

UNIVERSIDAD DE VALLADOLID

MÁSTER UNIVERSITARIO

Ingeniería Informática



TRABAJO FIN DE MÁSTER

Marco Técnico-Legal de Adquisición, Confección y Publicación de un Corpus de Datos Biométricos de Sensores.

Realizado por **D. David Jiménez Bermejo**

Universidad de Valladolid

18 de julio de 2025

Tutores:

Dr. Cesar Llamas Bello

D. Jesús María Vegas Hernández

Universidad de Valladolid



Máster universitario en Ingeniería Informática

Dr. Cesar Llamas Bello, profesor del departamento de Informática.

Expone:

Que el alumno D. David Jiménez Bermejo, ha realizado el Trabajo final de Máster en Ingeniería Informática titulado "Marco Técnico-Legal de Adquisición, Confección y Publicación de un Corpus de Datos Biométricos de Sensores".

Y que dicho trabajo ha sido realizado por el alumno bajo la dirección del que suscribe, en virtud de lo cual se autoriza su presentación y defensa.

En Valladolid, 18 de julio de 2025

Vº. Bº. del Tutor:

Dr. Cesar Llamas Bello

Vº. Bº. del co-tutor:

D. Jesús María Vegas Hernández

Resumen

La actual era digital basa su potencial en la información y la información está basada en datos, por lo que actualmente los datos son la materia prima necesaria para cualquier estudio y para los avances en la sociedad digital. Gracias a la evolución tecnológica, cada vez tenemos a nuestra disposición más datos, más elementos para recoger datos y más capacidad de cómputo para procesar y analizar dichos datos. La realización de este proyecto nace con el fin de analizar y comprender las diferentes opciones que disponen los investigadores para almacenar los datos que recogen para la realización de sus investigaciones, las implicaciones legales que conlleva la recogida de dichos datos y el análisis de investigaciones publicadas para saber cuál podría ser la cantidad y variedad de datos necesarios a recoger para realizar una investigación, centrando dicho análisis en investigaciones relacionadas con la identificación de usuarios a través de IoT.

Descriptores

Inteligencia ambiental, Identificación de usuarios, Control de Acceso, IoT, repositorios de datos.

Abstract

The current digital age bases its potential on information and information is based on data, so data is currently the raw material needed for any study and for advances in the digital society. Thanks to technological evolution, we have more and more data at our disposal, more elements to collect data and more computing capacity to process and analyze said data. The realization of this project was born with the aim of analyzing and understanding the different options available to researchers to store the data they collect to carry out their research, the legal implications that the collection of said data entails and the analysis of published research to know what could be the quantity and variety of data necessary to collect to carry out an investigation, focusing said analysis on research related to the identification of users through IoT.

Keywords

Ambient Intelligence, User Identification, Access Control, IoT, data repository.

Índice General

Índice General.....	III
Índice de figuras	V
Índice de tablas.....	VI
1. Introducción	1
1.1. Motivación	2
1.2. Objetivos del proyecto	2
1.3. Organización del documento	3
2. Conceptos teóricos	5
3. Técnicas y herramientas	7
4. Marco Ético y Legal	9
4.1. Marco Ético	9
4.2. Marco Legal	11
4.3. Marco Legal Universitario	18
4.4. Resumen Legislación aplicable.....	19
4.5. Marco Legal Internacional	20
4.6. Otras referencias normativas.....	23
5. Repositorios de datos.....	25
6. Revisión de literatura.....	51
6.1. Estado del arte de investigaciones con datos biométricos	51
6.2. Trabajos relacionados	54
6.3. Selección y revisión de la literatura	56
6.3.1. Criterios de elección de la literatura	57
6.3.2. Análisis inicial	60
6.3.3. Búsqueda adicional de literatura	65
6.3.4. Artículos seleccionados	67

6.3.5. Riesgo de sesgo.....	68
6.4. Resultados obtenidos.....	69
6.4.1. Análisis de los resultados obtenidos	81
7. Conclusiones y líneas de trabajo futuras	83
7.1. Conclusiones	83
7.2. Líneas de trabajo futuras.....	87
Bibliografía	89
Apéndices	103
Apéndice I: Plantilla genérica de consentimiento informado	103
Apéndice II: Tabla de datos obtenidos de los artículos revisados	111
Apéndice III: PRISMA 2020	115
Diagrama de flujo de PRISMA 2020.....	124
Apéndice IV: Información adicional	127
Declaración de conflicto de intereses.....	127
Disponibilidad de datos.....	127

Índice de figuras

Ilustración 1 Distribución del número de autores por artículo	70
Ilustración 2 Distribución del número de instituciones por artículo	71
Ilustración 3 Distribución del número de citaciones por artículo	72
Ilustración 4 Distribución del número de citaciones por año de publicación.....	73
Ilustración 5 Distribución del número de palabras clave por artículo.....	74
Ilustración 6 Ratio diferenciación por edad.....	77
Ilustración 7 Formas de diferenciar la edad	77
Ilustración 8 Ratio diferenciación por género	78
Ilustración 9 Ratio Hombre/Mujer	78
Ilustración 10 Ratio Hombre/Mujer por artículo	79
Ilustración 11 Ratio diferenciación por procedencia/nacionalidad.....	79
Ilustración 12 Ratio otras diferenciaciones.....	80
Ilustración 13 Otras diferenciaciones encontradas	80

Índice de tablas

Tabla 1 Comparativa de tipo de usuario de Figshare	39
Tabla 2 Comparativa de repositorios de datos	47
Tabla 3 Años de publicación y entrada en vigor de las leyes de protección de datos	58
Tabla 4 Búsquedas sin filtros	63
Tabla 5 Búsquedas con el filtro "palabra clave"	64
Tabla 6 Búsquedas con filtro "palabra clave", accesibilidad y tipo de artículo	64
Tabla 7 Búsquedas adicionales	66
Tabla 8 Artículos seleccionados	68
Tabla 9 Listado de datos biométricos utilizados	69
Tabla 10 Distribución del número de autores por artículo	70
Tabla 11 Distribución del número de instituciones por artículo	71
Tabla 12 Distribución del número de citaciones por artículo	72
Tabla 13 Distribución del número de palabras clave por artículo	73
Tabla 14 Datos estadísticos	111
Tabla 15 Datos sobre participantes	113
Tabla 16 Marco legal y Marco ético	113
Tabla 17 Lista de verificación PRISMA 2020	124

1. Introducción

Cada vez que queremos acceder a un espacio privado o usar un servicio privado y/o personal, se está llevando a cabo sobre nosotros un proceso de identificación y autenticación. Unas veces es un proceso simple y que realizamos de forma natural, como usar las llaves para entrar en casa, donde el usuario que posee la llave es el que está autorizado a entrar, y otras veces el proceso es más complejo como puede ser autorizar una compra con el PIN de la tarjeta de crédito/débito o desbloquear el teléfono inteligente mediante la huella dactilar o el reconocimiento facial. Con estos sencillos ejemplos podemos ver que hay diferentes métodos para identificar y autenticar una persona para permitir el acceso. Este acceso puede ser físico para entrar a un edificio o sala desbloqueando una puerta o digital al permitir el usuario pueda entrar a sus servicios en línea como su cuenta de correo electrónico, plataformas de streaming, aplicaciones móviles, etc.

Para la verificación del usuario existen tres factores que se pueden usar, **algo que sabe el usuario** (como una contraseña o un PIN), **algo que tiene el usuario** (como una llave o una tarjeta inteligente) y **algo que es el usuario** (datos biométricos como huellas dactilares, voz, iris o forma de la cara).

Cada uno de estos sistemas tiene sus ventajas y sus desventajas y unas necesidades técnicas para su uso y ninguno de ellos es perfecto, pero todos tienen algo en común, los datos, todos utilizan datos de una forma u otra para realizar la validación.

Con la aparición de nuevas tecnologías y su posterior democratización, se amplía la cantidad y variedad de datos que se pueden recoger, almacenar y procesar para todo tipo de tareas. Una de esas tareas es la identificación de usuarios que cada vez se está realizando en mayor medida con datos biométricos por considerarse más seguros.

En este contexto de recolección y uso de datos biométricos, hace que no solo tengamos que evaluar las capacidades técnicas para su uso, si no que nos hace preguntarnos por los aspectos legales y éticos que conlleva su uso.

En este contexto complejo, se ha desarrollado este proyecto para ver cuál es el estado del arte de las investigaciones que usan datos biométricos para la identificación de usuarios, además de analizar el contexto y las implicaciones legales y éticas que tiene el uso de este tipo de datos.

El proyecto se complementa con un estudio sobre los repositorios de datos existentes, que facilitan el almacenaje y acceso a los datos de las diferentes investigaciones científicas que se realizan, facilitando la consecución de la ciencia replicable y la ciencia abierta.

También se desarrollan una serie de anexos, en los cuales se redactará un modelo de consentimiento informado para la cesión de datos por parte de los participantes en proyectos de investigación de acuerdo a las necesidades legales aplicables.

1.1. Motivación

Este proyecto surge con el fin de recopilar información sobre las investigaciones que se llevan a cabo para identificar usuarios a través de datos biométricos recogidos mediante sensores IoT. A través de esta información recogida se quiere realizar un análisis para determinar el estado del arte y poder obtener unas conclusiones sobre que tipos de datos se recogen, como se recogen o cuantos usuarios participan en las pruebas para poder establecer unos patrones o unos parámetros mínimos a la hora de realizar futuras investigaciones.

1.2. Objetivos del proyecto

Con la realización de este trabajo de fin de máster, se quieren conseguir diferentes objetivos, acorde a cada uno de los apartados que se han analizado y estudiado. Estos objetivos son los siguientes:

1. Realizar un estado del arte de las investigaciones que usan datos de carácter biométrico procedentes de la interacción humana con objetos de la vida cotidiana.

- a. Averiguar las palabras clave más usadas.
 - b. Estudiar las relaciones existentes entre autores, grupos de investigación, universidades y entidades que realizan las investigaciones.
2. Analizar el marco legal competente sobre la recolección y uso de datos de carácter biométrico.
 3. Analizar el marco ético sobre el que deben regirse las investigaciones que hacen uso de datos de carácter biométrico.
 4. Analizar diferentes repositorios de datos existentes para la publicación de datos de carácter biométrico.
 5. Crear un marco de análisis y comparación para corpus de datos analizados.

1.3. Organización del documento

El presente documento se estructura de la siguiente forma:

- **Capítulo 1: Introducción:** Este capítulo introduce el problema que se aborda en el trabajo, aportando el contexto del problema, la justificación de la realización del trabajo y los objetivos que se pretenden conseguir.
- **Capítulo 2: Conceptos teóricos:** En este capítulo se definen una serie de conceptos que se utilizan a lo largo del documento. Estas definiciones se proporcionan para facilitar la comprensión de los conceptos y del documento.
- **Capítulo 3: Técnicas y herramientas:** Este capítulo está dedicado a la explicación de las diferentes técnicas que se han utilizado para recopilar la información y las herramientas que se han utilizado para el análisis de dicha información
- **Capítulo 4: Marco Ético y legal:** Este capítulo está dedicado al estudio de la ética que rige la investigación y el marco legal que atañe a la investigación que hace uso de datos biométricos.
- **Capítulo 5: Repositorios de datos:** Este capítulo se destina al análisis de los repositorios de datos donde se publican, almacenan y comparten los datos obtenidos en la fase de adquisición de datos de las investigaciones.

- **Capítulo 6: Revisión de literatura:** En este capítulo se concentran todas las labores que se han realizado sobre el análisis de la literatura existente, exponiendo el estado del arte, un análisis de trabajos relacionados, el proceso seguido para la selección, revisión y análisis de los artículos que crean su propio corpus de datos y una discusión sobre los resultados obtenidos.
- **Capítulo 7: Conclusiones y líneas de trabajo futuras:** En este capítulo se exponen las conclusiones a las que se han llegado en el trabajo y se proponen diferentes puntos en los que se puede seguir investigando.
- **Apéndices:** En esta sección del documento se proporciona diversa información adicional pero igualmente útil para la comprensión de todas las dimensiones estudiadas durante la realización del proyecto

2. Conceptos teóricos

Para la correcta comprensión de este trabajo es necesario aclarar una serie de conceptos teóricos, para que no haya dudas en la interpretación. Dichos conceptos son:

- **Datos de carácter biométrico:** Son datos personales relativos a las características físicas, fisiológicas o conductuales de las personas físicas que permitan la identificación o autenticación de una persona. Estos datos se generan a partir de características únicas y medibles del cuerpo humano o de su comportamiento y se recogen mediante diferentes sensores de IoT. Como explican Jain et al. [1], algunos datos biométricos que se pueden usar son reconocimiento facial, huella dactilar, el iris o la voz, pero también se pueden usar otro tipo de datos o la huella que producen nuestros dispositivos móviles al atravesar una red inalámbrica como puede ser una red WiFi.
- **Ética de la investigación científica:** Exige que la práctica de la ciencia se realice conforme a principios éticos que aseguren el avance del conocimiento, la comprensión y mejora de la condición humana y el progreso de la sociedad, respetando a la dignidad del ser humano, a la autonomía de su voluntad, protección de sus datos - privacidad, confidencialidad -, bienestar animal y preservación del medio ambiente. [2]
- **Repositorio de datos:** Es un espacio centralizado donde se almacena, organiza, mantiene y difunde información digital [3]. Según C. Ortega [4], puede entenderse como una gran infraestructura de bases de datos o varias bases de datos que recogen, gestionan y almacenan conjuntos de datos para su análisis, intercambio y elaboración de informes. Los datos pueden recogerse y almacenarse de diferentes maneras, como los datos agregados, que suelen recogerse de múltiples fuentes o segmentos de una empresa, y luego pueden almacenarse de forma estructurada o no estructurada y, posteriormente, pueden etiquetarse con diferentes metadatos.
- **Ciencia abierta:** Es el movimiento que promueve el acceso abierto a la investigación científica, incluidas las publicaciones, los datos, metodología,

código, entre otros, de manera que éstos sean accesibles a todos los niveles de la sociedad, tanto para aficionados como para profesionales [5].

