actualizando y poniendo a disposición los datos relativos a la producción científica del personal docente e investigador.
- Conocer y analizar herramientas tecnológicas relacionadas con el gobierno del dato para tener una visión general del panorama actual, dando a conocer las funcionalidades de las herramientas tanto de pago como gratuitas.
- Profundizar sobre temas de actualidad como son, en este caso, los datos, ya que, como bien se ha señalado en la introducción, las nuevas tecnologías han ayudado a aumentar la creación de datos dentro de cualquier organización, en consecuencia, es fundamental conocerlos y aprender a tratarlos para poder sacarles el máximo beneficio.

Finalmente, en caso de que los resultados de este Trabajo de Fin de Grado se pudieran aplicar sobre la *Universidade da Coruña*, se propondrían una serie de objetivos a alcanzar, más específicos, como son:

- Dar solución a un problema que se encuentra, en este caso, en la *Universidade da Coruña* que hace referencia, por una parte, a la manera en la que se gestionan los datos de la producción científica y, por otra parte, a la actualización de los datos sobre la producción científica del PDI, puesto que, la mayoría de ellos no tiene ni actualizada ni completa su producción científica.
- Romper con los silos de datos en una organización, más concretamente en la *Universidade da Coruña*, para evitar la aparición de ciertos problemas como recursos perdidos o aislamiento de la información.
- Crear unas pautas para el desarrollo de un proyecto de gobierno del dato real, para que, en un futuro, miembros de otras organizaciones las puedan utilizar para desarrollar la propuesta aquí mencionada.

### 3. Metodología

Para la realización de este trabajo se ha seguido la metodología investigación-acción, cuya característica principal es que “*conducts research and takes action at the same time*” (George, 2023), es decir, investigar y resolver un problema simultáneamente, siendo “apropiada para aquellos que realizan investigaciones en pequeña escala, preferentemente en áreas de educación, salud, asistencia social e incluso administración” (ARR, 2023). Cabe destacar que la investigación-acción ha desarrollado dos tendencias fundamentales: una sociológica y otra educativa (Colmenares y Piñeiro, 2008, 100), siendo esta última, la más conocida.

La metodología de investigación-acción se puede dividir en dos tipos principales: la participativa, dentro de la cual los participantes han de ser los miembros de la comunidad que se estudia, de modo que los participantes son coinvestigadores y sus experiencias resultan formativas para el trabajo, y la práctica, la cual se enfoca en la forma en la que se ha de llevar a cabo la investigación, por lo que está diseñada para abordar y resolver un problema determinado. Así, la metodología empleada a lo largo de este trabajo ha sido la investigación-acción práctica.

Con todo, la investigación-acción se aplica con el objetivo de mejorar prácticas específicas a partir de la evaluación, y análisis crítico de los datos recopilados. Además, se ve facilitada por la participación y colaboración entre diferentes personas, enfocándose en determinadas situaciones, teniendo en cuenta su contexto. Por tanto, se pueden extraer de este tipo de metodología una serie de ventajas, recogidas por Dudovskiy en el año 2023: “*high level of practical relevance of the business research*”<sup>9</sup>, “*can be used with quantitative, as well as qualitative data*”<sup>10</sup> y “*possibility to gain in-depth knowledge about the problem*”<sup>11</sup>.

Para la elaboración del proyecto aquí presente, por un lado, se planificaron una serie de tareas: la identificación del problema, la revisión bibliográfica y la redacción de la memoria y, por otro lado, también se planificaron dos ciclos de investigación-acción: el estudio de la situación actual de la *Universidade da Coruña* con respecto a la producción científica del personal docente e investigador y la elaboración del plan de

---

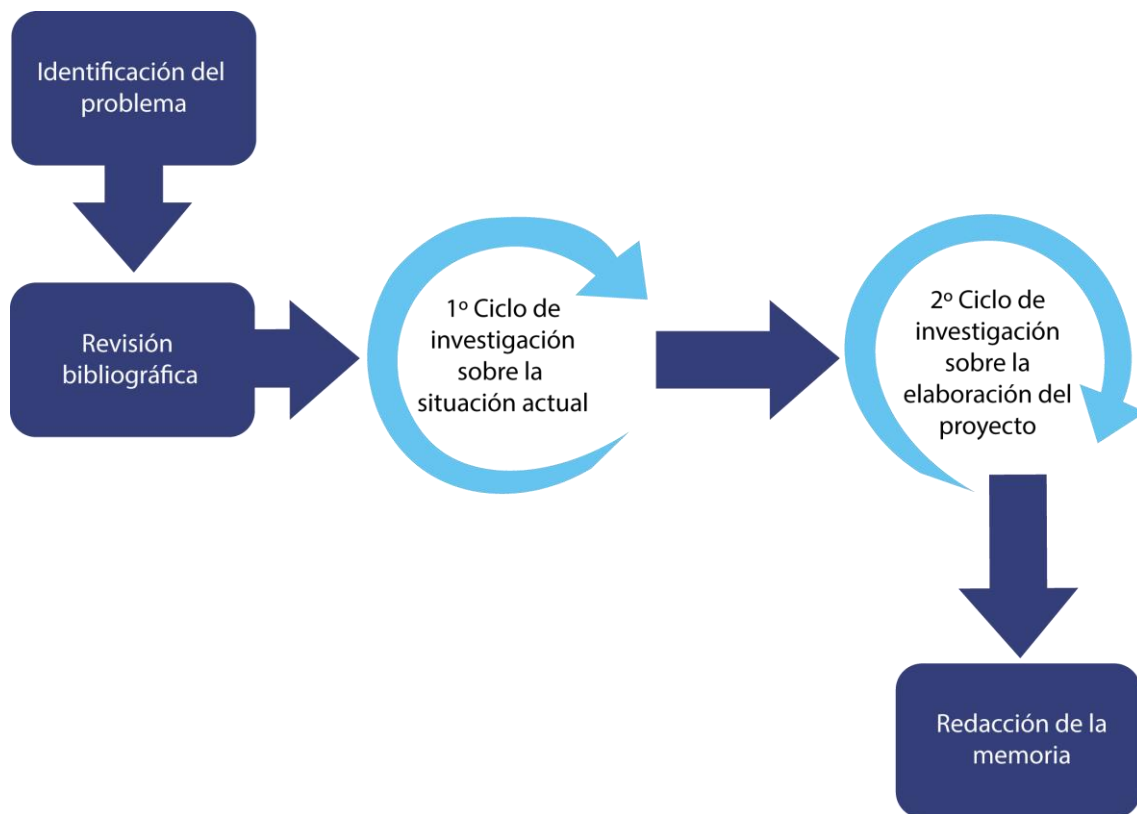
<sup>9</sup> Traducción: alto nivel de relevancia práctica de la investigación empresarial.

<sup>10</sup> Traducción: se puede utilizar con datos cuantitativos y cualitativos.

<sup>11</sup> Traducción: posibilidad de profundizar en el conocimiento del problema.



gobierno, gestión y calidad del dato para dar respuesta al problema planteado. Estas tareas y ciclos de investigación se encuentran recogidos en un diagrama de elaboración propia, en donde las primeras se representan mediante rectángulos, mientras que los segundos, se representan a través de círculos. Además, la dirección de las flechas indica el orden que se ha de seguir (Ilustración 2).



*Ilustración 2. Metodología investigación-acción. Elaboración propia.*

La primera tarea es la identificación del problema al que se le quiere dar respuesta con la elaboración de este trabajo. Por ello se analizan las dificultades a la hora de gestionar la producción científica del personal docente e investigador dentro de la *Universidade da Coruña*.

A continuación, la segunda tarea a abordar es la realización de una revisión bibliográfica de la cual se extrae la documentación relacionada con el gobierno, gestión y calidad del dato. A partir de aquí, se han de seleccionar los autores más destacados en dicho ámbito, lo cual sirve para seguir indagando sobre sus investigaciones, proyectos y modelos. Posteriormente se crea la primera parte del trabajo, en donde se recogen los aspectos teóricos del gobierno, gestión y calidad del dato, acompañado de sus actividades,

acciones, principios comunes, equipo de trabajo y actividades comunes que tiene todos los proyectos de esta índole.

Seguidamente, se realiza el primero de los ciclos de investigación-acción, en el cual desarrolla un estudio de la situación actual existente en la UDC. En este ciclo se ha de llevar a cabo una pequeña investigación para determinar si, dentro de la universidad, hay algún sistema que gestione los datos de la producción científica del personal docente e investigador de forma eficaz y eficiente. Por tanto, en este primer ciclo de investigación se estudia la plataforma SUXI en donde el PDI introduce sus investigaciones y datos curriculares. De modo que se analizan las problemáticas, las carencias y el funcionamiento de dicha herramienta

A continuación, determina el alcance del proyecto. En segundo lugar, se lleva a cabo un análisis sobre las plataformas en donde el PDI tiene publicados sus trabajos, acompañado de gráficos representativos del número de publicaciones en cada una de las plataformas para conocer, de este modo, la importancia de la UDC en cada una de ellas. Además, también se investiga sobre la forma de extraer datos para, posteriormente, introducirlos en la plataforma de gobierno del dato seleccionada.

Atendiendo al segundo ciclo de investigación, se analiza la estructura organizativa de la propia universidad, con el objetivo de encontrar cargos que pudiesen estar dentro del equipo de trabajo para el proyecto. No obstante, en este punto, se han de proponer una serie de comisiones que se han de crear junto con los cargos y personas que han de ocupar los puestos en cada una de ellas, tanto de nueva incorporación como los ya existentes en la universidad.

A continuación, se selecciona la información y los datos que se incorporarían en la herramienta tecnológica correspondiente. Para este fin, se crean tablas que representan distintas tipologías de datos. A partir de estas tablas, se genera un modelo relacional con el objetivo de unificar todos los datos y facilitar la realización de las consultas deseadas.

Para finalizar con este segundo ciclo de investigación, se seleccionan cinco herramientas para el gobierno, gestión y calidad del dato, buscando información sobre las características de las mismas recogidas en tablas independientes. Para poder realizar una comparativa entre todas ellas, se plasma, en una misma tabla, tanto los pros como los contras de cada una de las herramientas. Por último, se determinan una serie de beneficios

que supondría la implementación de un proyecto de gestión del dato para la producción científica del personal docente e investigador de la *Universidade da Coruña*.

Para acabar con la metodología de elaboración del trabajo, se lleva a cabo una última tarea que consiste en la propia redacción del Trabajo de Fin de Grado, teniendo en cuenta todo lo expuesto hasta el momento.

Finalmente, es importante señalar que el método empleado tanto para las citas como para las referencias bibliográficas elaboradas al final de dicho documento, se han hecho según el sistema Chicago, más concretamente, siguiendo el sistema autor-fecha, elaborado por la Universidad Autónoma de Madrid<sup>12</sup> (UAM, 2023).

---

<sup>12</sup> UAM.

## 4. Gobierno, Gestión y Calidad del dato

### 4.1. Gobierno del dato

El Gobierno del Dato es “una función organizadora responsable de rendir cuentas sobre el uso eficaz, eficiente y aceptable de los datos por parte de la organización” (Especificación UNE 0077:2023, 2023, 9), cuyo objetivo principal es maximizar el valor de los datos y gestionar los riesgos asociados mediante la elaboración y comprobación de una serie de objetivos estratégicos (Rodríguez, Gómez y Delgado, 2018, 28).

Por tanto, se puede afirmar que el gobierno del dato se refiere a la fase teórica primordial en cualquier proyecto de este tipo. Durante esta etapa se desarrollan los procesos a seguir, se establecen políticas a cumplir y se llevan a cabo diversas acciones con el propósito de implementarlas en las etapas posteriores, específicamente en la gestión de los datos.

En consecuencia, el gobierno del dato persigue “maximizar el valor de los datos de la organización y gestionar los riesgos asociados, mediante la elaboración y comprobación de una serie de objetivos estratégicos” (Gómez y Piattini, 2018, 7). Así mismo, cuando se habla del gobierno del dato se vuelve necesario diferenciar entre tres tipos:

- **Centralizado:** en este tipo de gobierno, todas las decisiones relacionadas con los datos se toman desde un punto de vista centralizado. Cabe destacar que, esta tipología está dirigida, sobre todo, hacia las grandes organizaciones.
- **Descentralizado:** aquí las decisiones relacionadas con los datos se delegan tanto a los departamentos como a los equipos individuales que componen la organización. A diferencia del anterior, este tipo de gobierno se suele utilizar en organizaciones más pequeñas.
- **Híbrido:** como bien indica su nombre, dicha tipología es una mezcla de los dos enfoques anteriores, por lo que, en este caso, se pueden tomar decisiones tanto centralizadas como descentralizadas o ambas combinadas.