- **Ciencia replicable:** También conocida como ciencia reproducible, es ir un paso más allá de la ciencia abierta. Consiste en detallar todos los pasos y parámetros de la investigación científica para que la sociedad pueda replicar la investigación y así confrontar los datos obtenidos en la investigación inicial con la investigación realizada por la sociedad. Es necesario que además de los datos utilizados, se comparta el software empleado, el hardware utilizado detallando como se han montado los prototipos, tanto a nivel físico como de configuración, etc. como expone Rodríguez-Sánchez et al. [6].
- **Intimidad:** Ámbito íntimo y más reservado de una persona o de un grupo de ellas.
- **Privacidad:** Ámbito de la vida privada.
- **Servicios de la sociedad de la información:** todo servicio prestado normalmente a cambio de una remuneración, a distancia, por vía electrónica y a petición individual de un destinatario de servicios [7].
- **Seudonimización:** el tratamiento de datos personales de manera tal que ya no puedan atribuirse a un interesado sin utilizar información adicional, siempre que dicha información adicional figure por separado y esté sujeta a medidas técnicas y organizativas destinadas a garantizar que los datos personales no se atribuyan a una persona física identificada o identifiable [8].

3. Técnicas y herramientas

Dada la naturaleza del trabajo propuesto, centrado en una revisión de la literatura, las herramientas utilizadas han sido un ordenador personal, conexión a internet, acceso a diferentes bases de datos en línea de artículos científicos gracias al acceso que provee la universidad o de acceso abierto y software ofimático, donde se ha utilizado el paquete office que está disponible como alumno de la Universidad de Valladolid. Para la gestión de referencias bibliográficas se ha usado la herramienta Zotero [9] que permite integración con Microsoft Word y con diferentes navegadores para la adquisición y uso de las referencias bibliográficas.

En cuanto a las técnicas utilizadas para la búsqueda de información de cara a desarrollar los diferentes apartados del trabajo de fin de máster, hay que señalar las siguientes técnicas:

- Buscar, leer y comprender las diferentes leyes que afectan a la recolección y el uso de datos.
- Buscar las leyes que tienen implicación en la ética de las investigaciones.
- Buscar los diferentes organismos y/o instituciones que velan por el cumplimiento de la ética en las investigaciones.
- Buscar en la literatura existente artículos que hablen sobre la ética de la investigación, comparen leyes de protección de datos de diferentes países y que utilicen datos biométricos en las investigaciones para obtener el estado del arte actual y encontrar si hay artículos similares a este Trabajo de Fin de Master. Para ello se han utilizado herramientas de búsqueda de artículos de investigación:
 - Consensus.app [10].
 - Google Scholar [11].
- Buscar diferentes repositorios de datos existentes y analizar cuál es su objetivo principal y las condiciones para publicar en ellos.
- Buscar, leer y analizar estudios ya realizados que usen datos de carácter biométrico procedentes de la interacción humana con objetos de la vida cotidiana. En el punto 6.3.1 Criterios de elección de la literatura se explican detalladamente los criterios seguidos, pero como resumen, se ha usado la

técnica de búsquedas de palabras clave en las bases de datos de artículos de investigación ([12], [13], [14]) y utilizar los artículos que se obtienen en los resultados. Las palabras clave utilizadas han sido:

- IoT
- Ambiental Intelligence / Ambient Intelligence / Aml
- User Identification
- Access Control
- Human Activity Recognition Systems / HARS

En este punto es necesario hablar de la herramienta o enfoque PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) [15] que aporta una serie de directrices que ayudan en las revisiones sistemáticas de literatura, mejorando la presentación del método utilizado y de los resultados obtenidos. El gran aporte de PRISMA es una lista de verificación de puntos para asegurarse de que está toda la información relevante y un diagrama de flujo para una sencilla visualización del proceso. Esta información propuesta por PRISMA para este trabajo se muestra en el Apéndice III: PRISMA 2020.

En cuanto a las herramientas concretas utilizadas han sido las siguientes:

- Ordenador personal.
- Conexión a internet.
- Revistas científicas accesibles gracias a las credenciales de la universidad como pueden ser:
 - ScienceDirect.
 - IEEE Xplore.
 - MDPI.
- Lector de PDF, concretamente el programa gratuito Adobe Acrobat Reader.
- Software Ofimático, en más detalle se ha usado Microsoft Word para la realización del presente documento y Microsoft Excel para la agrupación de datos, su posterior análisis y realización de gráficas. Ambos programas se obtienen gracias a la licencia proporcionada por la universidad.
- Herramienta Zotero para la gestión de referencias bibliográficas.

4. Marco Ético y Legal

Esta sección está dedicada al estudio tanto del marco ético como del legal a tener en cuenta a la hora de realizar investigaciones científicas con datos biométricos.

4.1. Marco Ético

Hay que tener en cuenta el marco ético en las investigaciones y en los datos de las investigaciones para equilibrar la investigación científica con los derechos y valores de la sociedad, los seres humanos, animales, otros seres vivos o agentes biológicos para que el proceso científico e investigador no atente contra esos derechos y valores.

Cada vez hay que tener más en cuenta este aspecto ético, ya que se imponen exigencias éticas, metodológicas y jurídicas en convocatorias de investigación, en revistas científicas para la publicación de artículos [16], en congresos y jornadas para la presentación de investigaciones [17] y en general a una serie de normas de obligado cumplimiento para actividades de investigación. Algunas de estas normas o leyes son la Declaración de Helsinki [18] de la Asociación Médica Mundial, la Ley 14/2007, de 3 de julio, de Investigación Biomédica, el Real Decreto 53/2013, de 1 de febrero, por el que se establecen las normas básicas aplicables para la protección de los animales utilizados en experimentación y otros fines científicos, incluida la docencia, o el Real Decreto 178/2004, de 30 de enero, por el que se aprueba el Reglamento general para el desarrollo y ejecución de la Ley 9/2003 de 25 de abril, por la que se establece el régimen jurídico de la utilización confinada, liberación voluntaria y comercialización de organismos modificados genéticamente.

Por este motivo los centros de investigación y las universidades han creado los Comités de Ética de la investigación. El que nos atañe en este trabajo, es el Comité Ético de Investigación de la Universidad de Valladolid [19] y su Código de Buenas Prácticas en Investigación (CBPI) [20]. Dicho Comité Ético de Investigación se encarga de controlar el cumplimiento de los principios deontológicos de las

diferentes actividades para evitar conductas inadecuadas y su corrección inmediata. También se encarga de vigilar y controlar los aspectos éticos de las investigaciones, especialmente si la investigación se aplica a seres humanos o la utilización de muestras de origen humano, la obtención y tratamiento de datos de carácter personal y que puedan afectar a los derechos fundamentales. Por lo tanto, ya que en las investigaciones que se están analizando se usan datos de carácter personal y en algunos casos de carácter biométrico, se tiene que obtener la verificación positiva del Comité Ético de Investigación y garantizar el uso correcto de dichos datos.

La Universidad de Valladolid tiene varios comités de ética que son específicos para temas concretos, en el contexto de este trabajo donde se tratan muestras biométricas el comité específico es el "Comité de Ética en la Investigación con Seres Humanos, sus datos o muestras" (CEISH) [21], constituido en Diciembre de 2024. Hay que enviar al comité, mediante el correo electrónico comites.eticos@uva.es, con carácter previo al inicio de la investigación de la siguiente información sobre la investigación:

- Formulario de Evaluación para Proyectos de Investigación CEISH UVa [22].
- El o los consentimientos informados pertinentes.
- Compromiso de confidencialidad de todos los investigadores.
- Memoria completa.
- Instrumentos de recogida de datos: cuestionarios, entrevistas, etc.
- Declaración de conflicto de intereses.

En su página web hay accesibles modelos de los documentos necesarios y también se especifica que se deben cumplir los requerimientos de protección de datos UVa [23].

Además de este consentimiento por el Comité Ético, hay otras consideraciones que podemos incluir dentro del marco ético, y hacer referencia a ellas en la publicación del corpus de datos, ya que podrían influir en el resultado de la investigación. Estos puntos a tener en cuenta son:

- **Financiación de la investigación:** Este dato puede llevarnos a pensar si el estudio puede estar sesgado ya que puede haber diferencias entre estudios realizados por doctores de una universidad en el ejercicio de su función

investigadora y estudios realizados por una empresa privada que busque obtener un beneficio futuro.

- **Conflictos de intereses:** Esta información sirve para identificar a simple vista si se ha producido algún tipo de conflicto en la realización de la investigación que pudiera interferir en el resultado.
- **Publicación de los datos:** Si un estudio indica todo el proceso de como se ha realizado la investigación y comparte los datos utilizados permitiendo que otros investigadores repliquen el estudio, esto pone de manifiesto que se prestan a que cualquier persona puede comprobar y verificar su estudio, dando a entender que todo es ético y correcto y no hay ningún punto que esconder. Para la publicación de los datos, hay que tener en cuenta el marco legal y realizar un paso previo de anonimizar o codificar los datos para preservar la intimidad de las personas.
- **Motivos de exclusión:** Si se ha producido la exclusión de algún dato, se necesitaría explicar que datos se han excluido y el motivo, para que en la revisión y corrección por pares se pueda determinar si dicha exclusión puede sesgar y falsear el resultado de la investigación.

4.2. Marco Legal

Este capítulo está enfocado a analizar la normativa aplicable en relación a la protección de datos de los voluntarios que participan en los proyectos de investigación. En dicho análisis se tendrá en cuenta la normativa española y la europea en orden cronológico.

- **Convenio Europeo para la Protección de los Derechos Humanos y de las Libertades Fundamentales:** El Convenio Europeo para la Protección de los Derechos Humanos y de las Libertades Fundamentales [24] de 4 de noviembre de 1950 y ratificado en España en el BOE de 10 de octubre de 1979, en su artículo 8 establece “Toda persona tiene derecho al respeto de su vida privada y familiar, de su domicilio y de su correspondencia” [24],

donde se reconoce el derecho a la vida privada, que puede entenderse como un derecho a la protección salvando la brecha temporal y de avances tecnológicos que separan esa fecha de la situación actual.

- Constitución Española: La Constitución Española de 1978 [25] incluyó el derecho a la protección de la intimidad personal, lo que conlleva el derecho a la protección de datos. Según se recogen en los artículos 18.1 y 18.4 donde se reconoce el derecho a la intimidad personal y a limitar el uso de la informática para garantizar dicho derecho.
- Convenio para la protección de las personas con respecto al tratamiento automatizado de datos de carácter personal: El Convenio para la protección de las personas con respecto al tratamiento automatizado de datos de carácter personal [26] de 28 de enero de 1981 promulgado por el Consejo de Europa, fue el primer instrumento internacional jurídicamente vinculante adoptado en el ámbito de la protección de datos [27]. Fue ratificado por España y dio lugar a la Ley Orgánica de Regulación del Tratamiento Automatizado de los Datos donde se atienden sus exigencias y previsiones que para la protección de datos.
- Ley Orgánica 5/1992 de Regulación del Tratamiento Automatizado de los Datos: La Ley Orgánica de Regulación del Tratamiento Automatizado de los Datos (LORTAD) [28] 5/1992, de 29 de octubre, de regulación del tratamiento automatizado de los datos de carácter personal (actualmente derogada). Es una ley de ámbito nacional que amplía el término de intimidad de la constitución al término privacidad. Esta Ley surge para “implantar mecanismos cautelares que prevengan las violaciones de la privacidad que pudieran resultar del tratamiento de la información” y para “limitar el uso de la informática y otras técnicas y medios de tratamiento automatizado de los datos de carácter personal para garantizar el honor, la intimidad personal y familiar de las personas físicas y el pleno ejercicio de sus derechos.” Bajo el mandato de esta ley se creó Agencia Española de Protección de Datos (AEPD).
- Real Decreto 1332/1994: El Real Decreto 1332/1994 [29], de 20 de junio de 1994 sirve para regular los artículos 15.1, 16.1, 17.1, 24.2, 38.3 y 47.1 de la Ley Orgánica 5/1992, concretando información sobre los ficheros de datos

y los derechos de los ciudadanos sobre dichos ficheros de datos, además de definir un procedimiento sancionador.