Además, en el gobierno del dato es importante llevar a cabo una serie de actividades elaboradas tanto por los órganos del gobierno del dato como por los directores de los sistemas de gestión de información. La primera de ellas es la evaluación, en donde se identifica el uso actual de los datos, es decir, la situación de la que parten los datos

dentro de una organización. Por tanto, se analiza su ubicación, las características que cumplen, las normativas que se le aplican, si se le aplica alguna, la calidad que estos tienen, así como también se analiza su actividad en el futuro. En segundo lugar, se encuentra la actividad relacionada con la dirección, en donde se preparan e implementan tanto estrategias como políticas de datos, cuyo objetivo principal es asegurar que estos, por un lado, satisfagan los objetivos planteados con anterioridad, y, por otro lado, que sean fiables y reales. Por último, se lleva a cabo la monitorización, la cual se encuentra relacionada con la anterior, ya que es aquí en donde se pone en marcha lo planteado hasta el momento, es decir, tanto las políticas como las estrategias definidas.

Finalmente, tras definir el gobierno del dato y analizar sus actividades, en este punto es fundamental la elaboración de un glosario de términos (Tabla 22), cuya finalidad es que tanto el creador o los creadores del proyecto como todas aquellas personas que lo lleven a cabo tengan un mismo vocabulario, con el objetivo de evitar problemas, ya que, ante un término que pueda tener diferentes significados o pueda ser entendido de diversas formas, a través del glosario, todos hacen uso de una misma definición, por lo que, dicho término, se va a emplear siempre y de la misma forma por ambas partes.

#### 4.2. Gestión del dato

La gestión del dato o también denominada *Data Management* es el “conjunto de actividades encaminadas a garantizar la entrega exitosa de datos relevantes y con niveles adecuados de calidad” (Especificación UNE 0078, 2023, 11), para procesar y analizar los datos siguiendo las directrices desarrolladas por el gobierno del dato. Además, a su vez, también seguirá los principios de gestión de calidad del dato.

Por tanto, el objetivo del *Data Management* es mantener los datos organizados de manera práctica y utilizable para garantizar que todo el cuerpo de datos de una organización sea preciso y coherente, de fácil acceso y protegido adecuadamente (KIO, 2023). En otras palabras, además de ser una forma de eliminar duplicados y estandarizar formatos, *Data Management* también sienta las bases para el análisis de datos.

Uno de los aspectos más importantes para la gestión del dato es el desarrollo y actualización de la arquitectura de datos, la cual hace referencia a la estructura que deben tener los datos y las operaciones, como son la ingesta, almacenamiento, integración, eliminación, etc., que se van a ejecutar sobre ellos.

La elaboración de una eficaz y eficiente gestión de datos, trae consigo una serie de beneficios, como son: la visibilidad, permite que los datos tengan una accesibilidad rápida y sencilla, fiabilidad, permite, como bien dice la palabra, que los datos sean fiables, actualizados y confiables reduciendo las posibilidades de cometer errores, la seguridad, protege contra posibles pérdidas o incluso robos de datos a partir de la implementación de herramientas de autenticidad y cifrado, y la escalabilidad, evita los costos de realizar acciones duplicadas.

Por otra parte, la gestión de los datos persigue varias funciones, entre las cuales destacan: *Data Governance* (Gobierno del Dato), definido anteriormente (véase la sección 4.1), *Data Architectures* (Arquitectura del dato), a partir de la cual se establecen modelos, reglas y políticas, *Data Modeling & Design* (Modelado y diseño del dato), encargado de diseñar la base de datos, de la implementación y del desarrollo del soporte, *Data Storage* (Almacenamiento de datos), función que determina cómo, cuánto y qué se almacena, *Data Security* (Seguridad de los datos), se encarga de todo lo relativo a la privacidad, confidencialidad y garantía de un acceso apropiado, *Data Integration & Interoperability* (Interacción e interoperabilidad de datos), por un lado, define la integración y, por otro, define la transferencia de los datos y *Data Quality* (Calidad de los datos), a través de la cual se define, controla y mejora la calidad de los datos.

Finalmente, es necesario destacar la diferencia existente entre el gobierno del dato y la gestión del dato, puesto que, este último engloba la parte práctica, es decir; la ejecución de los procesos y procedimientos, por lo que se encarga de recopilar, controlar y proteger los datos, mientras que, por el contrario, el gobierno del dato es el encargado de crear los procesos y asignarles a estos unos determinados responsables.

#### 4.3. Calidad del dato

Siguiendo con la especificación UNE 0079: 2023, la calidad del dato es el “grado en que las características de los datos satisfacen necesidades declaradas e implícitas cuando se utilizan en condiciones específicas” (Especificación UNE 0079, 2023, 10). Además, también se encarga de “proporcionar requisitos de calidad de datos a la capa de gestión de datos” (Gómez y Piattini, 2018, 8). Por tanto, la calidad del dato es importante para tener datos actualizados, fiables, íntegros y relevantes, puesto que, una mala calidad de datos puede afectar en la toma de decisiones o incluso provocar inconsistencias a través

de datos erróneos (Heredia y Vilalta, 2009, 45) entre otras dificultades duplicados o datos incompletos.

En consecuencia, es fundamental para cualquier organización tener una buena calidad de datos, cual se puede garantizar a partir de la aplicación de ciertas soluciones como son: la incorporación de procesos de perfilado, también denominado *data matching*, encargado de clasificar y cuantificar los problemas de calidad existentes en las fuentes de datos, como puede ser la redundancia de datos (PowerData, 2014), la limpieza de datos, también denominada *data cleaning*, a partir del cual se estandarizan los datos, se eliminan duplicados y se previene la aparición de valores nulos y la denominada “mejora de calidad del dato”, la cual permite realizar cambios tanto en el repositorio de datos como en su entorno para asegurar que se cumplan los requisitos de datos, eliminando o corrigiendo las no conformidades.

Con todo, la norma ISO/IEC 25012 determina que a la hora de desarrollar la calidad del dato se han de establecer una serie de características que sirvan para evaluar las propiedades de un determinado producto de datos. Por ello define 15 características (ISO/IEC 25012) a analizar, las cuales también se recogen, junto con sus correspondientes definiciones, presentadas a continuación, en la Especificación UNE 0079 del año 2023, más concretamente, entre las páginas 10 y 11:

- **Exactitud:** grado en el que los datos representan el verdadero valor del atributo seleccionado de un concepto en un determinado contexto de uso.
- **Compleitud:** grado en el que los datos asociados con una entidad tienen valores para todos los atributos e instancias en un determinado contexto de uso.
- **Consistencia:** grado en el que los datos están libres de contradicción y son coherentes con otros datos en un contexto determinado de uso.
- **Credibilidad:** grado en el que los datos tienen atributos que los usuarios consideran correctos y creíbles en un determinado contexto de uso.
- **Actualidad:** grado en el que los datos tienen se corresponden con una edad adecuada en un determinado contexto de uso.
- **Accesibilidad:** grado en el que los datos pueden ser accesibles por parte de las personas en un determinado contexto de uso.

- **Cumplimento:** grado en el que los datos se ajustan a las normas, políticas, reglamentos, etc. en relación con la calidad de los datos en un determinado contexto de uso.
- **Confidencialidad:** grado en el que se garantiza que los datos únicamente son accesibles por aquellos usuarios autorizados en un determinado contexto de uso.
- **Eficiencia:** grado en el que los datos pueden procesos y proporcionar niveles de rendimiento esperados en un determinado contexto de uso.
- **Precisión:** grado en el que los datos son exactos en un determinado contexto de uso.
- **Trazabilidad:** grado en el que los datos proporcionan un acceso auditado a los datos o cualquier cambio realizado en los datos en un determinado contexto de uso.
- **Comprensibilidad:** grado en el que los datos permiten ser leídos e interpretados por los usuarios haciendo uso de símbolos y lenguajes en un determinado contexto de uso.
- **Disponibilidad:** grado en el que los datos pueden ser recuperados por los usuarios o incluso por ciertas aplicaciones en un determinado contexto de uso.
- **Portabilidad:** grado en el que los datos pueden ser instalados, sustituidos o trasladados de un sistema a otro, manteniendo la calidad de estos en un determinado contexto de uso.
- **Recuperabilidad:** grado en el que los datos pueden mantener y preservar un nivel especificado de calidad en un determinado contexto de uso.

Por otra parte, siguiendo la Especificación UNE 0079 (Especificación UNE 0079, 2023, 12), dentro de la calidad del dato se distinguen tres componentes principales: la medición de la calidad del dato con la finalidad de determinar cuál es el grado de cumplimiento de los datos con relación a las características de la calidad del dato, la evaluación de la calidad del dato, la cual consiste en analizar si los datos presentan un nivel de calidad adecuado para el uso que se les quiere dar y la mejora de la calidad del dato, la cual hace referencia al conjunto de tareas que se han de llevar a cabo para mejorar los repositorios de datos a través de cambios realizados en los mismos con el objetivo de evitar todos los problemas posibles.



Por último, al igual que sucedía con el gobierno del dato, para la gestión del dato también se definen una serie de actividades que se deben llevar a cabo (Especificación UNE 0079, 2023, 13-14):

- **Planificar la calidad del dato:** identificar los requisitos de calidad del dato a partir de los objetivos desarrollados en la estrategia del dato, con el objetivo de planificar y desarrollar instrumentos necesarios que soportan la gestión de la calidad del dato.
- **Controlar y monitorizar la calidad del dato:** supervisar los valores de los datos a lo largo de un determinado tiempo, previamente establecido, para identificar y eliminar errores o no conformidades en los datos.
- **Asegurar la calidad del dato:** dotar a la organización de la confianza en que los requisitos de datos identificados durante la planificación de calidad del dato serán satisfechos.

#### 4.4. Componentes comunes

Para finalizar con la definición del gobierno, gestión y calidad del dato, es importante resaltar que las tres, siguiendo las correspondientes especificaciones (Especificación UNE 0077, Especificación UNE 0078 y Especificación 0079), mantienen los mismos componentes, los cuales se recogen en la siguiente ilustración:



Ilustración 3. Ilustración extraída de la Especificación UNE 077:2023 2023, 12.

Atendiendo a la figura aquí presente:

- Los **principios, políticas y marcos de referencia** han de ser realizados por la organización que implemente un sistema de gobierno del dato, más concretamente por el equipo de gobierno del dato que se ha de formar con los miembros de esta. Así, dichos principios se encuentran recogidos en la siguiente tabla:

Principio	Definición
Responsabilidad	El equipo de gobierno del dato debe adquirir las responsabilidades relacionadas con los datos y asegurarse de que la organización acepta y entiende las responsabilidades.
Estrategia	El equipo de gobierno del dato debe crear una estrategia del dato que concuerde con la global de la organización.
Adquisición	El equipo de gobierno del dato debe rendir cuentas acerca de la adquisición de los datos y debería asegurar que tales adquisiciones son propias mediante una consideración.
Desempeño	El equipo de gobierno del dato ha de identificar los indicadores de desempeño relevantes, asegurándose de que reciben la atención adecuada y que se llevan a cabo las acciones correctivas.
Conformidad	El equipo de gobierno del dato se ha de asegurarse de que los miembros de la organización cumplen y conocen tanto las obligaciones externas como las internas.
Comportamiento humano	Todos los miembros de la organización deben rendir cuentas en base a la utilización de los datos.

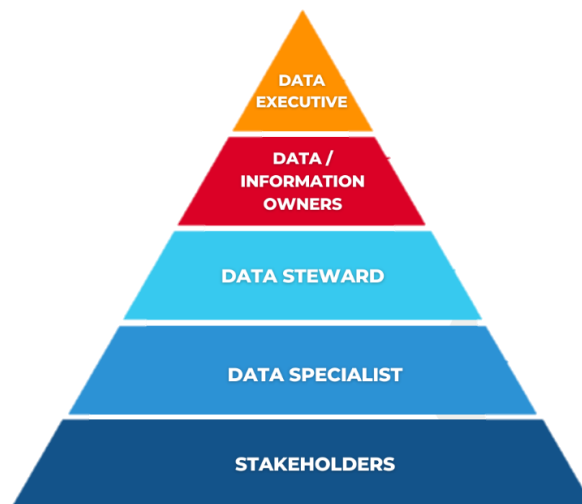
*Tabla 1. Tabla de principios con sus respectivas definiciones.*

- **Estructuras organizativas**, se encargan de identificar quiénes van a ser los responsables de los datos, las personas encargadas de diseñar e implementar la plataforma y aquellos que tomen las decisiones adecuadas para evitar el mayor número de problemas y riesgos posibles.
- **Procesos**, hacen referencia a todas aquellas actividades que se han de seguir para alcanzar los objetivos propuestos para el gobierno, gestión y calidad del dato, por ello, se ha de establecer la estrategia del dato, acompañada por una visión general, un propósito, los resultados y las tareas del proceso.

- **Información** generada a partir de cada uno de los procesos.
- **Cultura, ética y comportamiento.** Se refiere a las conductas individuales y colectivas del personal dentro de la organización).
- **Personas, habilidades y competencias,** de modo que es necesario tener dentro de proyectos de esta índole personal con ciertas habilidades y competencias que sean capaces de completar las actividades propuestas.
- **Servicios, infraestructuras y aplicaciones** sirven para llevar a cabo todas las acciones relacionadas con las tecnologías y sistemas de información, cuya responsabilidad será la de dar soporte para implementar y desarrollar la herramienta tecnológica seleccionada.

#### 4.5. Equipo de trabajo

En general, para todo proyecto, se vuelve necesaria la creación de un equipo de trabajo en el que cada persona tenga definido su rol, puesto que, de esta forma, se consigue que el trabajo se realice de una manera eficaz y eficiente sacando el máximo beneficio y obteniendo unos mejores resultados. En consecuencia, para los proyectos de esta índole, se establecen una serie principales cargos, los cuales se pueden ordenar por medio de una pirámide como bien se indica en la página web de Keyrus (Keyrus 2023):



*Ilustración 4. Pirámide de roles en proyectos de Gobernanza del dato extraída de la página web de Keyrus.*

De esta forma, los roles aquí presentes se pueden definir de la siguiente manera:

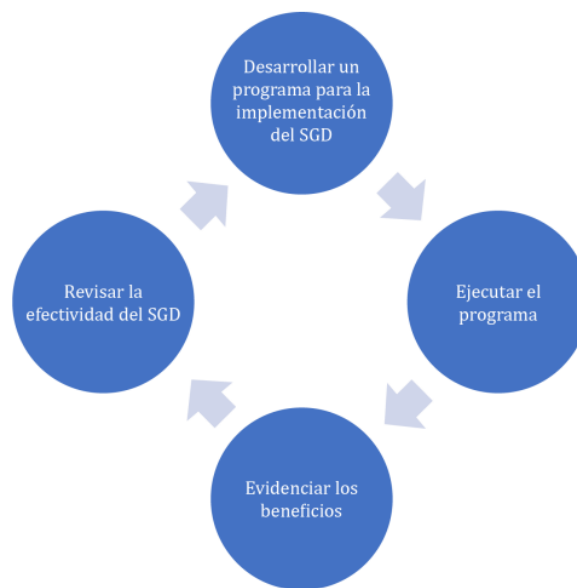
1. **Data executive** (ejecutivo de datos): es el líder del programa

2. **Data/Information Owner** (propietario de datos / información): son los responsables de la calidad, seguridad, acceso, privacidad, conservación y eliminación de los datos.
3. **Data Steward** (administrador de datos): son los responsables de la recopilación y registro de metadatos, revisión y verificación del modelo de datos, políticas de gestión de problemas y apoyan la calidad, uso y seguridad de los datos.
4. **Data Specialist** (especialista en datos): responsables del almacenamiento y disponibilidad de los datos, copias de seguridad, prevención de intrusión y de la pérdida o corrupción de datos.
5. **Stakeholders** (partes interesadas): son aquellas personas interesadas en poner en marcha dicho proyecto.

#### 4.6. Diseño y actividades de proyectos de gobierno, gestión y calidad del dato

Primeramente, es necesario tener en cuenta el diseño del proyecto. Por tanto, en primer lugar, se debe formar un equipo en el que cada uno tenga un determinado rol, como se muestra en la Ilustración 4, en función a las capacidades y conocimientos adquiridos. A continuación, se tiene que comprender el contexto de la organización, es decir, saber cuál es su estado actual. Posteriormente, se realiza el análisis de las carencias, problemas y la cantidad de ellas que son capaces de solventarse, encontrados en la fase previa de análisis. Esto implica definir el alcance inicial que supondrá la elaboración del proyecto y, por último, se tiene que refinar el diseño del sistema de gobierno a través de las estrategias y tecnologías que se van a emplear (Especificación UNE 0079, 2023, 34-25).

De esta forma, tras realizar el diseño del proyecto, se ha de llevar a cabo la implementación de este, desarrollando un programa de puesta en funcionamiento del gobierno, gestión y calidad del dato. Una vez desarrollado, se ha de ejecutar para, posteriormente poder evidenciar los resultados y revisar su efectividad, es decir, el éxito que ha supuesto la implementación de dicho sistema con el objetivo de evaluar tanto los beneficios como los posibles riesgos y solucionar errores que puedan suceder (Especificación UNE 0077, 2023, 36).



*Ilustración 5. Ilustración extraída de la Especificación UNE 0077:2023 sobre la implementación de un sistema de gobierno del dato.*

Así, en todo proyecto de esta índole, recogiendo tanto el diseño como la implementación, es importante saber qué actividades se han de seguir para desarrollar la estrategia del gobierno del dato. De esta forma y, siguiendo el artículo publicado por Elena Bello, en el año 2022 se pueden crear unas pautas, más concretamente, 6, generales, aplicables, como bien se ha dicho, a los proyectos relacionados con el gobierno, gestión y calidad del dato:

- **Identificación y priorización de los datos existentes**, de manera que se elabora un inventario en donde se recoja la información disponible con sus respectivos metadatos; una clasificación en datos estructurados y no estructurados; una conservación, a partir de la cual se organizan y gestionan los datos activamente; y se definen los roles de los miembros del equipo.
- **Selección de una opción de almacenamiento de metadatos**. Para ello se deben analizar diferentes alternativas para encontrar aquella que centralice los metadatos.
- **Transformación y actualización** de los datos correspondientes, a partir de la realización de ciertas actividades como: la limpieza y validación de datos, la transformación y actualización de los datos y la creación de un procedimiento o diccionario con el objetivo de que todos los miembros vean como están estructurados.

Una vez identificadas todas las actividades anteriores, lo siguiente es construir un modelo adecuado a las necesidades que se presenten. Finalmente, también se han de elaborar otras dos actividades: por un lado, establecer un proceso de distribución a partir de políticas de gobierno del dato relacionadas con las actividades, flujos de trabajo y herramientas habituales usadas por los miembros del equipo y, por otro lado, identificar posibles riesgos en la realización del proyecto.

## 5. Propuesta de proyecto de gestión del dato para la producción científica del personal docente e investigador de la Universidade da Coruña

### 5.1. Situación actual

Para comenzar con el propio proyecto de gestión de datos de la producción científica del personal docente e investigador de la *Universidade da Coruña*, se ha de conocer y analizar la situación actual de la misma, es decir, se debe realizar lo que se conoce como un *AS IS*, lo que en español significa “como es” y que consiste en definir el modo en el que la UDC, en este caso, almacena y trata los datos correspondientes a la producción científica de sus investigadores.

De esta forma, dentro de la propia universidad se ha desarrollado el *Sistema Unificado de Gestión de la Investigación*<sup>13</sup>, el cual permite tanto a los gestores como a los propios investigadores centralizar toda la información relativa a sus actividades de investigación, incluyendo los méritos recogidos en el *Curriculum vitae* del personal docente e investigador (Universidade da Coruña, 2015, 2).

No obstante, esta plataforma, a pesar de que su principal objetivo es el gestionar y mantener la información y los datos sobre la actividad investigadora del PDI actualizados, presenta una serie de problemas y carencias expuestas a continuación:

- **Falta de actualización**, debido a que el personal docente e investigador no lo lleva a cabo. En consecuencia, a la hora de visualizar los perfiles de estos, se van a encontrar datos faltantes acompañados por la frase “No figura en el gestor curricular de la UDC (SUXI)”.
- **Inexistencia de un plan de gobierno, gestión y calidad de datos** o, al menos, si existe, no se expone públicamente.
- **Duplicados**, a la hora de añadir datos informativos sobre la actividad científica del PDI, es posible que dos personas introduzcan el mismo mérito dos veces y, por tanto, en consecuencia, se publique información duplicada.

---

<sup>13</sup> Conocido por sus siglas SUXI-XC, en gallego, *Sistema Unificado de Xestión da Investigación – Xestión curricular*.

- **Falta de integración con terceros.** Atendiendo a la incorporación de datos, dicho sistema no presenta ningún tipo de integración con diferentes fuentes de datos existentes, como pueden ser *Scopus*, *Web of Science*, *Google Scholar*, *ResearchGate*, etc.
- **Visualización de la información** se realiza únicamente como un apartado de la página del investigador sin que tenga más relevancia en otras secciones del sitio web de la universidad.
- **Dificultad de recuperar información.** Es necesario resaltar que en ningún momento se permite la realización de búsquedas sobre la actividad científica del PDI de la *Universidade da Coruña*.
- **Información incompleta**, aunque los datos de la producción científica se muestran en el catálogo de grupos de investigación de la UDC de forma agregada, no se muestra toda la información disponible (por ejemplo, no se muestran los proyectos de investigación en los que participa el profesorado), ni se ofrece toda la funcionalidad (por ejemplo, no se muestran gráficas de resumen de información).

En consecuencia, a partir de este análisis previo sobre la situación actual de la universidad, a continuación, se ha propuesto una serie de actividades a seguir para implementar un proyecto que gobierne, gestione y mejore la calidad de los datos de la producción científica del personal docente e investigador dentro de la UDC.

## 5.2. Alcance de la propuesta

Para la elaboración del trabajo aquí presente, en primer lugar, cabe destacar que se hace uso de un gobierno del dato centralizado (Véase el apartado 4.1) puesto que, para comenzar, es decir, para incorporar el proyecto en la Universidad, las decisiones han de ser tomadas por un núcleo de personas que han de ser, principalmente, los directores que conforman el equipo de trabajo.

En segundo lugar, como se viene diciendo desde el inicio del trabajo, se ha decidido hacer uso únicamente de la producción científica del personal docente e investigador de la *Universidade da Coruña*. En este caso, solo se recogen los datos de dicha afiliación puesto que, por un lado, es de la que más conocimiento se tiene y, por otro lado, al tratarse de un Trabajo de Fin de Grado, si se planteara para las tres universidades gallegas, la información sería demasiado abundante y, en consecuencia, la



investigación duraría mucho tiempo. Es importante indicar que, a la hora de recoger datos, solo se han tenido en cuenta los relativos al PDI de la propia Universidad, asumiendo las consecuencias que conlleva no tener presente la información del personal externo. Además, se ha limitado la recolección de datos a cuatro tipos de publicaciones, es decir, los artículos en revistas, los libros, capítulos de libros y actas de congreso.

Por otra parte, como se ha expuesto en la sección 4 (Gobierno, Gestión y Calidad del dato) habría que definir procesos. Sin embargo, esta definición haría que el alcance del proyecto fuera muy superior al alcance de un Trabajo de Fin de Grado de la titulación, por lo que se ha decidido no abordarlo, llevando a cabo únicamente las tareas de definición de gobernanza de datos, de definición del modelo de datos y reglas de calidad del dato del proyecto.

No obstante, si en un futuro se pudiera llevar a cabo este proyecto o uno similar para gobernar la producción científica de una universidad, se podrían poner en marcha actividades, tareas y pautas complementarias a los primeros pasos propuestos a continuación.

### 5.3. Estructuras organizativas

Para la elaboración del proyecto es necesario contar con estructuras organizativas, dentro de las cuales, se han de crear miembros con roles determinados para cada una de las áreas. Eso supondría un aumento de beneficios, debido a que cada uno de ellos se centraría en la realización de sus correspondientes actividades y tareas, evitando la necesidad de estar pendiente de otros elementos de diferentes campos.

Tras analizar las estructuras organizativas de las que se compone la *Universidade da Coruña*, con el objetivo de determinar si alguno de los cargos existentes se puede aplicar al equipo de trabajo de este proyecto, se ha concluido que, para la creación de este equipo de trabajo, es necesario, por un lado, la creación de un Servicio de Gestión de Producción Científica, compuesto por personal de nueva contratación de administración y servicios cuyas titulaciones han de ser o bien Grado en Informática o bien Grado en Gestión Digital de Información y Documentación, cuya finalidad sería la de organizar, gestionar y manejar todos los datos relativos a la producción científica del personal docente e investigador de la UDC.