- Directiva 95/46/CE del Parlamento Europeo y del Consejo: La Directiva 95/46/CE del Parlamento Europeo y del Consejo [30] de 24 de octubre de 1995, es una directiva de ámbito europeo que establece una serie de disposiciones en relación a la “protección de las libertades y de los derechos fundamentales de las personas físicas, y, en particular, del derecho a la intimidad, en lo que respecta al tratamiento de los datos personales” [30]. Esta directiva estableció un marco legal común en los países miembros de la Unión Europea para proteger a los ciudadanos en lo que respecta al tratamiento de sus datos y la libre circulación de los mismos. A partir de esta directiva cada estado adaptó su legislación de protección de datos, dando como resultado en España la Ley Orgánica 15/1999, de Protección de Datos de Carácter Personal de Protección de Datos (LOPD) [31]. Los hitos más destacados de este decreto son:
 - Los ciudadanos obtienen una serie de derechos sobre sus datos personales como el derecho de acceso, de rectificación, cancelación y oposición (ARCO), por los cuales pueden controlar como se utilizan sus datos y solicitar correcciones o borrados.
 - Las empresas deben seguir la normativa para el tratamiento de los datos personales, lo que incluye un consentimiento explícito para procesar y almacenar los datos y garantizar que se manejan de manera segura.
 - El gobierno debe ejercer una supervisión sobre la protección de datos e imponer sanciones en caso de incumplimiento. En España se nombró a la Agencia Española de Protección de Datos (AEPD) para llevar a cabo esta labor.

Se deroga el 25 de mayo de 2018 por el reglamento 2016/679 que veremos más adelante.

- Real Decreto 994/1999 Reglamento de medidas de seguridad LOPD: El Real Decreto 994/1999, por el que se aprueba el Reglamento de medidas de seguridad de los ficheros automatizados que contengan datos de carácter personal [32], de 11 de junio es una ley de ámbito nacional que “tiene por objeto el desarrollo de lo dispuesto en los artículos 9 y 43.3.h) de la Ley

Orgánica 5/1992. El Reglamento determina las medidas de índole técnica y organizativa que garanticen la confidencialidad e integridad de la información con la finalidad de preservar el honor, la intimidad personal y familiar y el pleno ejercicio de los derechos personales frente a su alteración, pérdida, tratamiento o acceso no autorizado" [32]. Este Real Decreto es una ampliación de la Ley Orgánica 5/1992 y nos sirve para darnos cuenta de que según avanza la sociedad y la técnica se hace necesario adaptar la legislación para no perder garantías ni seguridad sobre los datos.

- Ley Orgánica 15/1999 de Protección de Datos de Carácter Personal: La Ley Orgánica 15/1999, de Protección de Datos de Carácter Personal de Protección de Datos (LOPD) [31], de 13 de diciembre de 1999 derogó la anterior Ley Orgánica de Regulación del Tratamiento Automatizado de los Datos (LORTAD). Es una ley de ámbito nacional, que al amparo de la Directiva 95/46/CE del Parlamento Europeo y del Consejo y basándose en el artículo 18 de la Constitución Española de 1978, aportó un marco legal para garantizar y proteger las libertades públicas y los derechos fundamentales de las personas en lo que respecta al tratamiento de sus datos personales. El artículo 4.2 de esta ley "Los datos de carácter personal objeto de tratamiento no podrán usarse para finalidades incompatibles con aquellas para las que los datos hubieran sido recogidos. No se considerará incompatible el tratamiento posterior de éstos con fines históricos, estadísticos o científicos" es importante para los investigadores ya que habilita el uso de los datos para usos científicos.
- Directiva 2000/31/CE de Comercio Electrónico: La Directiva 2000/31/CE del Parlamento Europeo y del Consejo, de 8 de junio de 2000 o Directiva sobre el comercio electrónico [33] de 17 de julio de 2000, es una directiva de ámbito europeo que surge con el objetivo de homogeneizar las legislaciones de la Unión Europea en lo referente al comercio electrónico y los servicios de la sociedad de la información eliminando los obstáculos que existían en el mercado interior europeo. Gracias a esta directiva se obtienen garantías en la protección de datos en el comercio electrónico.
- Ley 34/2002, de servicios de la sociedad de la información y de comercio electrónico: La Ley 34/2002, de servicios de la sociedad de la información y

de comercio electrónico (LSSI o LSSICE) [34] , de 11 de julio es una ley de ámbito nacional con la que se incorpora al ordenamiento jurídico español de la Directiva 2000/31/CE, del Parlamento Europeo y del Consejo. Con esta ley se regula “lo referente a las obligaciones de los prestadores de servicios incluidos los que actúan como intermediarios en la transmisión de contenidos por las redes de telecomunicaciones, las comunicaciones comerciales por vía electrónica, la información previa y posterior a la celebración de contratos electrónicos, las condiciones relativas a su validez y eficacia y el régimen sancionador aplicable a los prestadores de servicios de la sociedad de la información” [34] para garantizar la protección de datos en el comercio electrónico y la contratación electrónica entre otros objetivos.

- Ley 41/2002 De Autonomía del Paciente: La Ley 41/2002, básica reguladora de la autonomía del paciente y de derechos y obligaciones en materia de información y documentación clínica [35], de 14 de noviembre, es una ley de ámbito nacional que estable determinados derechos para los pacientes en las relaciones clínico-asistenciales. En su artículo 7 hace referencia a la intimidad y confidencialidad de los datos. Esta ley nos sirve para darnos cuenta de la importancia de la intimidad y confidencialidad de los datos, en todos los ámbitos, pero especialmente en las situaciones con datos más sensibles.
- Real Decreto 1720/2007 Reglamento de desarrollo de la Ley Orgánica de Protección de Datos: El Real Decreto 1720/2007 de 21 de diciembre de 2007 derogó el Real Decreto 1332/1994 y el Real Decreto 994/1999, de medidas de seguridad de los ficheros automatizados que contengan datos personales y desarrolló la Ley Orgánica 15/1999. Algunos de los artículos más importantes de este Real Decreto son el artículo 8 sobre la calidad de los datos, el artículo 9 que hace compatibles los datos con usos científicos, artículo 14 sobre el consentimiento para el tratamiento de los datos, título III sobre los derechos de acceso, rectificación, cancelación y oposición y el título VIII sobre las medidas de seguridad en el tratamiento de datos. Este Real Decreto es de gran importancia para asegurar que el tratamiento y el almacenamiento de los datos personales se hace con garantías de calidad y seguridad.

- Ley 14/2011, de 1 de junio, de la Ciencia, la Tecnología y la Innovación [36], es un ley de ámbito nacional, que derogó la Ley de Investigación Científica y Tecnológica de 1986 y su principal objetivo es la consolidación de un marco para el fomento de la investigación científica y técnica y sus instrumentos de coordinación general para contribuir al desarrollo económico sostenible y al bienestar social mediante la generación, difusión y transferencia del conocimiento y la innovación. Un punto importante para este trabajo es el artículo 37 que trata sobre la ciencia abierta e indica que los agentes públicos del Sistema Español de Ciencia, Tecnología e Innovación así como así como el personal de investigación del sector público o cuya actividad investigadora esté financiada mayoritariamente con fondos públicos deben hacer que las publicaciones científicas, datos, códigos y metodologías, estén disponibles en acceso abierto y en repositorios institucionales. También indica que hay que cumplir los principios internacionales FAIR (sencillos de encontrar, accesibles, interoperables y reutilizables) y promueve el desarrollo de infraestructuras y plataformas abiertas para fomentar la publicación de los resultados científicos en acceso abierto, y la participación abierta de la sociedad civil en los procesos científicos.
- Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo: El Reglamento (UE) 2016/679 o Reglamento General de Protección de Datos (RGPD) [8], publicado el 4 de mayo de 2016 y entró en vigor el 25 de mayo de 2018. Es una ley de ámbito europeo que sustituye a la Directiva 95/46/CE y tiene como objetivo mejorar el derecho de las personas sobre sus datos personales y el control que se hace sobre dichos datos. También aumenta el alcance de esta protección al incluir a todas las compañías de fuera de la UE siempre que se traten datos de personas residentes dentro de la UE. Este reglamento dejó obsoleta la anterior Ley Orgánica 15/1999, de Protección de Datos de Carácter Personal (LOPD).
- Directiva (UE) 2016/680 del Parlamento Europeo y del Consejo (2016): La Directiva (UE) 2016/680 o Directiva sobre protección de datos en el ámbito penal [37], de 4 de mayo de 2016, es una directiva de ámbito europeo cuyo objetivo es garantizar la protección de los datos personales de las personas implicadas en procesos penales. Esta directiva nos sirve como ejemplo para

saber que ciertos tipos especiales de datos, por su naturaleza, requieren una legislación concreta para ellos.

- Ley Orgánica 3/2018 de Protección de Datos Personales y Garantía de los Derechos Digitales: La Ley Orgánica 3/2018, de Protección de Datos Personales y garantía de los derechos digitales (LOPD-GDD) [38], de 5 de diciembre de 2018, es una ley orgánica de ámbito nacional que sirve para adaptar a la legislación española el Reglamento General de Protección de Datos (RGPD) o Reglamento (UE) 2016/679. Es ley derogó la Ley Orgánica 15/1999, de Protección de Datos de Carácter Personal (LOPD). La disposición adicional decimoséptima, aunque versa sobre datos de salud, la podemos tomar referencia y extrapolarlo a investigaciones científicas fuera del ámbito de la salud, para preservar la protección de datos. En esta disposición los puntos más destacables son la necesidad de un consentimiento para el uso de datos, la anonimización o seudonimización de los datos, el “compromiso expreso de confidencialidad y de no realizar ninguna actividad de reidentificación”, la investigación “deberá ser sometida al informe previo del comité de ética de la investigación” y “los comités de ética de la investigación ... deberán integrar entre sus miembros un delegado de protección de datos o, en su defecto, un experto con conocimientos suficientes del Reglamento (UE) 2016/679 cuando se ocupen de actividades de investigación que comporten el tratamiento de datos personales o de datos seudonimizados o anonimizados”.
- Reglamento (UE) 2022/868 del Parlamento Europeo y del Consejo de 30 de mayo de 2022 relativo a la gobernanza europea de datos (Reglamento de Gobernanza de Datos) [39], de 3 de junio de 2022, es un reglamento que crea un marco legal para facilitar la reutilización y el intercambio de datos dentro de organismos públicos, además de favorecer el alturismo de los datos mediante la cesión de estos de forma voluntaria por los individuos y organizaciones para el interés general. Este reglamento no introduce nuevas normas sobre el tratamiento de datos y en caso de duda siempre prevalece lo que dicta el RGPD. El reglamento quiere favorecer el uso de los datos para la investigación, pero siempre hay que respetar y cumplir lo determinado por el RGPD.

4.3. Marco Legal Universitario

Además de la legislación vista hasta ahora, las universidades tienen también leyes concretas que tienen que cumplir.

Actualmente las universidades españolas dependen del Ministerio de Ciencia, Innovación y Universidades y se rigen por la Ley Orgánica 2/2023, de 22 de marzo, del Sistema Universitario (LOSU) [40]. Esta ley orgánica, en su Título IV “Investigación y transferencia e intercambio del conocimiento e innovación” aborda lo relativo a la investigación, la transferencia e intercambio del conocimiento y la innovación, por lo que se debe estudiar su alcance en esta sección del presente trabajo. De los artículos que componen el título IV de la ley podemos destacar los siguientes puntos:

- Artículo 11. Normas generales:
 - Punto 1: “La investigación es una de las funciones fundamentales de las universidades.”
 - Punto 2: “La investigación, al igual que la docencia, es un derecho y un deber del personal docente e investigador.”
 - Punto 6: “Las actividades de investigación, y de transferencia e intercambio del conocimiento e innovación realizadas por el personal docente e investigador se considerarán conceptos evaluables a efectos retributivos y de promoción.”
- Artículo 12. Fomento de la Ciencia Abierta y Ciencia Ciudadana:
 - Punto 1: “El conocimiento científico tendrá la consideración de un bien común. Las Administraciones Públicas y las universidades promoverán y contribuirán activamente a la Ciencia Abierta mediante el acceso abierto a publicaciones científicas, datos, códigos y metodologías que garanticen la comunicación de la investigación.”
 - Punto 2: “El personal docente e investigador deberá depositar una copia de la versión final aceptada para publicación y los datos asociados a la misma en repositorios institucionales o temáticos de acceso abierto, de forma simultánea a la fecha de publicación.”

- Punto 3: "La versión digital de las publicaciones académicas se depositará en los repositorios institucionales, sin perjuicio de otros repositorios de carácter temático o generalista."
- Punto 5: "Los datos, entendidos como aquellas fuentes primarias necesarias para validar los resultados de las investigaciones, deberán seguir los principios FAIR (datos fáciles de encontrar, accesibles, interoperables y reutilizables) y, siempre que sea posible, difundirse en acceso abierto."

4.4. Resumen Legislación aplicable

Como se puede observar, el marco legal en lo referente a la protección de datos tiene una larga trayectoria a lo largo del tiempo y ha ido evolucionando desde menciones como la vida privada y la intimidad hasta las complejas y extensas leyes actuales basadas en directivas de la Unión Europea. Por este motivo para la realización de investigaciones científicas que requieran del uso de datos es conveniente contar con un experto en protección de datos que posea amplios conocimientos sobre en la última legislación vigente.

En lo referente a este trabajo, la vigente ley sobre protección de datos que se tiene que tener en cuenta es el Reglamento General de Protección de Datos (RGPD) [8], que es el reglamento europeo relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y la Ley Orgánica de Protección de Datos Personales y Garantía de los Derechos Digitales (LOPDGDD) [41], que es la ley española que adapta el ordenamiento jurídico español al RGPD y completa sus disposiciones. Además, se puede hacer uso de la Guía sobre tratamientos de presencia mediante sistemas biométricos [42] publicada por Agencia Española de Protección de Datos (AEPD) [43] donde explica el tratamiento que hay que hacer de los datos para los usos de identificación y autentificación y por lo tanto, el control de presencia y el control de acceso mediante datos biométricos.

A raíz de la legislación vigente, se consideran los siguientes puntos clave:

- Según el principio de minimización de datos, al tratarse de una investigación de técnicas biométricas, los datos biométricos forman parte de la finalidad y por lo tanto son necesarios.
- Los datos biométricos en este caso son de categoría especial al poder identificar a una persona física con ellos.
- La persona debe dar un consentimiento expreso que debe ser libre, específico, informado e inequívoco.
- Se deben anonimizar o seudonimizar los datos de forma que no sea posible una reidentificación.
- La investigación debe estar validada por un comité de ética.
- Una copia del corpus de datos debe estar publicada en un repositorio institucional de acceso abierto y que cumpla los principios FAIR.

En el Apéndice I: Plantilla genérica de consentimiento informado, se muestra un ejemplo de documento de consentimiento por parte de los participantes en la investigación para la cesión de sus datos.

4.5. Marco Legal Internacional

La protección de datos no es algo que solo tenga importancia en Europa, es una preocupación global que lleva tiempo preocupando a los dirigentes mundiales. Prueba de ello son las menciones a la privacidad que se mencionan en la Carta Internacional de Derechos Humanos, compuesta principalmente por la Declaración Universal de los Derechos Humanos (DUDH) [44], el Pacto Internacional de Derechos Civiles y Políticos (PIDCP) [45] y el Pacto Internacional de Derechos Económicos, Sociales y Culturales (PIDESC) [46]. Aunque no son tan específicas y concretas como las leyes actuales, hablan sobre la protección a las personas y su vida privada. Más concretamente el artículo 12 de la DUDH y el artículo 17 del PIDCP.

Recientemente debido a la cada vez más presente globalización de la economía y de toda la sociedad, la privacidad y la protección de datos toman mayor importancia y se expande a ámbitos supranacionales, como demuestra el RGPD.

Aunque en este trabajo nos vamos a centrar en la legislación de ámbito europeo, es conveniente nombrar algunas de las legislaciones de otros países fuera de Europa y hacer una breve comparativa entre dichas regulaciones y la RGPD europea. Por motivos de importancia económica, capacidad investigadora, cantidad de población o importancia histórica, algunas de las leyes de protección de datos a nivel global son:

- **Ley de Privacidad del Consumidor de California (CCPA)** [47] y su modificación, la CPRA, es una ley del estado de California en Estados Unidos. Esta ley tiene similitudes y diferencias respecto a la RGPD, pero es importante mencionarla debido a que el estado de California es donde se encuentra "Silicon Valley", también conocido como la meca de la tecnología, donde se encuentran algunas de las empresas tecnológicas más importantes de Estados Unidos y del mundo. Debido a esto se realizan gran cantidad de investigaciones y se recopilan gran cantidad de datos. W. G. Voss [48] hace un análisis comparativo entre esta ley y la RGPD, concluyendo que hay similitudes pero también grandes diferencias ocasionadas por los diferentes modelos jurídicos o las definiciones de tratamiento o datos personales. En lo referente a los datos biométricos en la CCPA no existe la categoría de datos sensibles, pero en ambas legislaciones son tratados como información personal a proteger.
- **Ley de Protección de Información Personal (PIPL)** [49] de China. Toma como base la RGPD europea, pero la adapta a la idiosincrasia del país, como exponen W. Li y J. Chen [50]. En ciertos artículos como [51] se cuestionan sobre si esta ley está más enfocada a los derechos de las personas o a los intereses nacionales en cuanto a seguridad y restricciones. En lo referente al ámbito de este proyecto, tenemos en cuenta los artículos 6, 13, 14, 15, 17, 25 y 26 donde se establecen los principios generales para el tratamiento de la información personal y los derechos de las personas, estableciendo como necesario el consentimiento expreso por parte de los individuos y se define la información personal sensible y su tratamiento especial. También es necesario tener en cuenta los artículos 13 y 51, en los que se aborda la investigación científica y académica. En resumen, la PIPL describe los datos biométricos como información personal sensible dándolos un nivel de protección más alto. Esto significa que se necesita un consentimiento

explicito, solo se pueden usar para un propósito específico y en un tiempo limitado y requieren de medidas de seguridad más estrictas. En lo referente a la investigación, se permite el uso de información personal sensible como los datos biométricos siempre que sean necesarios y relevantes, tomando las medidas de seguridad oportunas y promoviendo la anonimización y seudonimización. Por lo que en estos aspectos es similar a RGPD.

- **Ley de Protección de Información Personal (APPI)** [52] de Japón establece datos que son considerados información personal y otro información personal sensible, aunque no hace referencia explícita a datos biométricos. En el capítulo IV establece ciertas condiciones especiales para el tratamiento de datos personales en el ámbito de la investigación académica, la seudonimización y anonimización.
- **Ley General de Protección de Datos (LGPD)** [53] de Brasil estable que los datos biométricos son de carácter sensible y tienen unas condiciones específicas para su tratamiento. Permite el uso de datos sensibles para investigación bajo ciertas condiciones como puede ser la anonimización.
- **Ley de Protección de Datos Digitales Personales (PDDP)** [54] de India no hace referencia explícita a datos biométricos o consideraciones especiales para la investigación, pero existe la categoría de datos especiales y la necesidad de un consentimiento expreso.
- **Ley de Protección de Datos** o "Personal Information Protection Act" (PIPA) [55] de Corea del sur no hace una referencia explícita a los datos biométricos pero si a los datos sensibles y tiene en cuenta el uso de dichos datos para investigaciones científicas. Tiene en cuenta el consentimiento explícito y la seudonimización de los datos.

Como era de esperar, la preocupación por la privacidad tiene carácter global y de manera más individual o localizada como hemos visto hasta ahora o de manera global como analizan V. K. S. Anil y A. Babatope [56] o S. Lim y J. Oh [57], en la literatura se ha analizado frecuentemente este tema. Pero en lo que atañe a este trabajo, no se va ahondar más en ello porque ya tenemos una visión general con la que podemos enfrentarnos al verdadero problema que es la creación de un corpus de datos.

4.6. Otras referencias normativas

Además de todas las referencias legales en relación con la privacidad y la protección de datos que se han visto hasta ahora, es necesario incluir este apartado donde se mencionaran otro tipo de referencias, que, si bien no tienen el carácter de ley, en determinados momentos pueden ser de obligado cumplimiento. Estas referencias pueden estarndares, normas ISO, normas UNE, protocolos de comunicación, protocolos de adquisición de datos o protocolos de almacenado de datos entre otros. Algunos ejemplos prácticos pueden ser los siguientes:

- **Norma ISO/IEC 24745** [58]: Proporciona directrices para proteger la información biométrica bajo los requisitos de confidencialidad, integridad y renovabilidad/revocabilidad durante el almacenamiento y la transferencia. También establece requisitos y recomendaciones para el manejo y procesamiento seguro y conforme a la privacidad de la información biométrica.
- **Norma ISO/IEC 27001** [59]: Norma que establece un marco para los sistemas de gestión de la seguridad de la información, ciberseguridad y protección de la privacidad.
- **Norma ISO/IEC 29794-1** [60]: Norma que establece un marco para evaluar la calidad de las muestras biométricas. Define términos y métodos para interpretar las puntuaciones de calidad biométrica, normalizar datos y agregar resultados de algoritmos de calidad.
- **Protocolo TLS/SSL**: Protocolos criptográficos para transmisiones seguras.
- **Estándar X.509**: Es un estándar criptográfico utilizado en los certificados de claves públicas, algo esencial en protocolos de trasmisiones seguras como TLS/SSL para verificar la identidad de los participantes.
- **Protocolo HTTPS**: Es un protocolo seguro basado en HTTP que utiliza el protocolo TLS/SSL para encriptar las comunicaciones entre un servidor web y sus usuarios.

Sin ser específicamente normas, también es necesario nombrar los comités que se encargan de desarrollar, mantener y promover el uso de dichas normas y estándares:

- **Comité Técnico Conjunto 1** (ISO/IEC JTC 1) [61] que es el Comité Técnico de la Organización Internacional para la Estandarización (ISO) y de la Comisión Electrotécnica Internacional (IEC) encargado de los campos de la Tecnología de la información (TI) y las Tecnologías de la información y la comunicación (TIC).
- **Subcomité Técnico Conjunto sobre biometría** (ISO/IEC JTC 1/SC37) [56] que es el subcomité encargado específicamente del campo de la biometría, que a su vez se compone de diferentes grupos de trabajo específicos para cada estudio como el formato de intercambio de datos, la implementación técnica de los sistemas, las interfaces o el vocabulario específico.

Con estas y otras normas, protocolos, estándares y certificaciones conseguimos proteger la privacidad de los datos en todas las fases de la investigación, desde su adquisición, hasta su publicación, pasando por la transmisión, el procesado y el almacenamiento.

5. Repositorios de datos

Esta sección está destinada al estudio realizado sobre los repositorios de datos donde se publican, almacenan y comparten los datos que se han recogido a través de los participantes de la investigación.

Los repositorios de datos son de vital importancia en muchos aspectos, incluida la investigación científica, ya que permiten la puesta en común de las investigaciones. Gracias a los repositorios es posible publicar los artículos finales de las investigaciones, los datos utilizados y el código utilizado o desarrollado y el resto materiales utilizados.

Las ventajas que ofrecen los repositorios en general son:

- Centralización de la información.
- Organización lógica de la información.
- Facilitar el descubrimiento de información.
- Facilitar la identificación y citación de la información.
- Preservar los datos a largo plazo.
- Compartir la información.

Debido a la gran importancia que tienen los repositorios de datos, junto con la obligación de las universidades de publicar sus investigaciones de forma accesible, muchas universidades tienen su repositorio institucional pero además existe la Red de Bibliotecas Universitarias y Científicas Españolas (REBIUN) [62], que está formada por las bibliotecas de las 76 universidades que forman parte de la Conferencia de Rectores y Rectoras de las Universidades Españolas (CRUE) [63] (49 de ámbito universitario público y 27 de ámbito universitario privado) y el CSIC (Consejo Superior de Investigaciones Científicas) [64]. Gracias a REBUIN los miembros de la comunidad universitaria de cualquiera de las universidades que forman parte puede tener acceso al catálogo de cualquiera de las bibliotecas universitarias y por lo tanto tener acceso a las investigaciones que se han realizado.

Además de los repositorios, existen herramientas que permiten el manejo y administración de repositorios bibliográficos, que pueden contener datos de diferente índole como libros, tesis, artículos de investigación, corpus de datos o videos entre otros. Algunos ejemplos de herramientas de este tipo son DSpace [64] o GreenStone Digital Library [65]. Estas herramientas están enfocadas en los gestores de los repositorios no tanto en los usuarios de los repositorios.

Centrándonos específicamente en los repositorios podemos encontrar gran cantidad de repositorios de datos almacenados en la nube y que nos permiten obtener o almacenar datos de todo tipo. Al ser imposible analizarlos todos, hay que realizar una selección de los repositorios a analizar. Para esta selección se van a tener en cuenta diversos criterios de diversa índole. Los criterios seguidos para seleccionar los repositorios han sido:

- Repositorios usados durante el máster en alguna de las asignaturas.
- Repositorios usados mí mismo de forma personal o profesional.
- Repositorios usados por colegas de profesión a los que se ha consultado.
- Repositorios nombrados en la bibliografía analizada.
- Repositorios en encontrados gracias a búsquedas en la red.

Algunos de estos repositorios en línea son:

- **DataCite** [66]: Repositorio con la finalidad de mejorar la efectividad a la hora de localizar, identificar y citar datos de investigación gracias al uso del código DOI (Digital Object Identifier) [67].
- **European data portal** [68]: Recopila una diversidad de datos, entre otras publicaciones, del sector público de la Unión Europea. Cumple con la Regulación General de Protección de Datos (GDPR).
- **Fairsharing** [69]: Es un repositorio de estándares, datos y políticas, con datos categorizados.
- **Google dataset search** [70]: No es un repositorio como tal, es una herramienta de Google que sirve para hacer una búsqueda de datos y que te lleva al sitio web donde estén contenidos esos datos.
- **Huggin Face** [71]: Ofrece gran variedad de datos, y de modelos, enfocados al aprendizaje automático agrupados en categorías.