Por otro lado, también se han de definir roles más concretos, tanto de nueva creación como los ya existentes, así para finalizar con la creación del equipo de trabajo para este proyecto, se han identificado los siguientes puestos:

Puesto	Rol
Técnicos de sistemas y nuevas tecnologías.	Personal de Administración y Servicios <sup>14</sup> de la UDC que ocupe alguno de los siguientes puestos: Jefatura del Área de Operaciones y Administración Electrónica. Coordinador del Área de Servicios de información y gestión de indicadores. Analista del Área Gestión, investigación y bibliotecas.
Director de transformación digital.	Docente de la Facultad de Informática o de Humanidades y Documentación perteneciente a una de las siguientes áreas: Biblioteconomía y Documentación. Organización de empresas. Arquitectura y tecnología de computadores. Ciencia de la computación e inteligencia artificial. Ingeniería telemática. Lenguajes y sistemas informáticos.
Director del dato	Podría ser o bien el adjunto al Vicerrector para Investigación o bien el jefe de Servicio de Gestión de Producción Científica (puesto de nueva creación).
Director de tecnología.	Podría ser o bien el Adjunto del rector para tecnologías de la información y las Comunicaciones o un docente de la Facultad de Informática perteneciente a una de las siguientes áreas: Arquitectura y tecnología de computadoras. Ciencia de la computación e inteligencia artificial. Ingeniería telemática. Lenguajes y sistemas informáticos.

<sup>14</sup> PAS

Arquitectos del dato.	Personal del Servicio de Gestión de Producción Científica de nueva contratación
Director de seguridad de los datos.	Personal miembro de la UDC especializado en ciberseguridad de nueva contratación.
Director de protección de los datos.	Adjunta la Secretaría General.
Director de calidad del dato.	Vicerrector o vicerrectora de Política Científica, Investigación y Transferencia.
Propietarios del dato.	Personal de Servicio de Gestión de Producción Científica de nueva creación.

Tabla 2. Tabla de cargos y roles del equipo de trabajo del proyecto. Elaboración propia.

Cabe destacar que en el personal de Servicio de Gestión de Producción Científica deben existir dos grupos diferenciados, los arquitectos del dato y los propietarios del dato. De modo que, una misma persona solo puede pertenecer a uno de los dos grupos y no a los a la vez. Por otra parte, alguno de los roles aquí descritos se puede relacionar con los puestos de trabajo representados en la pirámide descrita en secciones anteriores (Ilustración 4):

<b>Puesto</b>	<b>Rol en la pirámide</b>
Director del dato	<i>Data executive</i>
Arquitectos del dato	Pueden ser: <i>Data specialist</i> <i>Data steward</i>
Propietarios del dato	<i>Data executive owners</i>
PDI y PAS	<i>Stakeholders</i>

Tabla 3. Relación entre puestos y roles representados en la pirámide (Ilustración 4). Elaboración propia.

Usando los roles que se han enumerado, se han definido estructuras organizativas, organizadas en tres comités. Por tanto, es necesaria la creación de tres comisiones: comisión de gobierno del dato, comisión de gestión del dato y comisión de calidad del dato. Por lo que, a continuación, de cada una de ellas se presentan sus componentes.

La comisión del gobierno del dato está compuesta por un grupo de responsables encargados de supervisar, soportar y monitorizar las iniciativas llevadas a cabo dentro de la organización, denominado “consejo de gobierno del dato” y los encargados de



coordinar el desarrollo de los diferentes componentes del sistema de gobierno y gestión del dato. Por tanto, esta primera comisión quedaría configurada del siguiente modo:

Comisión	Departamentos	Roles
Comisión del Gobierno del dato	Consejo de Gobierno del dato	Director de informática
		Técnicos de sistemas y nuevas tecnologías
		Director de transformación digital
	Oficina del dato	Personal del servicio de gestión de producción científica

Tabla 4. Composición de la comisión del gobierno del dato. Elaboración propia.

La segunda comisión es la de gestión del dato, en donde son fundamentales los directores, por lo que, entre ellos, se encuentra el director del dato que se corresponde con la misma persona que la definida en la comisión de gobierno del dato, el director de tecnología, quien dirigirá las iniciativas relacionadas con la infraestructura tecnológica del dato, arquitectos del dato, encargados de desarrollar y mantener la arquitectura de datos de la organización, el director de seguridad de la información y el director o delegado de protección de datos, encargado de las tareas relacionadas con la protección del dato. Esta segunda comisión de gestión del dato estaría compuesta del siguiente modo:

Comisión	Roles
Comisión de la gestión del dato	Director del dato (CDO)
	Director de tecnología (CTO)
	Director de seguridad del dato.
	Director o delegado de protección del dato (DPO)
	Arquitectos de los datos

Tabla 5. Composición de la comisión de gestión del dato. Elaboración propia.

Finalmente, en la tercera comisión de calidad del dato es fundamental tener, por un lado, a un director de la calidad del dato, que se encargue de dirigir las iniciativas de gestión de calidad del dato y, por otro lado, los propietarios de los datos, quienes deben conocer la semántica de los datos en el contexto de la organización y explicar qué

características deben de tener para ser usados con garantías de éxito. De esta forma, la comisión de calidad del dato estaría compuesta por:

Comisión	Roles
Comisión de calidad del dato	Director de calidad del dato
	Propietarios del dato

Tabla 6. Composición de la comisión de calidad del dato. Elaboración propia.

#### 5.4. Identificación de plataformas de extracción de información

Una vez se distribuyan los cargos dentro del gobierno del dato, de modo que cada persona involucrada tenga un determinado rol asignado, se han de identificar los lugares de los cuales se han de extraer los datos necesarios para poder poner en marcha el proyecto. Así, como se ha indicado anteriormente, debido a que muchas veces los investigadores o bien no actualizan o bien no incorporan todas sus contribuciones en la propia plataforma proporcionada por la UDC (SUXI-XC), se debe recurrir a otras fuentes, en este caso *online*, para completar la información lo máximo posible.

De esta forma, en primer lugar, se ha realizado un breve análisis con la finalidad de obtener los lugares en donde los miembros de la comunidad universitaria de A Coruña tengan una mayor afiliación. Así, entre ellas, destacan cuatro: *Scopus*<sup>15</sup>, *Dialnet*<sup>16</sup>, *WebOfScience*<sup>17</sup> y *Google Scholar*<sup>18</sup>. No obstante, es importante destacar que, en las cuatro plataformas aquí mencionadas, el formato más fácil al que se pueden extraer los dato y que, a la vez, sirva para cualquier tipo de herramienta tecnológica de gobierno del dato que se decida escoger, es el CSV<sup>19</sup>.

Primeramente, *Scopus*, una base de datos de referencias bibliográficas y citas de Elsevier<sup>20</sup>, con literatura *peer review*<sup>21</sup> y herramientas que permiten el análisis del investigador (FECYT, 2023). Es la plataforma que presenta un mayor número de afiliaciones, contando con un total de 4 176 afiliaciones de autores de la *Universidade da*

<sup>15</sup> Página de acceso a *Scopus*: <https://www.scopus.com/home.uri>

<sup>16</sup> Página de acceso a *Dialnet*: <https://dialnet.unirioja.es/>

<sup>17</sup> Página de acceso a *Web of Science*: <https://www.webofscience.com/wos/woscc/basic-search>

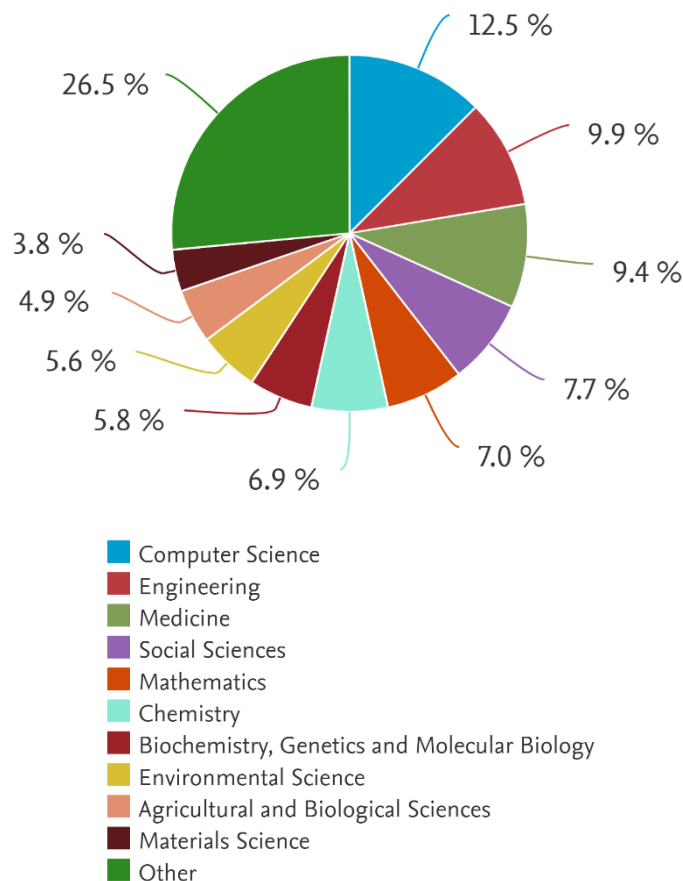
<sup>18</sup> Página de acceso a *Google Scholar*: <https://scholar.google.es/>

<sup>19</sup> *Command Separate Values*, se trata de un archivo de texto cuyos valores se encuentran separados por comas divididos en tablas.

<sup>20</sup> “Empresa de análisis de información global que asiste a instituciones y profesionales en el progreso de la ciencia” (Elsevier, 2023).

<sup>21</sup> Revisión por pares, proceso a través del cual se revisan y aprueban las investigaciones antes de ser publicadas por terceras personas.

Coruña y un total de 19 545 publicaciones. Cabe destacar que, el gráfico de afiliados por áreas temáticas no ha sido posible incorporarlo debido a que la página web no proporciona dicha información. Por tanto, a continuación, se muestra la gráfica de las 19 545 publicaciones clasificadas por áreas temáticas.



*Ilustración 6. Gráfica representativa de las publicaciones por áreas temáticas de los afiliados de la UDC a Scopus. Imagen extraída de la página web de Scopus.*

A partir de esta gráfica, se puede ver como la categoría de otros, en donde se incluyen área de conocimiento como arte, física y astronomía, economía, psicología, enfermería, odontología, etc. es en donde se clasifican la mayoría de las publicaciones de los afiliados a *Scopus* de la UDC, seguido por el área de ciencias de la computación y matemáticas.

Así mismo, en dicha plataforma no se pueden exportar todos los datos a la vez, sino que, únicamente permite la exportación de 800 publicaciones. No obstante, dicha acción puede ser realizada por varias personas durante un determinado periodo de tiempo, de modo que, buscando por la universidad, aparecen todos los autores afiliados y, a partir de estos, se extraen sus publicaciones. Cabe destacar que, dicha herramienta, a diferencia

de Dialnet, la cual se explicará a continuación, ofrece la posibilidad de exportar directamente a un formato CSV.

En segundo lugar, se encuentra Dialnet, una base de datos gratuita de literatura científica hispana, entre los cuales se incluyen artículos de revistas, libros, capítulos de obras colectivas, tesis doctorales, etc.) que, en el año 2023 contiene un total de 2 487 autores de la UDC afiliados a la plataforma, en donde destacan por encima de los demás los pertenecientes a las áreas de investigación de Tecnología (744 autores, que representa el 24% del total), seguida del área de Psicología y educación.

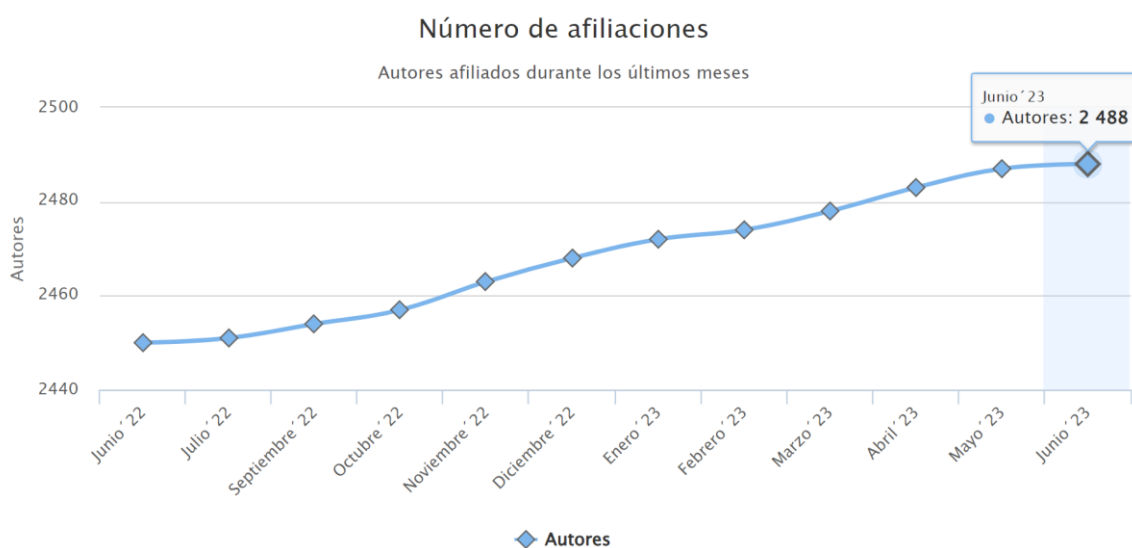


Ilustración 7. Gráfica del número de afiliados a Dialnet de la UDC. Imagen extraída de Dialnet métricas.