- **Kaggle** [72]: Repositorio con gran variedad de tipos de datos abiertos y accesibles. Gran parte de su utilidad está en el campo del aprendizaje automático.
- **OpenDOAR** [73]: Repositorio de acceso abierto y gratuito con garantía de calidad para acceder a recursos académicos.
- **Paperswithcode** [74]: Repositorio donde encontrar artículos enlazados con su código, datos, métodos y evaluación. Tiene licencia CC-By-SA y está especializado en astronomía, física, informática, matemáticas y estadística.
- **Re3data** [75]: Es un registro de repositorios de datos de investigación de diferentes disciplinas, es decir, actúa como buscador al tener indexados otros repositorios.
- **Zenodo** [76]: Repositorio de propósito general orientado a investigadores de todos los campos. Cumple con la política de ciencia abierta [5], [77], [78].
- **GitHub** [79]: Plataforma enfocada en la creación de código mediante control de versiones pero al permitir la subida de archivos que no sean código propiamente dicho, permite la publicación de datos.
- **Mendeley Data** [80]: Repositorio de datos de investigación de la editorial Elsevier donde los investigadores publican los datos en los que se basan sus investigaciones.
- **Centro Nacional de Información Biotecnológica** (National Center for Biotechnology Information - NCBI) [81]: Una de sus funciones es actuar como repositorio de datos de carácter médico o biomédico como proteínas, genoma humano.
- **VIVLI** [82]: Plataforma de intercambio de datos de investigación clínica. Permite compartir y acceder a los datos de los ensayos clínicos.
- **IEEE DataPort** [83]: Repositorio de datos de investigación de la editorial IEEE Xplore enfocado en la comunidad académica, donde los investigadores publican los datos en los que se basan sus investigaciones.
- **Figshare** [84]: Es un proveedor de repositorios de investigación abiertos.
- **Dataverse** [85]: Es un repositorio de datos de investigación de código abierto diseñado para la gestión, almacenamiento, publicación y preservación de datos científicos y académicos.

- **Dryad** [86]: Es una plataforma para la publicación de datos abiertos orientada a la disponibilidad abierta y la reutilización de los datos de investigación.
- **Open Science Framework** (OSF) [87]: Herramienta de gestión de proyectos gratuita y de código abierto para gestionar el flujo de trabajo, organizar materiales y compartir la totalidad o parte de un proyecto con la comunidad investigadora en general.
- **4TU.ResearchData** [88]: Es un repositorio internacional de datos para la ciencia, la ingeniería y el diseño que busca la preservación y la conservación de conjuntos de datos de investigación.
- **UVaDOC** [89]: Repositorio documental de la Universidad de Valladolid donde se publican los trabajos de los miembros de la comunidad universitaria. Tiene una sección dedicada a Datasets y corpus de datos [90].

Además de estos repositorios existen otros pertenecientes a diferentes administraciones públicas gracias a la política de datos abiertos como la Secretaría de Estado de Digitalización e Inteligencia Artificial [91] o la Junta de Castilla y León [92]. También podemos encontrar repositorios de datos en las universidades como en la UVa [93]. Otra manera de encontrar los datos que se puedan necesitar para nuestras investigaciones o averiguar en qué repositorio de datos deberíamos publicarlos es utilizar alguno de los diferentes buscadores de repositorios de datos existentes como por ejemplo [94] o [95].

De esta lista de repositorios nombrados, para realizar un análisis se van a descartar los buscadores de repositorios, ya que no son repositorios por sí mismos, aunque nos ayuden en las labores de búsqueda. También se van a descartar los repositorios especializados en campos específicos no tan relacionados con los objetivos de este trabajo como GitHub por estar especializado en código, NCBI por estar especializado en datos médicos o biomédicos y VIVLI por estar dedicada a ensayos clínicos. Otra categoría que se va a descartar es la de los repositorios centrados en aprendizaje automático como Huginn Face o Kaggle. El análisis va a estar centrando el análisis en los repositorios de investigaciones científicas por lo que se van a analizar Mendeley Data, IEEE DataPort, Zenodo, Figshare, Dryad, Open Science Framework (OSF) y UVaDOC por ser el repositorio institucional de la universidad.

Los objetivos del análisis de los repositorios son los siguientes:

- Saber si el repositorio está vivo y activo para conocer si se publican corpus de datos actualmente y con regularidad o no y datos los corpus publicados.
- Saber si el acceso y la publicación de corpus de datos es libre. En este apartado queremos saber si se requiere un registro, si requiere el pago de alguna cuota (única/mensual/anual).
- Conocer el tipo de datos soportados. Qué tipos de datos se pueden publicar como imágenes, archivos CSV, JSON, bases de datos, XLS, XLSX, DOC, PDF, etc.
- Saber si cumplen con las reglas FAIR y facilitan la ciencia abierta.
- Quien es el titular del repositorio.
- Saber cómo fomentan la visibilidad y el acceso a los datos
- Licencia de los datos.

Para encontrar la información necesaria, se consultarán las diferentes secciones de las webs de los repositorios, como las secciones FAQ, About, Support, Policies o similares que haya disponible en la web y se simulará la creación de un repositorio de datos para ver las opciones disponibles.

La información recogida de los repositorios es:

- **UVaDoc:**

UVaDoc es el repositorio documental multidisciplinar de la Universidad de Valladolid. Permite el acceso de forma abierta a la documentación que se genera en la universidad tanto por la actividad científica, como docente e institucional. Algunas de las publicaciones son trabajos de fin de estudios, documentos instruccionales o producción científica.

El repositorio busca garantizar el libre acceso a la documentación a la vez que pretende aumentar la visibilidad y difusión de los contenidos, para ello utiliza la plantilla o estructura de metadatos *Dublin Core* (DC) [96] y asigna un DOI a las publicaciones. Gracias a estas políticas está alineada con los principios FAIR.

Al publicar en este repositorio el autor conserva los derechos de propiedad intelectual, pero el elemento publicado estará bajo la licencia *Creative Commons*

que el autor escoja y permitiendo a la universidad difundir en abierto y preservar la publicación.

Este repositorio forma parte del consorcio BUCLE (Bibliotecas Universitarias de Castilla y León) [97] y de la red de bibliotecas REBIUM [62]. El contenido de su catálogo se publica en catálogos, agregadores o portales como Hispana [98], Recolecta [99], Europeana [100], OpenAire [101], EROMM [102], ROAR [103], BASE [104], Google Scholar [11] y WorldCat [105]. Además, al tener soporte para OAI-PMH hace que la interconexión y el envío de metadatos sea más simple.

El repositorio está montado sobre OA 5.10 y por lo tanto dispone de la API que viene integrada en DSpace para gestionar el repositorio con soporte para SWORD v2 [106].

El repositorio está ordenado en diferentes secciones en función del tipo de entrada publicada, estas secciones son documentos institucionales, producción científica, fondo antiguo, publicaciones UVa, objetos de aprendizaje y trabajos fin de estudio. A su vez cada una de estas secciones está divida en subsecciones. En total hay seis secciones y veintiocho subsecciones. Para este estudio la sección que nos interesa es la de producción científica y la subsección datos de investigación.

Se recomienda el uso de formatos de archivo fácilmente legibles y formatos abiertos como PDF, XML, RTF, TIFF, JPG, MP3, MP4 o ZIP frente a archivos propietarios como WORD, GIF. Se pueden consultar todos los formatos admitidos en el documento de formatos soportados [107].

Los encargados de publicar en el repositorio son los encargados del repositorio, también publicar el Personal Docente e Investigador (PDI) de la universidad solicitando ese permiso para su usuario, por lo cual como estudiante no es posible realizar una publicación y no se ha podido simular una para estudiar los datos solicitados.

Al acceder a la pantalla de un corpus de datos, hay dos opciones una simple y una ampliada, por defecto se muestra la simple donde podemos encontrar diferentes metadatos como el título, autor, editor, año de publicación, resumen, departamento, DOI, idioma, URI dentro UVaDoc, tipo de versión, derechos, subsección en la que se encuentra, los ficheros y la licencia del elemento

publicado. Dentro de la parte dedicada a los ficheros se encuentra el nombre, el tamaño, el formato y la descripción, además de la opción de visualizar o abrir el archivo, que dependiendo del tipo de archivo también sirve para descargarlo. Además, encontramos varias opciones y formatos de exportar la cita del elemento y datos estadísticos sobre las veces que se ha citado el elemento.

En la versión ampliada de la pantalla de visualización aparecen más metadatos como las fechas completas de adhesión y disponibilidad, proyecto al que está asociado o la fuente de financiación.

Al realizar una búsqueda existen varias opciones de búsqueda, filtrado y ordenación. Permite filtrar en función de la tipología, autor, titulación, fecha (con un rango de años), formato o encuadernación. También opciones de búsqueda como título, autor, editor, materia, fecha de publicación o formato entre otros. Las opciones de ordenación o listado son comunidades, fecha de publicación, autores, materias y títulos.

A día 20 de mayo de 2025, hay publicados un total de 71.795 elementos. Si nos centramos en la sección de producción científica el total de elementos publicados es de 16.443 (aproximadamente un 23% del total). Si nos fijamos en la subsección de datos de investigación, donde se publican los datos usados para las investigaciones, el total de elementos publicados es de 65. Siendo el más reciente de 22 de abril de 2025.

- **Mendeley Data:**

A través de la información publicada en la web, encontramos que los datos se pueden publicar con un acceso público o restringido, en caso de ser restringido se puede optar por la opción de que los interesados soliciten al autor los datos totales. La publicación está compuesta por una serie de metadatos que siguen los formatos de Dublin Core y Schema.org conforme al estándar de Google Dataset. Se nos proporciona más información, por ejemplo, que el servidor donde se almacenan los datos es en Amazon S3, que disponen de la certificación ISO/IEC 27001:2013 relativa a las prácticas seguras de gestión de la información o que las comunicaciones y el almacenamiento son seguros a través de cifrado entre otros datos.

Tienen una sección dedicada al tipo de archivos preferidos y permitidos [108] donde se puede ver que admite los tipos de datos más usados en documentos, imágenes, hojas de cálculo, video, audio, etc. y que se prefieren archivos de tipo libre como ODS o SVG a formatos propietarios como XLSX o AI (Microsoft o Adobe). El conjunto total puede tener un tamaño máximo 10 GB, pero no especifican si cada fichero tiene un límite individualmente. Los límites de datos pueden variar en función del tipo de cuenta de que se disponga, este límite es el establecido con la cuenta usada a través del convenio con la universidad.

Disponen de una API para acceso y búsqueda de información, dedicando una sección de la web a la explicación de su funcionamiento [109].

No se ha encontrado una referencia exacta de precios o cuotas por los servicios. Existe la posibilidad de obtener una cuenta gratuita y la posibilidad de ampliar la cuenta con un plan premium previo pago, por lo que si no se dispone de una cuenta institucional es muy probable que se necesite el pago de una cuota para acceder a ciertos servicios o tener más capacidad en las publicaciones.

Cuando se publica un conjunto de datos, no se pierde la propiedad y control sobre el mismo, y puedes elegir bajo qué licencia publicar los datos dentro del listado que permiten. Estas licencias se agrupan en función del tipo de información publicada. Para todos los tipos disponen de las licencias CC0 1.0, CC BY 4.0, CC BY-SA 4.0, CC BY-NC 4.0, CC BY-NC-SA 4.0, CC BY-ND 4.0, CC BY-NC-ND 4.0 y CC BY NC 3.0. Cuando se publica software las licencias específicas disponibles son MIT, Apache-2.0, BSD 3-clause, BSD 2-clause, GPLv3, GPLv2, LGPL, MPL-2.0, CeCILL y CeCILL-B. Cuando la publicación se refiere a hardware las licencias específicas disponibles son CERN OHL y TAPR OHL.

Los conjuntos de datos publicados son agregados al índice de metadatos de DataCite [66] y portal OpenAire [101].

No se ha encontrado ninguna mención directa a los principios FAIR, pero como el repositorio forma parte de *Generalist Repository Ecosystem Initiative* (GREI) [110], el repositorio debe estar alienado con dichos principios.

A fecha 20 de mayo de 2025, sin aplicar filtros nos arroja un resultado de 53.006.082 elementos, siendo la publicación más reciente del 20 de mayo de 2025.

Al comenzar a crear la información para publicar un conjunto de datos, lo primero nos solicita un nombre para el conjunto de datos. En ese momento se reserva un DOI de DataCite, que se activará tras la publicación del conjunto de datos. En la pantalla de creación del conjunto de datos nos permite añadir ciertos metadatos y los datos propiamente dichos. Los metadatos son los autores, la licencia, instituciones, categorías de datos, quien ha financiado el conjunto de datos, los pasos para reproducir la investigación y por último añadir links relacionados.

Al entrar como usuario para buscar repositorios, encontramos una barra de búsqueda y la opción de usar diferentes filtros (año desde, año hasta, tipo de dato, tipo de fuente, fuente y entidad financiadora), mostrando los resultados en el centro de la pantalla y permitiendo ordenarlos por relevancia y por fecha (más antiguos o más nuevos).

- **IEEE Dataport:**

Las opciones accesibilidad de este repositorio vienen dadas en función del tipo de los tipos de usuario establecidos que van incrementando sus posibilidades de acceso. Un usuario sin registrar solo puede realizar búsquedas y ver la información general de los corpus de datos, no tiene acceso a los datos. Un usuario registrado de forma gratuita puede ver los datos de los corpus de datos que sean de acceso abierto y un usuario de pago, bien de forma individual o institucional puede ver todos los corpus datos.

Los corpus de datos alojados en este repositorio pueden ser de acceso público o restringido, esto lo elige el propietario del corpus a la hora de subirlo. Si se sube de manera gratuita el acceso es restringido, para que sea de acceso público hay que pagar por ello un importe de 1.950\$. Si el corpus de datos se registra de manera gratuita, su acceso será restringido y solo tendrán acceso los usuarios de pago.

La tarifa individual de un usuario de pago es de 40\$ mensuales. La tarifa institucional básica va entre 4.500\$ y 22.500\$ al año, varía en función del tamaño de la institución. Por último, la tarifa institucional premium va desde 5.995\$ hasta 29.995\$ al año en función del tamaño de la institución. Las principales diferencias están en el almacenamiento de datos donde el usuario individual tiene un límite

de 2 TB por cada corpus de datos y los usuarios institucionales tienen 10 TB. Además, los institucionales tienen un panel de control institucional y como añadido para la suscripción institucional premium, disponen de corpus de datos privados y la posibilidad de publicar de forma abierta 20 corpus de datos de forma gratuita.

Se admiten gran variedad de datos, en la web oficial se mencionan específicamente CSV, TXT, ORC, Avro, RC, XML, SQL y JSON, que es una lista escasa. Posteriormente, cuando se realiza una simulación de subir un archivo, por cada tipo de archivo se facilitan los formatos admitidos, para imágenes son los tipos png, gif, jpg, jpeg y webp. Para documentación pdf, txt, docx, doc, md y html. Para scripts zip, gz, gzip, tbz2, 7z, txt, rmd, py, r, m, pig, java y sas. Finalmente, para los datos los formatos admitidos son zip, gz, gzip, csv, json, txt, sql, xml, tsv, ebs, avro, orc, parquet, hdf5, 7z, tbz2, iso, tar, bz2, z, xls, xlsx, graph, properties, offsets, flac, ogg, wav, aac, mp3, gif, jpg, jpeg, png, avi, mov, mp4, mpg, m4v, yaml, dat, mat, fig y rar.

No se ha encontrado ninguna referencia a la existencia de un API de acceso a los corpus de datos, aunque existe una API para acceder a IEEE Xplore. Se menciona que al estar los datos alojados en servidores AWS mediante el servicio S3, se permite subirlos directamente dicho servicio.

Cuando se publica un conjunto de datos, no se pierde la propiedad y control sobre el mismo. Los datos se publican bajo licencia "Creative Commons" Attribution (CC-BY). Cuando se simula la creación de una entrada al repositorio, no existe la opción de elegir que licencia usar dentro de las opciones existentes de CC-BY.

Se explica que gracias a los metadatos que se establecen en los corpus de datos a la hora de incluirlos en el repositorio se cumplen con los requisitos FAIR.

No indica si se agrega la entrada del repositorio en algún buscador o indexador adicional.

Cuando se sube un corpus de datos al repositorio, se solicita diversa información como el/los autores, permitiendo referencias ORCID ID, título, categoría (hay 35 disponibles), resumen, palabras clave, formato de los datos, agencia de financiación, número de financiación, documentación e instrucciones, enlaces asociados y los datos propiamente dichos. A cada corpus de datos o publicación

se le asigna un DOI para su referencia. Una vez publicado el corpus de datos, desde su página se puede hacer uso del botón “Citar” que permite obtener la cita en diversos formatos.

Al entrar como usuario para buscar repositorios, encontramos una barra de búsqueda que indica que podemos buscar por título, autor, resumen, categoría, formato de los datos o DOI. Además, existe la opción de usar diferentes filtros (categoría, tipo de conjunto de datos (libre o restringido), palabras clave, colección o formato de los datos), mostrando los resultados en el centro de la pantalla. No existe ninguna opción de ordenación de resultados.

A fecha 20 de mayo de 2025, sin aplicar filtros nos arroja un resultado de 13.023 elementos, siendo la publicación o actualización más reciente del 20 de mayo de 2025.

- **Zenodo:**

Es un repositorio de propósito general abierto a todo el mundo cuyo principal objetivo es fomentar la ciencia abierta.

Este repositorio es uno de los resultados de la política de ciencia abierta desarrollada por la Unión Europea a través de su programa openAIRE [101]. Ha sido desarrollado por CERN [111] como uno de los socios de openAire, que también lo gestiona. Además, forma parte de Generalist Repository Ecosystem Initiative (GREI).

Al estar tan enfocado en el acceso abierto y la ciencia abierta, dan ejemplo con ellos mismos y el código del proyecto está de forma abierta en GitHub donde la comunidad puede contribuir con mejores. También admiten sugerencias de nuevas funcionalidades y mejoras. Toda esta información está accesible bajo la licencia CC0. Fruto de esta filosofía, tiene una integración con GitHub en cuanto a usuarios y publicar desde repositorios de GitHub.

Este repositorio usa el protocolo de metadatos OAI-PMH para permitir recopilar información, por lo que está alineado con los principios FAIR.

Dispone de una API bien documentada que permite publicar investigaciones, buscar referencias del repositorio y cargar y descargar archivos.

Tiene una funcionalidad en la que se pueden cargar datos e investigaciones de forma privada, para tener todos los datos almacenados y reunidos mientras se realiza la investigación y poder hacer públicos una vez que se ha finalizado la investigación. Otra funcionalidad añadida, es poder dar acceso a determinadas personas a los datos para la revisión por pares de las investigaciones.

Cuando se carga un corpus de datos, se le asigna un DOI obtenido de DataCite con el prefijo de Zenodo. Permite el uso de ORCID para identificar a los autores.

Los elementos cargados en Zenodo son indexados por Google Scholar y son indexados en openAIRE lo que permite su descubrimiento y acceso desde todos los proyectos de openAIRE.

Las limitaciones en cuanto a los archivos es que no puede haber más de 100 archivos y en total no pueden superar los 50GB de tamaño. No se indica que haya un tamaño máximo por archivo. No permite la organización en carpetas o directorios, para organizar los archivos hay que usar archivos comprimidos. No proporciona una lista de tipos de archivos admitidas, ya que indica que soporta todos los tipos de archivos.

Al cargar contenido, el propietario no varía, es decir el autor o institución sigue siendo el propietario de la información publicada. Para compartir la información, Zenodo soporta un gran número de licencias con las que publicar el corpus de datos, tiene un total de 444 licencias, etiquetando algunas específicas para datos o para software. Las licencias que recomienda son Apache License 2.0, Creative Commons Attribution 4.0 International, Creative Commons Attribution Share Alike 4.0 International, Creative Commons Zero v1.0 Universal, GNU General Public License v3.0 or later o MIT License.

Cuando se va a realizar la carga de un corpus de datos vía web para publicarlos en Zenodo, se muestra una pantalla con diferentes secciones para cargar los datos y establecer los metadatos. Las diferentes secciones son “Files” para cargar los datos, “Basic Information” para establecer el DOI, indicar el tipo de dato que se va a subir (proporciona un amplio listado de 44 opciones como “dataset”, “image” con diferentes subtipos, “publication” con diferentes subtipos, “software” o

"Video/Audio" por ejemplo), el título, la fecha de publicación, los creadores, la descripción, la licencia y el copyright, la siguiente sección es "Recommended information" para añadir los metadatos de colaboradores, palabras clave y temas, idioma, fechas importantes (con un listado de tipos de fechas), la versión y la editorial, la siguiente sección es "Funding" para informar de donde viene la financiación, la siguiente sección es "Alternate identifiers" para añadir otros identificadores alternativos, la siguiente sección es "Related works" para añadir trabajos relacionados, después tenemos la sección "References" para añadir las referencias, la siguiente sección es "Software" para incluir el repositorio del código fuente, el lenguaje de programación y el estado del repositorio, la siguiente sección es "Publishing information" para incluir la información de donde se ha publicado por ejemplo una revista con su correspondiente ISSN o libro con su ISBN si es el caso, la siguiente sección es "conferencce" para incluir los datos de la conferencia donde se ha presentado si es el caso y por último está la sección "Domain specific fields" para incluir metadatos adicionales como por ejemplo dispositivos de captura o configuraciones. Por último, podemos seleccionar si el corpus de datos será público o restringido.

Al realizar una búsqueda, no requiere ningún usuario para realizar y ver los resultados, encontramos una barra de búsqueda donde escribir los términos de búsqueda. En la parte izquierda de la pantalla existen diferentes filtros para acotar los resultados (versión, tipo de acceso, tipo de recurso, temas y tipo de archivos). Los resultados se muestran en el centro de la pantalla y permite ordenarlos de diferentes maneras como más nuevos, más viejos, más vistos, etc.

A fecha 20 de mayo de 2025, sin aplicar filtros nos arroja un resultado de 4.689.694 elementos, siendo la publicación o actualización más reciente del 20 de mayo de 2025. Si limitamos la búsqueda a mostrar solo los conjuntos de datos, hay un total de 436.300 siendo la última incorporación el 20 de mayo de 2025.

Cuando se accede a la sección de un corpus de datos, permite ver la información contenida en los metadatos, datos estadísticos como el número de veces que se ha visto el corpus de datos y el número de veces que se ha descargado y obtener la cita del corpus de datos en diversos formatos.

- **Figshare:**

Es un repositorio abierto enfocado en la investigación. No está enfocado solo en corpus de datos, si no que se pueden publicar otros contenidos como los resultados de las investigaciones, imágenes o videos.

No es solo un repositorio, también es un software para gestionar repositorios usado y financiado por multitud de instituciones y forma parte de *Generalist Repository Ecosystem Initiative* (GREI).

Es un repositorio comprometido con la ciencia abierta y los principios FAIR, siendo una de sus convicciones que las publicaciones de su repositorio deben ser legibles y consultables tanto por personas como por máquinas, favoreciendo de este modo que las publicaciones sean encontrables y accesibles, esto se ve favorecido gracias al protocolo de metadatos OAI-PMH. Como software buscan que dentro del mismo sistema la misma información solo tenga que ser introducida por los investigadores una única vez.

Es una solución de la empresa Digital Science [112] y como empresa comercial, su software es propietario, pero buscando que la ciencia publicada con él sea abierta y accesible. Los datos se almacenan en el servicio Amazon S3.

Dispone de una API bien documentada que permite publicar investigaciones, buscar referencias del repositorio y cargar y descargar archivos.

Cuando se carga un corpus de datos, se le asigna un DOI obtenido de DataCite. Permite el uso de ORCID para identificar a los autores. Tienen integración con GitHub, GitLab y NitBucket para cargar datos para las publicaciones.

Tiene una versión gratuita y una versión premium de pago llamada Figshare Plus [113] y dependiendo del tipo de cuenta los límites en cuanto a espacio y licencias disponibles varían. En la siguiente tabla se muestran las diferencias entre las dos versiones.

Característica	Figshare gratuito	Figshare Plus
Límite de almacenamiento	20GB	Ilimitado
Tamaño máximo de archivo individual	20GB	5TB
Número máximo de archivos por artículo	500	500 (Ampliable por petición)

Licencias disponibles	Creative Commons CC0 CC BY 4.0	CC0 CC BY 4.0 CC BY-SA 4.0 CC BY-ND 4.0 CC BY-NC 4.0 CC BY-NC-SA 4.0 CC BY-NC-ND 4.0
Otras licencias disponibles	MIT GPL GPL 2.0+ GPL 3.0+ Apache 2.0	MIT GPL GPL 2.0+ GPL 3.0+ Apache 2.0 BSD 3-Clause MPL-2.0 CDLA-Sharing-1.0
Precio	Gratis	Desde 875\$ en función del tamaño del almacenamiento

Tabla 1 Comparativa de tipo de usuario de Figshare

Tiene una sección de su web donde explica el tipo de archivos y tipos de extensiones admitidas [114], donde expone todos los formatos admitidos, entre los cuales se encuentran los formatos más utilizados. Tiene otras secciones dedicadas a ayudar al usuario como una explicación de las licencias [115] para guiar en la selección de una de ellas y otra sección a modo de manual para cargar datos [116].