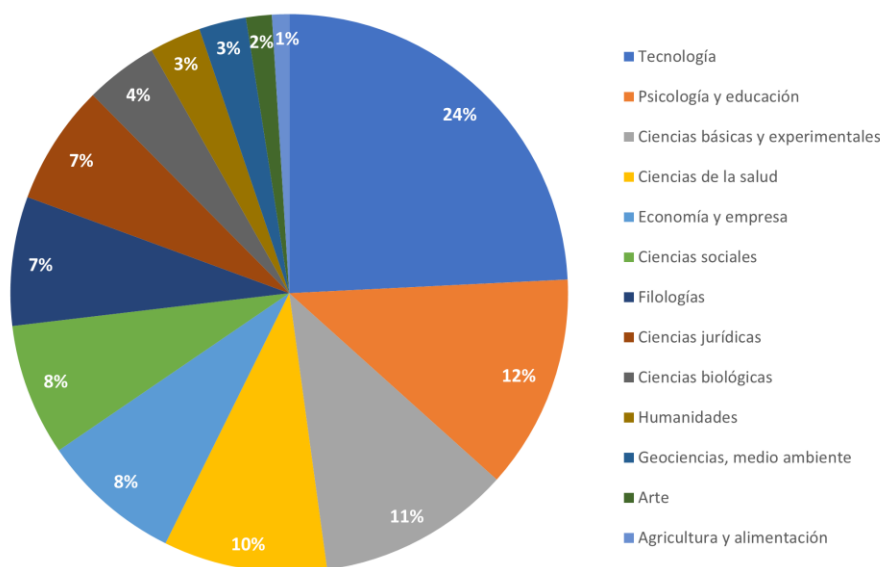


Ilustración 8. Gráfico correspondiente al número de autores afiliados a la UDC por materia. Gráfica de elaboración propia en base a la representada en Dialnet Métricas.

Por otra parte, para extraer los datos de dicha plataforma, en primer lugar, se debe realizar una búsqueda y filtrar por editores a “universidade da coruña”, una vez realizado este paso, se seleccionan todas las publicaciones existentes, se va al apartado “selección” y se exportan en formato *RIS*<sup>22</sup>, para a continuación, importarlos a *Zotero*<sup>23</sup> y, a partir de esa plataforma, volverlos a exportar al formato que se está buscando que, en este caso es CSV.

En tercer lugar, se encuentra la *Web of Science*, una base de datos de referencias bibliográficas y citas que recogen información desde 1 900 hasta la actualidad (FECYT, 2023), la cual cuenta con 20 218 publicaciones de los autores afiliados a dicha plataforma, de las cuales, en el año 2023, el área de ciencias de la computación, es decir, informática es la que cuenta con un mayor número de publicaciones. Es importante señalar que en este caso no se ha extraído el número total de afiliados a la plataforma, debido a que la cantidad de autores afiliados encontrados no es una cifra realista, puesto que es demasiado pequeña. Por ello, en la gráfica presente a continuación se muestra el número de publicaciones por área temática.

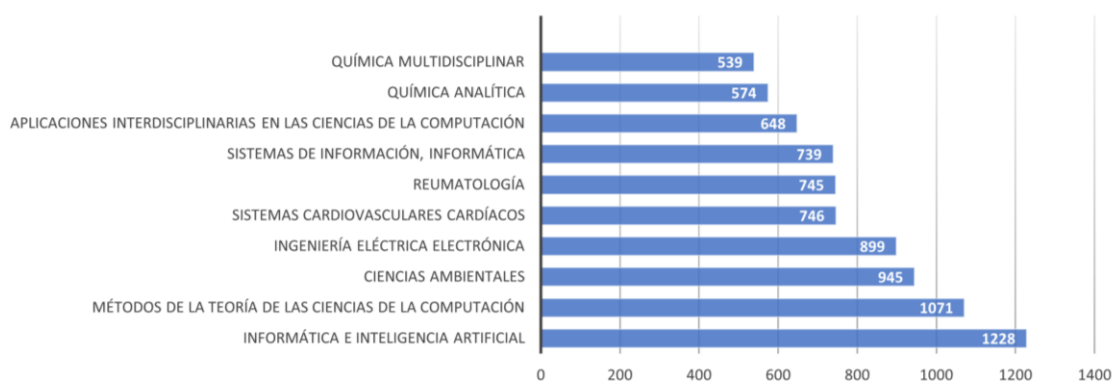


Ilustración 9. Gráfica representativa del número de publicaciones de los afiliados a la UDC por áreas temáticas en la *Web Of Science*. Gráfica de elaboración propia en base a la extraída de la *Web Of Science*.

Así mismo, para la extracción de datos de esta plataforma, es importante resaltar que solo permite extraer 1 000 datos de cada vez. No obstante, al igual que sucedía con *Scopus*, este proceso puede ser realizado por varias personas durante un determinado periodo de tiempo, por lo que, la duración de dicha acción se vería notablemente reducida. Por otra parte, tampoco permite exportar a un formato CSV, sin embargo, se puede

<sup>22</sup> *Research Information Systems* o Sistema de Información de Investigación.

<sup>23</sup> Gestor de referencias bibliográficas.



descargar como un fichero Excel para, posteriormente, una vez dentro de la plataforma, se guarde como un fichero .CSV.

Por último, con relación al análisis de lugares de publicación, se encuentra *Google Scholar*, en inglés, Google Académico en español, el cual cuenta con un total de 949 perfiles de docentes e investigadores afiliados de la *Universidade da Coruña*. Cabe destacar que, en este caso, no se ha podido extraer ninguna gráfica referente a dicha plataforma debido a que, a diferencia de las demás, *Google Scholar* no realiza una clasificación por áreas temática, por tanto, no se pueden extraer los datos necesarios.

A diferencia de los demás, la forma de extracción de datos de *Google Scholar* resulta más complicada, por lo que es necesario tener unos conocimientos mínimos previos sobre las herramientas que se van a utilizar. Por ende, una forma de hacerlo es a partir del *web scraping online*, haciendo uso de la aplicación *import.io*, de pago, aunque presenta una versión gratuita, la cual analiza de forma automática la estructura de la página web correspondiente, mostrando los datos en ordenados en tablas. No obstante, el principal problema es que dicho proceso no reconoce las páginas siguientes, por lo que el tiempo de extracción de datos se ve aumentado (Murillo, Saavedra y Quintero, 2018, 97).

Por otra parte, se podrían utilizar paquetes en R a través de diversos algoritmos para *scrapear*<sup>24</sup> datos, los detalles y las publicaciones de los perfiles de una determinada afiliación (Murillo, Saavedra y Quintero, 2018, 100-101).

Así, tras este análisis, cabe destacar que para completar la recogida de datos y obtener así la información lo más completa posible, es necesario que se incluyan todas las publicaciones existentes en las cuatro plataformas, ya que, a pesar de que algunas presenten casi los mismos contenidos, por ejemplo, el caso de *Scopus* y *Web of Science*, puede darse el caso de que en una esté recogida una publicación que en la otra no y si en un primer momento se descartara la selección de alguna de ellas, sería información pérdida.

### 5.5. Identificación de datos a incorporar en el proyecto

Atendiendo a los datos, como este proyecto se centra en la producción científica de los investigadores, los datos a identificar serían los relacionados con la misma, es decir;

---

<sup>24</sup> Forma de extraer información de páginas web a través del uso de código.

información relativa al investigador; a los diferentes tipos de publicaciones científicas, como son, los artículos de revistas, actas de congresos, libros y capítulos de libros; y a los lugares de publicación estas como son, por ejemplo, las revistas.

Así mismo, es importante señalar que tanto los grupos de investigación como sus miembros ya son gestionados por el propio servicio de investigación de la UDC, en consecuencia, se encontrarían fuera del alcance de este plan. También se considera que la universidad gestiona al personal docente e investigador y, en consecuencia, además de tener los datos personales de cada uno de ellos, sabe, para cada investigador, si está asociado a un grupo de investigación o, por el contrario, si se trata de un investigador individual. Además, es importante señalar que se asumen los inconvenientes que supone no recoger la información de los autores externos a la *Universidade da Coruña*, sino que, en este trabajo, únicamente se cuenta con los miembros de dicha afiliación.

Por ello, asumiendo esto, se han creado una serie de tablas en donde se recogen todos los datos necesarios para su posterior implementación en la plataforma tecnológica seleccionada. Así, en primer lugar, se recogen los datos relativos a los identificadores de los investigadores en la UDC:

<b>Datos de la producción científica de los investigadores</b>	
<b>Datos</b>	<b>Definición</b>
ID_investigador	Código identificativo único del investigador dentro de la universidad.
ORCID	Código identificativo numérico único que tienen cada uno de los investigadores compuesto por cuatro conjuntos de cuatro dígitos separados por guiones.
<i>Scopus ID</i>	Código identificativo numérico único que tienen cada uno de los investigadores compuesto por 10 dígitos.
<i>Researcher ID</i>	Código identificativo alfanumérico único que tienen cada uno de los investigadores compuesto por una letra mayúscula, un guion, cuatro dígitos, otro guion y otros cuatro dígitos.
Índice H	Es una métrica a nivel de investigador encargado de medir su calidad profesional en base a la cantidad de citas



	que han recibido sus artículos. Se calcula ordenando los artículos de mayor a menor en cuanto al número de citas, siendo el índice h aquel en el que coincide el número de citas y el número de orden.
--	--

Tabla 7. Tabla de recolección de datos de los identificadores de los investigadores. Elaboración propia.

En segundo lugar, se presentan los datos relativos a las investigaciones:

<b>Datos de las publicaciones</b>	
<b>Datos</b>	<b>Definición</b>
Título	Título completo en el idioma correspondiente de la publicación.
Año de publicación	Año de publicación de la investigación.
Lugar de publicación	Nombre completo del lugar (ciudad) de publicación de la investigación.
País	Nombre del país en el que se ha realizado la publicación.
URL en <i>Scopus</i>	Enlace a la publicación en <i>Scopus</i> .
URL en <i>Web of Science</i>	Enlace a la publicación en la <i>Web of Science</i> .
URL en Dialnet	Enlace a la publicación en Dialnet.
URL en <i>Google Scholar</i>	Enlace a la publicación en <i>Google Scholar</i> .
Tipo de publicación	Nombrar el tipo de publicación del que se trata.
Idioma de la publicación	Idioma original en el que está escrito la publicación.
Número de citas	El número total de citas que ha recibido, si se dispone de él.

Tabla 8. Tabla de recolección de datos de las publicaciones. Elaboración propia.

En tercer lugar, se presentan los datos sobre los artículos de revista, revistas y sus respectivos indicadores:

<b>Datos de los artículos de revista.</b>	
<b>Datos</b>	<b>Definición</b>
Volumen de la revista	Volumen de la revista en el que se encuentra publicado dicho artículo.
Número de la revista	Número de la revista en el que se encuentra publicado el artículo.



Páginas	Número de página en donde empieza el artículo y número final de página en donde termina.
---------	--

Tabla 9. Tabla de recolección de datos de los artículos de revista. Elaboración propia.

<b>Datos sobre las revistas</b>	
<b>Datos</b>	<b>Definición</b>
Nombre	Nombre completo de la revista en donde se ha publicado el trabajo.
Tipología	Área de conocimiento de la revista en base a una lista controlada establecida en los requisitos de los datos.
Editorial	Nombre completo de la editorial de la revista.
ISSN	Código identificativo de la revista
Idioma	Idioma en el que se escribe la revista.
Periodicidad	Periodicidad con la que salen nuevos números de la revista (anual, mensual, semanal).
Nacionalidad	Indicar si dicha revista se publica a nivel nacional o internacional.
URL	Enlace a la página web en la que está disponible

Tabla 10. Tabla de recolección de datos de las revistas. Elaboración propia.

Así mismo, con relación a las revistas, se recogen los datos relativos a sus indicadores, es decir, factor de impacto y cuartil.

<b>Datos de indicadores de revistas.</b>	
<b>Datos</b>	<b>Definición</b>
Tipo de indicador	Nombre del indicador al que se está haciendo referencia. Estos indicadores se encuentran en una lista controlada en el apartado de requisitos del dato.
Factor de impacto	Indicador que mide la frecuencia con la que una revista ha sido citada a un determinado año.
Cuartil	Indicador de posicionamiento de una revista en relación con todas las de su área.
Año	Año en el que se están registrando los datos.