Como usuario gratuito, una vez que te registras desde el perfil puedes incluir de ORCID y en la sección Upload puedes elegir entre subir datos, crear proyectos (individuales o grupales) y colecciones (para agrupar diferentes publicaciones). Cuando se procede a la carga de un corpus de datos, el proceso a seguir es primero cargar un archivo y una vez cargado, podemos seguir cargando más datos y estructurarlos en carpetas. También podemos asignar los siguientes metadatos: título, un tipo de elemento (hay 12 disponibles entre ellos *dataset*, *software*, *journal contribution* o *conference contribution*), otros autores, categoría (se pueden elegir entre 23 categoría que están compuestas a su vez por varias subcategorías), palabras clave, descripción, financiación, elementos relacionados y licencia. Una vez con esta información completa, podemos realizar varias

acciones como establecer una fecha de publicación, reservar un DOI, guardar la información o publicarla.

Desde la página principal hay una barra de búsqueda, que, al realizar una búsqueda, nos muestra los resultados en una pantalla dividida en dos zonas, a la izquierda aparecen los filtros que podemos aplicar (tipo de contenido, tipo de artículo, tipo de licencia, entidad financiadora, año de publicación, categoría, fuente o institución) y a la derecha los resultados, que tienen varias formas de ordenación (relevancia, citaciones, fecha de publicación, etc.) y visualización (lista o cuadrícula).

Al acceder a una publicación se ve una previsualización del archivo, los metadatos como el título, el autor, la descripción, palabras clave, categoría o licencia. Además, se muestran una serie de datos estadísticos como número de veces que se ha visto, se ha descargado o se ha citado y se permite descargar el contenido. También hay un botón para obtener la cita en varios formatos.

A día 20 de mayo de 2025, al realizar una búsqueda sin filtros hay un total de 7.431.309 elementos publicados, siendo el más reciente del 20 de mayo de 2025. Si restringimos la búsqueda al tipo *dataset*, para mostrar corpus de datos, la cantidad total es de 2.157.646 y la última publicación es del 20 de mayo de 2025.

- **Dryad:**

Dryad es un repositorio de acceso abierto enfocado en los datos de las investigaciones (principalmente investigaciones científicas y médicas como biología o genética) para que estén disponibles de forma abierta y favorecer la reutilización de los mismos.

Buscan crear una comunidad comprometida con prácticas abiertas en la investigación y fomentar la integridad y la calidad de los datos. Esto se refleja en su código de conducta y en firmar los Principios de la infraestructura académica abierta (Principles of Open Scholarly Infrastructure (POSI)) [117].

En su código de buenas prácticas establecen algunas directrices para que los corpus de datos publicados cumplan con los principios FAIR y obtener la máxima repercusión posible.

El repositorio se integra con DataCite, ORCID, ROR [118] y un registro de entidades financiadores de investigaciones, para poder identificar y conectar los corpus de datos con un código DOI, con los autores, las instituciones y las entidades que financian la investigación. También se integra con los flujos de trabajo de las editoriales para favorecer la interconexión entre artículos publicados en esas editoriales y los datos publicados en Dryad.

Dispone de una API para realizar todas las operaciones necesarias.

En lo referente a los archivos que componen el corpus de datos, Dryad acepta todo tipo de archivos, pero tienen preferencia por tipos de datos abiertos como CSV, JSON, XML, MD, PDF, PNG, WAV, MP3, MP4 o ZIP, por ejemplo. Recomiendan que cada archivo individual no supere los 10GB y tiene un límite de 300GB por corpus de datos.

Todos los corpus de datos se publican con licencia CC0.

Dryad forma parte de *Generalist Repository Ecosystem Initiative* (GREI).

Aunque es una entidad sin ánimo de lucro que fomenta la ciencia abierta y que intenta financiarse con donaciones de sus colaboradores, para asegurarse su supervivencia establece una serie de cuotas para los miembros y las publicaciones. Existen tres tipos de cuotas, una cuota anual que debe pagar cada miembro, una cuota por publicación y otra cuota por tamaño del corpus de datos (en GB). El importe de las cuotas depende de si el autor de la publicación está asociado a una de las entidades colaboradores o no y del tamaño del corpus de datos. Si un autor está asociado a una de las entidades colaboradoras no tiene que pagar la cuota de publicación ni la del tamaño siempre que sea menor de 10GB, si es mayor tiene que pagar una cuota de tamaño (el autor o la organización). Si el autor es independiente tiene que pagar una cuota de publicación, que va desde 150\$ para corpus de menos de 5GB hasta 12.162\$ para corpus de datos de entre 1 y 2 TB, es decir, a mayor tamaño más cuota de publicación. Además, hay que pagar una cuota por tamaño que va desde 30,00\$/GB cuando el corpus de datos es menor de 5GB hasta 6,08\$/GB para corpus de datos de más de 500GB, es decir, cuanto mayor tamaño tiene el corpus de datos menos cuesta el precio del GB. Para autores independientes disponen

de una solicitud de exención de la cuota de publicación si el corpus de datos es menor de 50GB que será estudiada por el equipo de Dryad.

Al requerir una cuota anual a los miembros no se ha procedido a registrarse y simular la carga de un corpus de datos para ver como es el proceso.

Al realizar una búsqueda, en la parte izquierda de la pantalla se muestran los filtros que se pueden utilizar que son palabras clave, revista, institución, entidad financiadora y extensiones de archivo. Permite ordenador los resultados por relevancia o por fecha y permite modificar el número de elementos que se muestran por página. Las publicaciones aparecen en la parte central de la pantalla mostrando el título y los autores. Al acceder a una publicación ya vemos toda la información disponible de los metadatos (título, autores, instituciones asociadas al autor, fecha de publicación, DOI, palabras clave, trabajos relacionados o licencia), los datos, un resumen explicativo, datos estadísticos como número de veces que se ha visto, descargado o citado y la nomenclatura para citarlo.

A día 20 de mayo de 2025 hay un total de 64.301 corpus de datos publicados, el último es del 20 de mayo de 2025.

- **Open Science Framework (OSF):**

Es un proyecto de infraestructura de código abierto desarrollado por la organización sin ánimo de lucro Center of Open Science (COS) [119] con el objetivo de apoyar durante todo el ciclo de vida de las investigaciones (planificación, realización, presentación de informes, publicación de resultados y datos y su posterior descubrimiento) para fomentar la apertura y reproducibilidad de la investigación científica. Forma parte de Generalist Repository Ecosystem Initiative (GREI).

OSF es gratuito y se financia por subvenciones públicas y donaciones privadas.

La información se guarda encriptada en diversos servidores como por ejemplo Google Cloud. Hay servidores en diversos países como Estados Unidos, Canadá, Alemania y Australia.

Inicialmente no había restricciones de tamaño, pero a partir del 3 de noviembre de 2020 se establecieron una serie de límites. El tamaño máximo de un archivo se estableció en 5GB, el tamaño máximo de proyectos privados se estableció en 5GB y el tamaño máximo de un proyecto público se estableció en 50GB. Estos límites están establecidos para el almacenamiento propio en OSF de forma genérica, es posible superar estos límites si se utiliza almacenamiento externo de las plataformas con las que tienen integración, estas plataformas son Amazon S3, BitBucket, Box, Dataverse, Dropbox, Figshare, GitHub, GitLab, Google Drive, OneDrive y OwnCloud. También se puede superar ese almacenamiento propio de OSF mediante el pago de cuotas en función de la cantidad de espacio necesario que van desde 500\$ por 100GB.

Las licencias que se pueden utilizar para publicar los corpus de datos son *Academic Free License 3.0*, *Apache License 2.0*, *Artistic License 2.0*, *BSD 2-Clause "Simplified" License*, *BSD 3-Clause "New"/"Revised" License*, *CC-By Attribution 4.0 International*, *CC0 1.0 Universal*, *Eclipse Public License 1.0*, *GNU General Public License (GPL) 2.0*, *GNU General Public License (GPL) 3.0*, *GNU Lesser General Public License (LGPL) 2.1*, *GNU Lesser General Public License (LGPL) 3.0* y *MIT License*. Son un total de 13 licencias más la posibilidad de publicar sin licencia, lo que significa que para realizar cualquier acción sobre el corpus de datos tienen que pedir autorización al autor.

OSF utiliza un perfil de metadatos propios alineados con los principios FAIR para que los elementos publicados sean accesibles y fáciles de encontrar. Además de esta plantilla de metadatos basada en el esquema de *Datacite 4.4*, se pueden usar otras plantillas como *Human Cognitive Neuroscience Data* creada por el *Center for Expanded Data Annotation and Retrieval (CEDAR)* [120], *LDbase Project Metadata Form*, *metaBUS* o *Psych-DS Official Template*.

No especifican los tipos de datos que admiten, pero tienen una guía de buenas prácticas como utilizar carpetas para tener un orden, separar el código de los datos, separar los datos originales de los derivados o de las salidas de los procesos, hacer los datos originales de solo lectura o crear archivos README explicativos. También aconsejan usar tipos de archivos accesibles y no propietarios como TXT, CSV, PDF, MPEG-4 o PNG con codificaciones estándar como UTF-8 o ASCII.

Dispone de una API bien documentada con diversas funcionalidades.

Cuando se hace una publicación se le asigna un DOI.

Al realizar una búsqueda global, en la pantalla de resultados en el lado izquierdo aparecen una serie de filtros (autor, fecha de creación, entidad financiadora, tema, licencia, tipo de recurso, Institución, proveedor y colección) para limitar la búsqueda. En la parte central tenemos los resultados obtenidos que muestra el tipo de publicación, el título, el autor, la fecha de creación y la fecha de modificación. Encima de los resultados aparece otro filtro para determinar el tipo de publicación que queremos visualizar (Todo, Proyectos, Registros, Preprints, Archivos y Usuarios) y el orden de los resultados (relevancia, fecha de creación/más nuevo, fecha de creación/más viejo, fecha de modificación/más nuevo, fecha de modificación/más viejo).

Los pasos a seguir para la publicación de un corpus de datos una vez registrados son los siguientes, primero hay que crear un proyecto, el proyecto debe tener un título, una descripción, un país donde alojarse y de forma opcional una plantilla. Con el proyecto creado, podemos añadir metadatos complementarios y añadir una plantilla de metadatos. Posteriormente se añaden los archivos necesarios. Cuando ya tenemos toda la información completa y todos los archivos subidos, se crea un registro del proyecto que será lo que se publique en el repositorio.

Al seleccionar un corpus de datos, en la pantalla se muestran diferentes secciones. Una sección superior con el título y debajo hay tres secciones, a la izquierda un menú de opciones, en la parte central la información de la opción seleccionada y a la derecha una serie de información de metadatos. Estos metadatos son el título, el autor, un resumen, la descripción, el tipo de registro, la fecha de registro y de publicación, el proyecto asociado, el DOI, los temas, la licencia y posibles formatos para citar la publicación. En la sección izquierda podemos escoger entre las opciones Información general, metadatos, archivos, recursos, wiki, componentes, enlaces, analítica, actividad reciente y comentarios, en función de lo que seleccionemos en la parte central vemos una información u otra, siendo las opciones más destacables la de archivos para poder descargar los archivos o los metadatos para obtener más información sobre la publicación.

Realizando una búsqueda, a día 20 de mayo de 2025, sin ningún seleccionado, no muestra el número total de resultados (cuando no se utiliza ningún filtro, al usar un filtro si muestra el dato total de ese filtro), pero si se suman los resultados de los diferentes tipos de recursos da un total aproximado de 370.000 siendo el más reciente del 15 de mayo de 2025. De esos resultados si filtramos por corpus de datos (Dataset) tenemos un total de 3.035 siendo el más reciente del 14 de mayo de 2025.

Búsquedas con filtros para datos y metadatos	Sí	Sí	Sí	Sí	Sí	Sí	Sí
Visionado online	Sí	Sí	Sí	Sí	Sí	Sí	Sí
Descarga directa	Sí	Sí	Sí	Sí	Sí	Sí	Sí
Acceso mediante API	Sí	Sí	Sí	NO	Sí	Sí	Sí
Precio/Cuotas							
Cuota de creación	Gratis	Sí	Gratis	Opcional	Opcional	Gratis	Gratis
Cuota de mantenimiento	Gratis	Sí	Opcional (según capacidad)	Opcional	Opcional	Gratis	Gratis

Tabla 2 Comparativa de repositorios de datos

Tras el análisis realizado, se ha podido comprobar que todos los repositorios están activos y que tienen actualizaciones constantes y frecuentes. Como era de esperar por el ámbito de acción, el repositorio más pequeño y con menos frecuencia de actualización es el repositorio UVaDoc ya que está limitado a los investigadores asociados a la universidad en vez de tener un ámbito mundial como los otros repositorios. Pese a ello, debido a la legislación aplicable, es necesario publicar los corpus de datos en el repositorio instruccional de la universidad, que es compatible con la publicación en otros repositorios, pero teniendo en cuenta las limitaciones sobre versiones y derechos de propiedad intelectual que establezcan los otros repositorios donde se vaya a publicar.