Tabla 11. Tabla de recolección de datos de los indicadores de revistas. Elaboración propia.

Por último, se recogen los datos relativos a los restantes tipos de publicaciones que se van a incorporar, es decir, libros, actas de congreso y capítulos de libros.

<b>Datos de los libros</b>	
<b>Datos</b>	<b>Definición</b>
Título	Nombre completo del libro
Editores	Nombre de los editores del libro.
Editorial	Nombre de la editorial del libro.
ISBN	Código identificativo del libro.

Tabla 12. Tabla de recolección de datos de los libros. Elaboración propia.

<b>Datos de las actas de congreso</b>	
<b>Datos</b>	<b>Definición</b>
Título del acta	Título del acta de congreso.
Nombre del congreso	Nombre completo del congreso del que se ha extraído el acta.
Organizador	Nombre del organizador del congreso.
Páginas	Número de página inicial en la cual empieza el acta y número de página final en donde termina el acta.

Tabla 13. Tabla de recolección de datos de actas de congreso. Elaboración propia.

<b>Datos de capítulos de libros</b>	
<b>Datos</b>	<b>Definición</b>
Título	Título completo del capítulo del libro.
Páginas	Número de página inicial en donde empieza el capítulo del libro y número de página final en donde termina el capítulo del libro.

Tabla 14. Tabla de recolección de datos de capítulos de libros. Elaboración propia.

## 5.6. Modelo conceptual

Tras la búsqueda y elección de los datos identificados en el apartado anterior, a continuación, toda la información recopilada anteriormente, se recoge en un modelo relacional, representado a través de un modelo conceptual (diagrama Entidad-Relación, también conocido como ER), compuesto por: entidades, es decir objetos en la base de datos representados por medio de cajas y atributos, los cuales hacen referencia a la información sobre una entidad.

Además, las entidades se conectan a través de líneas, las cuales tienen notaciones, también conocidos como indicadores para explicar el tipo de relación existente entre ambas. Atendiendo a las entidades, para los artículos de revista, actas de congreso y capítulos de libro se ha de crear un nuevo atributo que sea un identificador que, a su vez, funcione como clave primaria de cada una de ellas. Por otra parte, para su representación se usa la notación denominada *crow's foot notation*, en donde la cardinalidad adquiere una gran relevancia, puesto que, existe un número mínimo y máximo situado en los extremos de las relaciones, con la finalidad de definir la relación. Así, esto se puede representar a través de un modelo conceptual:

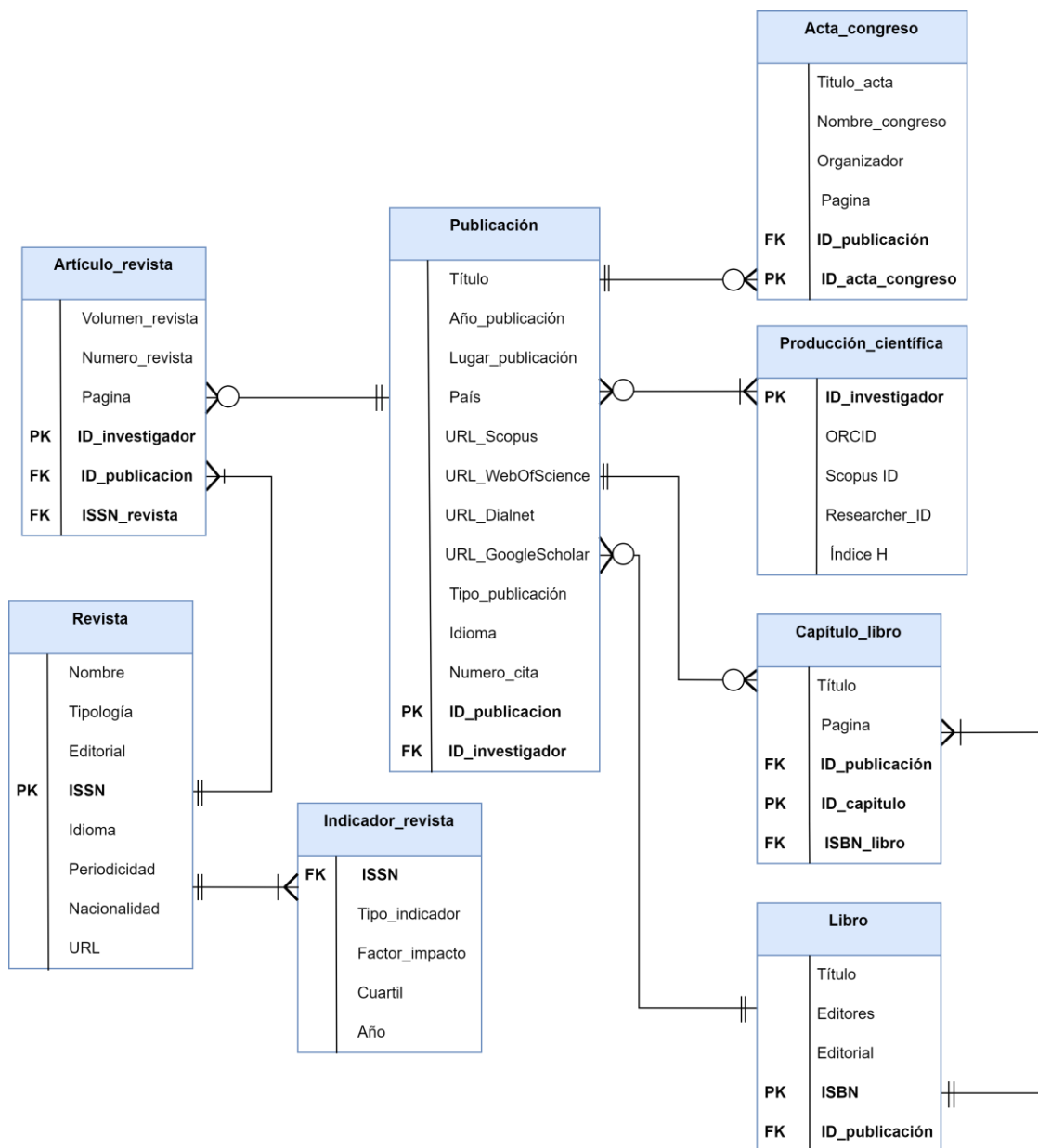


Ilustración 10. Modelo conceptual del proyecto. Elaboración propia.

### 5.7. Requisitos de los datos

Los requisitos de los datos se expresan a partir de las características de calidad del dato, las cuales sirven para juzgar la calidad de un conjunto de datos (Especificación UNE 0079, 2023, 10). Con todo, tras la definición de los datos, las tablas en las cuales se han de recolectar y las relaciones existentes entre ellas, se han elaborado una serie de requisitos de los datos con el objetivo de que estos adquieran una buena calidad. De forma que, a continuación, se presenta una lista de requisitos que se han de tener en cuenta a la hora de incorporar los datos para, por un lado, como bien se ha dicho, tener una cierta calidad y, por otro lado, para que estos estén controlados con el objetivo de que siempre se mantengan las mismas estructuras. Así, dicha lista quedaría del siguiente modo:

- El identificador ORCID debe mantener la estructura: 0000-0000-0000-0000.
- El identificador *Research ID* debe cumplir con la siguiente estructura: A-0000-0000
- El identificador *Scopus ID* debe contener un total de 10 dígitos.
- El índice H debe ser únicamente un dígito sin incluir letras.
- Los años se deben escribir completos, es decir, aaaa.
- Los códigos identificativos (ID) de todos los tipos de publicaciones, es decir, artículo de revista, acta de congreso y capítulo del libro solo pueden contener números.
- El ISBN ha de estar compuesto por 13 cifras, divididas en cuatro partes de extensión variada y separadas con un guion, precedidas de las letras ISBN.
- Las páginas se han de representar a través de dígitos, el primero de ellos hace referencia a la página de inicio y el segundo a la página de fin y ambos irán separados mediante un guion sin espacios, por lo que, quedaría del siguiente modo: 1-1.
- Se ha creado una lista controlada para los siguientes campos:
  - o Nombre completo de los editores de los libros.
  - o Para el cuartil, cuyos valores son:
    - Q1.
    - Q2.
    - Q3.
    - Q4.

- Para el área de conocimiento de las revistas, cuyos valores son:
  - Agricultura y alimentación.
  - Arte.
  - Biología.
  - Ciencias de la computación.
  - Ciencias de la salud.
  - Ciencias jurídicas.
  - Ciencias sociales.
  - Economía y empresa.
  - Filología.
  - Humanidades.
  - Ingeniería.
  - Matemáticas.
  - Medicina.
  - Psicología y educación.
  - Química.
  - Tecnología.
  - Otros.
- Para los tipos de indicadores, cuyos valores son:
  - *SCImago Journal & Country Rank (SJR)*<sup>25</sup>.
  - *Journal Citation Reports (JCR)*<sup>26</sup>.
  - Dialnet Métricas.
- Para la periodicidad de la publicación de las revistas, cuyos valores son:
  - Diaria.
  - Semanal.
  - Mensual.
  - Trimestral.
  - Anual.
  - Otra.

---

<sup>25</sup> Indicador de la influencia de las revistas científicas midiendo el número de citas obtenidas en cada publicación en un periodo de 3 años, dando mayor importancia a aquellas procedentes de revistas con alto impacto.

<sup>26</sup> Indicador más antiguo y valorado que muestra la relevancia de la revista en su determinada área, accesible a través de Fundación española para la Ciencia y la Tecnología.



## 5.8. Análisis comparativo de herramientas tecnológicas

Finalmente, para desarrollar todo lo planteado hasta el momento se ha de seleccionar una herramienta tecnológica, en este caso, propia del gobierno del dato. Por eso, es necesario realizar un análisis comparativo entre algunas de ellas que están a disposición de los usuarios en el mercado, tanto de pago como gratuitas. De esta forma, de cada una de ellas se han analizado ciertos puntos, considerados relevantes a la hora de tomar decisiones. A continuación, se presentan las diferentes herramientas, con sus respectivos nombres, acompañados de una breve introducción en donde se define lo que son y qué hacen y, seguidamente, se representan en una tabla cuatro aspectos considerados relevantes que son:

- Las características que presenta, es decir; las funcionalidades que se pueden utilizar en cada uno de los Softwares.
- Los conectores de datos con los que se puede integrar.
- El precio, se indicará tanto si lo tiene como si no.
- La madurez de la aplicación, es decir, el nivel de desarrollo que tiene.

Por último, tras el análisis individual se ha elaborado una tabla en donde se exponen las ventajas y desventajas de cada una de ellas. Así, se han analizado un total de 5 herramientas tecnológicas:

*Azure Purview*, es una herramienta de solución de Gobierno del Dato basada en la nube, creada por Microsoft, cuyo objetivo es facilitar tanto la administración como el control de los datos en un entorno local, en la nube y de software como servicio SaaS<sup>27</sup>, permitiendo, de este modo, que todas aquellas personas que deseen consumir dicha información tengan la seguridad de que están accediendo a datos fiables y de gran valor (Microsoft Azure, s.f.).

Características	Integración de datos	Precio	Madurez
Descubrimiento de datos Gobierno del dato	SQL <sup>28</sup> , <i>Google BigQuery</i> <sup>29</sup> , <i>AmazonS3</i> <sup>30</sup> , Amazon	Consultar con la empresa proveedora de la herramienta,	Baja (en proceso de desarrollo)

<sup>27</sup> Software como servicio basado en la nube que se pone a disposición del cliente a través de Internet.

<sup>28</sup> Structure Query Language, es un lenguaje de consulta

<sup>29</sup> Almacén de datos administrado por Google.

<sup>30</sup> Servidor de almacenamiento de archivos.

Conectividad de origen de datos	RDS <sup>31</sup> , <i>Oracle</i> , <i>Hive</i> <sup>32</sup> ,	<i>Azure</i> , <i>Power</i>	puesto que se paga por los servicios usados
Colaboración	MySQL <sup>33</sup> ,		
Catalogación y metadatos	<i>BI</i> <sup>34</sup> , CSV, etc.		
Seguridad			
Catálogo			

Tabla 15. Tabla análisis de la herramienta Microsoft Purview. Elaboración propia.

La siguiente herramienta analizada es *Talend*, se trata de una solución de Big Data en la nube fácil, potente, flexible, en donde se combina la integración de datos, la calidad de datos y el gobierno de datos bajo código que se integra con prácticamente cualquier fuente y arquitectura de datos (Talend, 2023).

Características	Integración de datos	Precio	Madurez
Gobierno del dato	AWS <sup>35</sup> , <i>RedShift</i> <sup>36</sup> ,	Acceso Abierto.	Alta
Integración en la nube	<i>Microsoft Azure</i> ,	Integración de datos	(Desde
Integración de datos	procesos ETL,	en la nube de Talend:	2005)
Compatibilidad con otras fuentes de datos	<i>Snowflake</i> <sup>37</sup> , CSV, etc.	1170\$/usuario/mes	
Catálogo de datos			
Administración de datos			
Inventario de datos			

Tabla 16. Tabla análisis de la herramienta Talend. Elaboración propia.

La tercera herramienta analizada es *Apache Atlas*, creada por Hortonworks, permite cumplir de manera efectiva y eficiente los requisitos de los datos, así como también permite realizar una integración con el ecosistema de datos, proporcionando capacidades de administración y gobierno de metadatos con la finalidad de crear un catálogo de datos (Apache Atlas, s.f.).

<sup>31</sup> Servicio web de bases de datos relacionales.

<sup>32</sup> Tecnología de almacenamiento de datos desarrollada por Apache, pensado para gestionar grandes cantidades de datos.

<sup>33</sup> Sistema gestor de bases de datos.

<sup>34</sup> Plataforma de inteligencia empresarial (Business Intelligence) dirigida a grandes empresas.

<sup>35</sup> Servidor de almacenamiento de datos desarrollado por Amazon (*Amazon Web Service*).

<sup>36</sup> Igual que AWS.

<sup>37</sup> Herramienta de análisis de datos en la nube (*SaaS*).

Características	Integración de datos	Precio	Madurez
Metadatos	<i>Hive</i> , <i>Hbase</i> <sup>38</sup> ,	Acceso abierto	Media
Linaje de datos	<i>Siqoop</i> <sup>39</sup> , <i>Storm</i> ,		(desde
Seguridad de los datos	<i>Kafka</i> <sup>40</sup> , CSV, etc		2015)
Búsqueda de datos			

Tabla 17. Tabla análisis de la herramienta Apache Atlas. Elaboración propia.

En cuarto lugar, se encuentra la herramienta tecnológica analizada *Atlas*, una plataforma que posibilita una protección y democratización de los datos, personalización de políticas de gobierno y creación de *bots* personalizados para clasificar automáticamente los datos confidenciales. Todo ello basado en una arquitectura de API abierta, lo cual permite la integración con herramientas de datos modernas y populares (Atlas s.f.).

Características	Integración de datos	Precio	Madurez
Catálogo	<i>Amazon S3</i> , <i>Athenea</i> <sup>41</sup> ,	Consultar con la	Baja
Activo de datos 360	<i>Redshift</i> , <i>Google</i>	empresa para obtener	(desde
Linaje de datos	<i>BigQuery</i> , <i>Azure</i> ,	más información	2018)
Colaboración integrada	<i>Kafka</i> , SQL, MySQL,		
Glosario empresarial	<i>Oracle</i> , <i>PostgreSQL</i> <sup>42</sup> ,		
Personalización y curación	<i>Power BI</i> , CSV, etc.		
Metadatos			
APIs abiertas			

Tabla 18. Tabla análisis de la herramienta Atlas. Elaboración propia.

Posteriormente, se ha analizado *Tableau*, una de las plataformas más populares de análisis de datos, a través de la cual, por un lado, se pueden visualizar los datos y, por otro lado, se pueden resolver los problemas que surjan en relación con los mismos (Tableau, 2023).

<sup>38</sup> Sistema gestor de bases de datos no relacionales.

<sup>39</sup> Herramienta en línea desarrollada por Apache que permite transferir datos a bases de datos relacionales.

<sup>40</sup> Plataforma que permite la transmisión de datos en tiempo real, igual que *Storm*.

<sup>41</sup> Servicio de código abierto de Amazon para analizar datos a partir de un lago de datos.

<sup>42</sup> Sistema gestor de bases de datos relacionales.



Características	Integración de datos	Precio	Madurez
Seguridad de los datos	Excel, SQL, <i>RedShift</i> ,	Tableau Creator: 70\$	Alta (desde 2003).
Facilidad de uso	<i>PostgreSQL</i> , <i>Google</i>	persona/mes	
Conectividad de datos en tiempo real	<i>Sheets</i> , <i>Google</i>	Tableau Explorer: 42\$ persona/mes	
Alta compatibilidad con otras fuentes de datos	<i>BigQuery</i> , <i>Hive</i> ,	Tableau Viewer 15\$ persona/mes	
Gobierno del dato	<i>Snowflake</i> , <i>Amazon</i> <i>Athenea</i> , <i>Google</i> <i>Analytics</i> , CSV, etc.		

Tabla 19. Tabla análisis de la herramienta Tableau. Elaboración propia.

Finalmente, tras analizar las diversas herramientas individualmente, se ha recogida en la siguiente tabla tanto los aspectos positivos, denominados “pros”, como los aspectos negativos “contras” de cada una de ellas, con la finalidad de que puedan ser analizados de forma rápida y sencilla.

Pros de las herramientas tecnológicas				
<i>Azure Purview</i>	<i>Talend</i>	<i>Apache Atlas</i>	<i>Atlan</i>	<i>Tableau</i>
Fácil arranque sin necesidad de implementación	Es de acceso abierto	Es de acceso abierto	Fácil integración con terceros.	Prueba gratuita.
Permite la catalogación de los datos.	Compatibilidad con otras fuentes de datos	Permite el linaje de datos	Permite el linaje de datos	Permite el filtrado de datos
Permite realizar búsquedas de datos	Fácil de usar	Permite realizar búsqueda de datos	Permite realizar búsquedas de datos	Permite realizar búsquedas de datos
Garantiza la seguridad de los datos	Garantiza la seguridad de los datos	Garantiza la seguridad de los datos	Garantiza la seguridad de los datos	Garantiza la seguridad de los datos
Permite añadir etiquetas personalizadas	Manejo de múltiples			Permite realizar



	fuentes de datos			comentarios en las tablas
Permite construir un glosario				Permite crear tablas interactivas

Tabla 20. Tabla de pros de las herramientas tecnológicas analizadas anteriormente. Elaboración propia.

<b>Contras de las herramientas tecnológicas</b>				
<i>Azure Purview</i>	<i>Talend</i>	<i>Apache Atlas</i>	<i>Atlan</i>	<i>Tableau</i>
Lista de fuentes de datos	No hay prueba gratuita	Elevado número de herramientas para ponerlo en marcha	Tiempo elevado de resolución de problemas	No hay prueba gratuita
No soporta automatización ni alertas	Mal soporte técnico		Límite a la hora de integrar con terceros	Mal soporte técnico
La relación entre archivos se establece manualmente	No hay servicios premium de integración o consultoría		Operaciones faltantes por estar en proceso de desarrollo	Precio elevado

Tabla 21. Tabla de contras de las herramientas tecnológicas analizadas anteriormente. Elaboración propia.

### 5.9. Beneficios de la implementación

Finalmente, tras todo lo analizado anteriormente, es importante destacar los beneficios que supondría la implementación de una eficaz y eficiente estrategia de gestión del dato para la producción científica del personal docente e investigador de la *Univerisdade da Coruña*.

- Aumenta el valor de los datos, puesto que, a partir de este momento, se sabe con certeza que estos son fiables y están actualizados.
- Ahorro de tiempo al personal docente e investigador, tanto para incorporar publicaciones como para buscarlas.

- Ahorro de tiempo en la elaboración de tareas debido a la creación de unas comisiones en donde se indican las personas junto con sus roles que serán las encargadas de realizar las actividades correspondientes.
- Se evita tener datos duplicados o la incorporación del mismo dato por la misma persona.
- Aumento en la calidad del dato debido a la implementación de requisitos que indican la estructura obligatoria que estos deben de tener.

## 6. Conclusiones

Finalmente, para acabar con el trabajo aquí presente, se han desarrollado una serie de conclusiones, en las cuales se identifican aquellos objetivos cumplidos, las lecciones aprendidas y los trabajos futuros que se pueden llevar a cabo a partir de los conocimientos aquí presentes.

Así, atendiendo a los objetivos previamente establecidos al inicio del proyecto (véase la sección Objetivos). Se ha alcanzado la elaboración de un análisis de las fuentes de datos en donde el personal docente e investigador publica sus trabajos a partir de una pequeña investigación en sus respectivas páginas *web*, de las cuales se ha extraído tanto el número de afiliados y como el número de publicaciones de los autores de la *Universidade da Coruña*. Ambas cifras han sido acompañadas por gráficos, algunos extraídos de las propias plataformas y otros de elaboración propia, en base a los existentes en las páginas *web*, para obtener una mejor visualización de los resultados. La principal función de exponer las gráficas ha sido la de ilustrar dichos resultados para conocer la importancia de la Universidad en esas plataformas. Además, también se ha averiguado el modo de extracción de datos a un formato CSV, para su futura implementación en la herramienta tecnológica seleccionada. Esta última investigación también, por un lado, se ha elaborado en base a los conocimientos previamente adquiridos durante la titulación, pero, por otro lado, se han hecho búsquedas autónomas para aquellas que requerían el uso de materiales externos.

El siguiente objetivo cumplido ha sido la definición de un modelo de datos para almacenar la producción científica del PDI en la UDC a través de la creación de tablas, en las cuales se han identificado qué datos se han de guardar. Posteriormente, a partir de estas tablas se ha elaborado un modelo conceptual representado en un diagrama entidad-relación para visualizar, de forma rápida y sencilla, tanto las clave primarias y foráneas como las relaciones entre las entidades. Por último, para acabar de definir el modelo de datos, se han establecido una serie de requisitos, representados a partir de una lista, que se han de cumplir a la hora de introducir datos para poder mantener su calidad, evitando duplicaciones, datos erróneos y manteniendo la normalización entre ellos.

Además, también se ha conseguido analizar la situación actual de la Universidad con relación a la gestión de la producción científica. De modo que, se ha investigado

sobre la existencia de alguna plataforma en la cual se recogieran este tipo de datos, dando como resultado el SUXI. A partir de la documentación proporcionada por la página web de la *Universidade da Coruña* sobre dicho sistema, se han identificado sus características y definido sus carencias.

Siguiendo con los objetivos, también se han identificado herramientas tecnológicas relacionadas con el gobierno del dato a partir del análisis de plataformas de esta tipología disponibles en el mercado. Así se han seleccionado cinco de ellas que fueran variadas, es decir, que tuvieran diferentes grados de madurez y que fuesen tanto gratuitas como de pago. A partir de aquí se realizaron búsquedas para obtener información más concreta, la cual se ha representado en tablas individuales. Finalmente, a través de toda la información recolectada, se realizó un análisis comparativo entre todas ellas exponiendo tanto las ventajas como las desventajas.

El último objetivo cumplido ha sido la profundización sobre temas de actualidad, en este caso sobre el gobierno, gestión y calidad del dato a partir de una revisión bibliográfica. Por tanto, es importante mantenerse actualizado puesto que, como bien se ha señalado, hoy en día, los datos, así como la tecnología que los rodea, se encuentran en pleno desarrollo, por lo que, en consecuencia, las herramientas y proyectos también evolucionan, de modo que, lo que hoy sirve mañana ya no o puede haber otras alternativas mejores. Por ende, estar al tanto, sobre todo dentro de este mundo de los datos y de los tiempos que corren, resulta fundamental y esencial cuando se pretende desarrollar un proyecto de dicha tipología.

Además de los objetivos alcanzados, también se han identificado una serie de lecciones aprendidas entre las cuales destacan: el aprendizaje de un plan de gobernanza del dato no visto durante la titulación, el enfrentamiento ante un trabajo de gran complejidad y gran valor, a día de hoy, para aquellas organizaciones que manejan una gran cantidad de datos o que están experimentando un crecimiento en la creación de datos digitales, el aprendizaje autónomo que ha supuesto la creación del proyecto y la formación autónoma en materia de gobierno, gestión y calidad del dato.

Con todo, es importante tener en cuenta que este tipo de proyectos no es algo que tenga un final próximo, primeramente, debido a que los investigadores docentes están publicando investigaciones cada cierto tiempo, y, en consecuencia, por la actualización



de los datos. Este tipo de proyectos se deben mantener en el tiempo, es decir, no se trata de trabajos que tengan un final próximo, sino que, se tiene que ir actualizando y completando a medida que pasa el tiempo. Así, como bien señala el *Business Application Research Center*<sup>43</sup>, “*data governance is a highly complex, ongoing program*”, es decir, el gobierno del dato es un programa continuo muy complejo (Olavsrud, 2023). En consecuencia, es muy importante mantenerse al día, una vez la integración ya se haya realizado, siendo el equipo de datos mencionado anteriormente, los encargados de mantener los datos actualizados.

Finalmente, para concluir este Trabajo de Fin de Grado, cabe destacar el planteamiento de trabajos futuros con el objetivo de, por un lado, completar el proyecto aquí iniciado, abordando los puntos comprendidos en la sección Gobierno, Gestión y Calidad del dato y definiendo los procesos que sean necesarios. Por otro lado, se puede realizar una ampliación de los datos de producción científica, recogiendo, además de los artículos de revistas, actas de congreso, libros y capítulos de libros, tesis, direcciones de proyectos, patentes, etc. Así como también se podría seguir aumentando el alcance de modo que, en vez de abordar únicamente la *Universidade da Coruña*, llevar el proyecto de gestión del dato para la producción científica del personal docente e investigador al Sistema Universitario Gallego.

---

<sup>43</sup> BARC es una empresa de analistas de software la cual trabaja en las áreas de datos, inteligencia empresarial (BI) y análisis (BARC, 2023).



## 7. Referencias

- Apache Atlas. Data Governance Metadata framework for Hadoop. Apache Atlas. <https://atlas.apache.org/#/> (Consultado el 24 de abril de 2023).
- ARR. 2023. Investigación-acción. Universidad de Colima. El portal de la tesis. Universidad de Colima. [https://recursos.ucol.mx/tesis/investigacion\\_accion.php](https://recursos.ucol.mx/tesis/investigacion_accion.php) (Consultado el 13 de junio de 2023).
- Asociación Española de Normalización. 2023. *Calidad del dato*. Especificación UNE 0079:2023. Madrid: AENOR.
- Asociación Española de Normalización. 2023. *Gestión del dato*. Especificación UNE 0078. Madrid: AENOR.
- Asociación Española de Normalización. 2023. *Gobierno del dato*. Especificación UNE 0077. Madrid: AENOR.
- Atlan. Active Data Governance. Atlan. <https://atlan.com/active-data-governance/> (Consultado el 24 de abril de 2023).
- Azure. Gobierno de datos unificado con Microsoft Azure. Microsoft Azure. <https://azure.microsoft.com/es-es/products/purview> (Consultado el 24 de abril de 2023).
- BARC. About BRAC. BARC. <https://barc-research.com/about-us/> (Consultado el 5 de mayo de 2023).
- Bello, Elena. 2022. Data Governance: Definición, beneficios y cómo implementarla. *IEBS*, <https://www.iebschool.com/blog/data-governance-big-data/> (Consultado el 5 de mayo de 2023).
- Biblioguias UAM. Citas y elaboración de bibliografía: el plagio y el uso ético de la información: Estilo Chicago. Universidad Autónoma de Madrid. [https://biblioguias.uam.es/citar/estilo\\_chicago](https://biblioguias.uam.es/citar/estilo_chicago) (Consultado el 20 de abril de 2023).
- Caballero, Ismael, Moisés Rodríguez y Carlos Manuel Fernández. 2017. Calidad de datos digitales certificada. *AENOR: Revista de la normalización y la certificación*, 33: 36-39.



- Colmenares, Ana Mercedes y María Lourdes Piñeiro. 2008. La investigación acción. Una herramienta metodológica heurística para la comprensión y transformación de realidades y prácticas socioeducativas. *Revista de educación* 14, 27.
- Cortes Vargas, Daniel. 2007. Medir la producción científica de los investigadores universitarios: la bibliometría y sus límites. *Revista de la educación superior* 36, 142: 43-65. <https://www.scielo.org.mx/pdf/resu/v36n142/v36n142a3.pdf> (Consultado del 10 de mayo de 2023).
- Dudovskiy, John. Action Research. *Business Research Methodology*, <https://research-methodology.net/research-methods/action-research/> (Consultado el 13 de junio de 2023).
- Elsevier. 2023. Acerca de Elsevier. Elsevier <https://www.elsevier.com/es-es/about> (Consultado el 4 de junio de 2023).
- FECYT. 2023. Base de datos de Scopus. FECYT. <https://www.recursoscientificos.fecyt.es/licencias/productos-contratados/scopus> (Consultado el 4 de abril de 2023).
- FECYT. 2023. Base de datos Web of Science. FECYT. <https://www.recursoscientificos.fecyt.es/licencias/productos-contratados/wos> (Consultado el 4 de abril de 2023).
- Flores Nussy, Eddymar María, Jenny Macarena Méndez y Riczely Lissetch Mendoza. 2019. Producción científica como medio para la transformación social desde las universidades. *Revista Científica* 4, 14: 62-84.
- George, Tegan. 2023. *What Is Action Research? Definition & Examples*. Scribbr, <https://www.scribbr.com/methodology/action-research/> (Consultado el 13 de junio de 2023).
- Gómez Carretero, Ana Isabel y Mario Piattini Velthuis. 2018. Importancia de la calidad de los datos en la transformación digital. *RUIDERAE: Revista de universidades de Información* 13, 1-15. (Consultado el 15 de mayo de 2023).
- Heredia Rico, Jobany José y José Alberto Vilalta Alonso. 2009. La calidad de los datos: Su importancia para la gestión empresarial. *Libre empresa*, 11: 43-50.



- ISO/IEC 25012. 2008. *Software engineering- Software product Quality Requirement and Evaluation (SQuaRE). Data Quality model*. ISO/IEC 25012:2008.<https://www.iso.org/standard/35736.html> (Consultado el 3 de junio de 2023).
- Keyrus. Pasos para implementar un modelo de gobierno del dato de manera efectiva. Keyrus. <https://keyrus.com/sp/es/insights/pasos-para-implementar-un-modelo-de-gobierno-del-dato-de-manera-efectiva> (Consultado el 10 de mayo de 2023).
- KIO. 2023. ¿Qué es Data Management y por qué es importante?, KIO. <https://www.kionetworks.com/blog/data-center/que-es-data-management> (Consultado el 10 de mayo de 2023).
- Murillo, Danny, Dalys Saavedra y Erika Quintero. 2018. Extracción de datos de perfiles en Google Scholar utilizando un algoritmo en el lenguaje R para hacer minería de datos. *I+D Tecnológico 14*, 1: 94-104, <https://revistas.utp.ac.pa/index.php/id-tecnologico/article/view/1807/pdf> (Consultado el 06 de junio de 2023).
- Olavsrud, Thor. 2023. What is data fovernance? Best practices for managing data assets. CIO. <https://www.cio.com/article/202183/what-is-data-governance-a-best-practices-framework-for-managing-data-assets.html> (Consultado el 5 de mayo de 2023).
- Oracle. ¿Qué es IoT? Oracle. <https://www.oracle.com/es/internet-ofthins/what-is-iot/> (Consultado el 24 de abril de 2023).
- Piattini Mario, Moisés Rodríguez, Ismael Caballero y Carlos Manuel Fernández. 2018. Evaluación y certificación de la Calidad de Datos. *Oracleando. Revista de Spain Oracle Users Group*, 5-6
- PowerData. Big Data: ¿En qué consiste? Su importancia, desafíos y gobernabilidad. PowerData. <https://www.powerdata.es/big-data> (Consultado el 30 de abril de 2023).
- PowerData. 2014. El Ciclo de Vida en la Calidad de los Datos. PowerData. <https://n9.cl/calidadedatos> (Consultado el 30 de abril de 2023).

- Ramón Reyro, Eva y Ascensión Gil Martín. 2021. Desinformación e infoxicación, dos falsos sinónimos fronte a la estrategia de información de la Comisión Europea. *Comunicación y Hombre*, 17: 103-118.
- Rodríguez Moisés, Ana Isabel Gómez y Boris Delgado. 2018. Nueva certificación para el gobierno, gestión y calidad de los datos. *AENOR: Revista de la evaluación de la conformidad*, 340: 28-33.
- Tableau. About Tableau. Tableau. <https://www.tableau.com/> (Consultado el 10 de mayo de 2023).
- Talend a Qlik Comapany. Talend, A Complete Scalable Data Management Solution, Talend: A Leader in Data Integration & Data Integrity. Talend. <https://www.talend.com> (Consultado el 10 de mayo de 2023).
- Universidade da Coruña. 2015. Sistema Unificado de Gestión de la Investigación. Gestión Curricular. Procedimiento de exportación del Currículum Vitae en formato CVN. Universidade da Coruña. [https://cvn.fecyt.es/sites/default/files/Manual\\_Exportacion\\_CVN\\_UDC.pdf](https://cvn.fecyt.es/sites/default/files/Manual_Exportacion_CVN_UDC.pdf) (Consultado el 1 de junio de 2023).



## 8. Anexo

### 8.1. Glosario de términos

<b>Término</b>	<b>Definición</b>
Acta de Congreso	Tipo de publicación que recoge los temas establecidos en una conferencia, además también pueden ser resúmenes o informes de investigaciones presentadas en el congreso. Dichas actas se encuentran publicadas en una revista o libro
Arquitectos del dato.	Personal encargado de diseñar e implementar la infraestructura y los sistemas necesarios para gestionar y gobernar los datos dentro de una organización
Artículo	Texto escrito publicado en revistas digitales escritos por un investigador
Artículo científico	Texto escrito por un investigador en donde se expone el desarrollo de una investigación junto con sus respectivos resultados
Bibliometría	Ciencia que mide los productos y componentes que están relacionados directamente con la productividad o cantidad de documentos que elaboran los investigadores universitarios
<i>Big Data</i>	O datos masivos se refiere a un conjunto de datos de gran tamaño y complejidad
Calidad del dato	Características y requisitos impuestos para que los datos sean correctos, sigan unas mismas pautas y se mantengan actualizados
Característica del dato	Criterios que se han de seguir para analizar y evaluar la calidad del dato
Capítulo de libro	Parte o sección dentro de un libro escrita por un investigador
Congreso	Lugar en donde los investigadores presentan sus investigaciones con el objetivo de divulgar avances en la ciencia
Cuartil	Indicador o medida de posición de una revista en relación con todas las de su área



Dato	“Representación reinterpretable de la información de una manera formalizada adecuada para la comunicación, la interpretación, el procesamiento o el análisis” (Especificación UNE 0079:2023 2023, 37)
Director de calidad del dato.	Responsable de supervisar las políticas y procesos para garantizar la calidad de los datos en una organización
Director del dato	Líder ejecutivo responsable de la estrategia y gestión de los datos en una organización. Además, se encarga de supervisar todas las actividades relacionadas con el gobierno, gestión y calidad del dato
Director de seguridad de la información	Responsable de establecer y asegurar las medidas de seguridad de los datos en una organización
Director de protección de datos.	Responsable de garantizar el cumplimiento de las políticas y regulaciones de protección de datos en una organización
Director de tecnología	Responsable de gestionar la tecnología a utilizar dentro de una organización
Gobierno del dato	Encargado de desarrollar las políticas, roles, normativas, etc. que se han de seguir en la gestión del dato para implementar un proyecto de gobernanza del dato
Gestión del dato	Parte práctica del proyecto en donde se llevan a cabo las medidas, políticas, normativas, etc. impuestas durante el desarrollo del gobierno del dato
Índice H	Índice de los investigadores medido a partir de las citas recibidas en los artículos de estos
Investigador	Individuo de cualquier disciplina, cuyo trabajo es estudiar, recoger y analizar datos con el fin de obtener resultados que sean comunicados al resto de personas
Libro	Tipo de publicación especializada escrito por uno o más autores
Metadatos	Datos sobre datos
ORCID	Código alfanumérico único que identifica a un investigador

Producción científica	Conxunto de publicacións científicas, independentemente de su tipoloxía, de un investigador
Propietarios del dato.	Personal encargado de manipular y custodiar los datos en nombre de la universidad. Además, tienen la responsabilidad de tomar decisiones relacionadas con la gestión, uso y protección de datos
Publicación científica	Texto o comunicación científicos encargados de divulgar artículos científicos o información relativa a investigaciones en un determinado campo científico
Revista científica	Es un tipo de publicación, en la cual se recogen artículos de carácter científico para exponer investigaciones o novedades en dicho campo
Rol	Papel que tiene una persona o miembro de un equipo dentro de una organización
<i>SCimago Journal &amp; Country Rank (SJR)</i>	Indicador de la influencia de las revistas científicas encargado de medir el número de citas obtenidas en cada publicación en un periodo de 3 años
<i>Scraping</i>	Proceso de extracción de datos en la <i>web</i> a través de un sistema automático
<i>Journal Citation Reports (JCR)</i>	Indicador más antiguo y valorado encargado de mostrar la relevancia de la revista en su área correspondiente
Técnicos de sistemas y nuevas tecnoloxías.	Personal especializado en la implementación, mantenimiento y soporte de los sistemas informáticos y las tecnoloxías

Tabla 22. Glosario de términos