Las conclusiones que se pueden obtener tras el análisis son las siguientes:

- La ciencia abierta es el objetivo de los repositorios, pero no todos la implementan de la misma manera, mientras algunos no requieren del pago de ninguna cuota y el acceso a las publicaciones es público, otros requieren siempre el pago de cuotas o el pago de cuotas en función de si el acceso es público o restringido.
- Todos están alineados con los principios FAIR, ya que resulta esencial para que sus publicaciones se puedan encontrar y citar fácilmente. Su éxito depende de ello.
- Salvo UVaDOC y IEEE DataPort, el resto de repositorios forman parte de *Generalist Repository Ecosystem Initiative* (GREI), lo que facilita la interconexión entre ellos y el intercambio de información. Esto está alineado con los principios FAIR.
- En todos los repositorios los metadatos tienen mucha importancia, ya que es lo que facilita los principios FAIR y que las búsquedas tengan éxito. Todos tienen algún tipo de esquema o protocolo de metadatos como OAI-PMH o DublinCore.
- Todos los repositorios utilizan algún tipo de licencia para sus publicaciones, siendo las más populares las *Creative Commons*.
- A título personal, tras comprobar que encontrar cierta información es bastante difícil, creo que debería existir un estándar o protocolo donde haya una estructura similar en el diseño de los sitios webs de los repositorios y una información mínima que publicar para que el público

tenga accesible dicha información de forma fácil a la hora de elegir donde quiere publicar su investigación.

6. Revisión de literatura

Este capítulo trata sobre la revisión de la literatura que se ha realizado, tanto para establecer el estado del arte como para encontrar trabajos relaciones y finalmente la identificación y análisis de literatura donde los investigadores crean sus propios corpus de datos.

6.1. Estado del arte de investigaciones con datos biométricos

Sin perder de vista el contexto de este trabajo, crear un corpus de datos biométricos que permitan la identificación de personas y su publicación, hay que analizar el estado del arte desde diferentes puntos de vista, cada uno en función de los objetivos que se buscan en este trabajo como el uso de datos biométricos, el uso para identificación de personas, la publicación del corpus de datos en repositorios de datos y las consideraciones éticas y legales.

Hay mucha literatura existente sobre datos biométricos, por lo que puede resultar complicado seleccionar artículos que sean relevantes en nuestro contexto, pero gracias a las búsquedas realizadas para encontrar trabajos similares y las búsquedas para encontrar los artículos a analizar se obtiene un amplio abanico de artículos, que junto con sus referencias proporcionan una visión general con la que obtener el estado del arte.

Respecto al uso de datos biométricos y su uso para identificación y autentificación de personas, publicaciones como Fairhurst et al. [121], Arman et al. [122], Dargan y Kumar [123], Mandalapu et al. [124], Alrawili et al. [125] o Adil et al. [126] nos permiten conocer el estado actual de las investigaciones sobre de datos biométricos, tanto fisiológicos como conductuales. Dichas investigaciones muestran un conocimiento muy amplio sobre los datos biométricos, como tipos de datos que se pueden utilizar, qué características podemos inferir de cada uno de ellos, por ejemplo, procedencia, nacionalidad, edad, género o enfermedades de las personas. También se analizan las ventajas y desventajas de cada tipo de dato respecto a diferentes usos que pueden tener o las posibles debilidades que

pueden tener frente a ataques. Además, nos permite conocer diferentes métodos de identificación que existen y la precisión o robustez de cada uno de ellos entre otros muchos datos. Varias de estas publicaciones son revisiones de literatura existente, como el presente trabajo, pero al no tener los mismos objetivos, no se pueden considerar trabajos similares para el siguiente apartado, con los que comparar los resultados que obtengamos.

Respecto al marco ético y legal, los artículos encontrados solo hacen referencia al marco legal, no se han encontrado artículos que hagan un análisis detallado sobre el marco ético. Los artículos Makrushin et al. [127], Natgunanathan et al. [128], Jasserand [129] o Migdal y Rosenberger [130] sobre el marco legal hacen referencia al RGPD y la privacidad de los datos biométricos por ser considerados datos sensibles en las diferentes de protección de datos existentes en diferentes países. Otros artículos como Whiskerd et al. [131] analizan como influyen las leyes de privacidad sobre los datos biométricos, haciendo énfasis en que determinados datos además de identificar a las personas pueden revelar información los posibles problemas o enfermedades. Englobados dentro del marco legal, con el objetivo de salvar la privacidad de las personas, hay algunos artículos como Makrushin et al. [127] o Migdal y Rosenberger [130] que exponen la creación y el uso de datos sintéticos como sustitución de datos de personas reales para no estar sujetos a las leyes sobre protección de datos y poder hacer un uso libre de estos datos artificiales.

Dentro de la literatura existente, hay artículos como Klie et al. [132], Fallahi et al. [133] o Kasprowski y Rigas [134] que hablan sobre la importancia de la calidad de los datos y los metadatos para evitar errores y sesgos en los corpus de datos. También se señala la importancia de diferentes características que hay que tener en cuenta en la recogida de los datos como la limpieza y calibración de los dispositivos de recogida de datos, el número de participantes, las repeticiones por cada importancia o la influencia del lapso de tiempo entre las muestras de cada participante. Al no exponer unos resultados concretos de número mínimo de participantes o repeticiones por cada participante, no es posible comparar nuestras conclusiones con las de estos estudios.

En la literatura también artículos como Fairhurst et al. [121], Mandalapu et al. [124], Migdal y Rosenberger [130], Fallahi et al. [133], Ren et al. [135] o Nnamoko

et al. [136] que exponen o nombran repositorios de datos concretos para un tipo de dato biométrico concreto o los conjuntos de datos concretos que han usado en su investigación, explicando sus características y posibles usos. Pero no realizan un análisis de los repositorios de datos.

Debido a la legislación actual que obliga a que las investigaciones de personal público o investigaciones financiadas mayoritariamente con fondos públicos se publiquen de forma abierta y promueven la creación de repositorios institucionales, han surgido a lo largo del tiempo organismos enfocados en mejorar las infraestructuras tecnológicas como la European Open Science Cloud (EOSC) [137] o la Red de Bibliotecas Universitarias y Científicas Españolas (REBIUN) [62] que permite el acceso a los repositorios de todas las bibliotecas asociadas. A su vez desde la Conferencia de Rectores y Rectoras de las Universidades Españolas (CRUE) [63] también se han creado una serie de documentos [138], guías [139] e informes [140] o [141] cuyo objetivo es fomentar y desarrollar la ciencia abierta y ver su evolución a lo largo del tiempo realizando mediciones sobre los repositorios institucionales. En Estados Unidos también existe preocupación por la ciencia abierta y por parte del National Institutes of Health (NIH) (Institutos Nacionales de la Salud en español) [142] se creó la Generalist Repository Ecosystem Initiative (GREI) (Iniciativa del Ecosistema de Repositorios Generalistas en español) [110] con el fin de desarrollar enfoques colaborativos para la gestión y el intercambio de datos mediante la inclusión de repositorios generalistas en el ecosistema de datos de los NIH, también busca facilitar la búsqueda y el descubrimiento de datos financiados por los NIH y ayudar a los investigadores a cumplir los principios FAIR. Los repositorios que actualmente forman parte de esta iniciativa son Dataverse, Dryad, Figshare, Mendeley Data, Open Science Framework (OSF), Vivli, Zenodo, de los cuales ya se ha hablado anteriormente en el apartado de repositorios.

Tras el análisis de parte de la literatura existente, se puede afirmar que la investigación sobre los datos biométricos está muy extendida, lo que ha facilitado un trasvase de conocimiento a la sociedad para que se puedan emplear gran variedad de datos biométricos para múltiples usos. Estas investigaciones no solo han tenido en cuenta la parte más técnica, sino que también se han investigado sus implicaciones legales y éticas y los problemas que puede acarrear su uso. Debido a esta situación, los investigadores también han puesto su esfuerzo en los retos futuros que implica el uso de datos biométricos, siendo uno de los

principales retos los problemas de seguridad, ya que, al no poder alterarse los datos biométricos, si se produce una fuga de información, las personas afectadas no pueden reemplazar esos datos como si cambiaran una contraseña o anularan una tarjeta.

También se puede afirmar que desde hace tiempo se viene dando mucha importancia a la ciencia abierta, fomentando el acceso público a las investigaciones y los datos para favorecer el trasvase de información desde la parte investigadora a la sociedad para mejorarla aplicando los avances y descubrimientos realizados.

6.2. Trabajos relacionados

Se han realizado varias búsquedas tanto en Google Scholar como en Consensus.app con el objetivo de encontrar trabajos relacionados con el que se está desarrollando para poder tener referencias sobre lo que ya se ha llevado a cabo y para poder comparar nuestros resultados son los resultados que hubieran obtenido dichos trabajos. Pero no se ha obtenido un resultado satisfactorio, ya que los trabajos que se han encontrado tratan sobre temas relacionados con el contexto de este trabajo, pero ninguno de los trabajos habla sobre la cantidad de datos que tiene que tener un corpus de datos o analiza los repositorios donde publicar dichos datos.

Algunas de las búsquedas realizadas han sido:

- How to create a biometric dataset
- How to publish a biometric dataset
- Create biometric dataset
- Characteristics of a biometric dataset
- Biometric dataset
- Publish biometric dataset
- Compare data repository

Todas estas búsquedas arrojan muchos resultados relacionados con datos biométricos, pero ninguno de ellos tiene el mismo enfoque global que el presente trabajo.

En la revisión de la literatura para conocer el estado del arte, se han encontrado muchas publicaciones que están hasta cierto punto relacionadas con este trabajo, pero cada una de esas publicaciones está enfocada en uno o dos puntos concretos, por lo que no se ha conseguido identificar alguna publicación que recoja todos los puntos tratados o que proporcione unas pautas o conclusiones concretas para las características que tiene que tener un buen corpus de datos.

Si lo tratamos por partes, si se han encontrado trabajos relacionados con los repositorios, por ejemplo comparativas de repositorios como Stall et al. [143] o [144] o análisis sobre la calidad de los corpus de datos como Kasprowski y Rigas [134] que pueden servir para comparar parcialmente sus características con las conclusiones que se obtengan en este estudio.

Respecto a los trabajos relacionados sobre el análisis de repositorios, tanto el presente análisis como los de la literatura han analizado diversos repositorios y se ha producido una coincidencia con los repositorios Dryad, Figshare, Mendeley Data, OSF y Zenodo.

Respecto a los datos analizados de cada repositorio también existen coincidencias en ciertos campos como en el formato de los archivos admitidos y el tamaño máximo permitido, la existencia o no de cuotas o tarifas, la asignación de DOI o las licencias. Otros datos solo se comparten con uno de los otros estudios, como la existencia de una API, y otros datos no se comparten con ninguno como el número total de publicaciones o la fecha de la última publicación, aunque estos datos son muy volátiles ya que cambian prácticamente a diario, pero pueden dar una idea de lo activo que es un repositorio a lo largo del tiempo.

En la información que se ha analizado por los tres estudios, con mayor o menor explicación sobre los resultados todos tienen el mismo resultado a excepción los límites de tamaño de los archivos individuales o del corpus de datos completo, donde si hay variaciones que pueden deberse a cambios de políticas a lo largo del tiempo.

6.3. Selección y revisión de la literatura

Realizar una revisión bibliográfica es una tarea ardua y compleja, para realizar esta revisión de una forma que sea clara, transparente y reproducible, es necesario que el proceso sea estructurado y sistemático para que el proceso de búsqueda y selección de artículos este correctamente documentado y explicado con el objetivo de que los resultados sean óptimos y se minimicen los sesgos. Para este trabajo se ha utilizado el enfoque PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) en su versión 2020 [145].

Con la revisión de la literatura se quiere obtener una idea de lo que se puede entender como un buen corpus de datos, estableciendo para ello unas características que debe cumplir el corpus de datos. Dichas características comprenden desde el número de individuos y características de los mismos o repeticiones de la prueba a realizar por individuo, hasta el total de datos recogidos pasando por datos necesarios para establecer el marco legal como que normativas ha cumplido o datos necesarios para establecer el marco ético como si ha sido supervisado por un comité de ética, quien lo ha financiado o si ha publicado los datos.

De esta forma la lista de características o preguntas a responder es la siguiente:

- Datos estadísticos:
 - Año de publicación.
 - Número de autores.
 - Número de Organizaciones implicadas.
 - Palabras clave.
 - Tipo de dato recogido.
 - Número de participantes.
 - Repeticiones por participante.
 - Diferenciación de participantes:
 - Edad.
 - Genero.
 - Procedencia.
 - Otros.

- Marco Legal:
 - Menciona sobre qué ley de privacidad se rige.
 - Menciona normas o protocolos utilizados.
- Marco Ético
 - Menciona si ha sido supervisado por un comité de ética.
 - Especifica el comité de ética.
 - Menciona quien ha financiado la investigación.
 - Menciona si existe conflicto de intereses.
 - Menciona si ha publicado los datos utilizados o son accesibles.

6.3.1. Criterios de elección de la literatura

Para seleccionar la literatura a revisar, se va a usar el acceso a revistas y repositorios que están disponibles gracias al acceso que proporciona la universidad. Estas revistas son:

- ScienceDirect
- MDPI
- IEEE Xplore

Además, para encontrar investigaciones relacionadas con nuestro ámbito de interés, se van a realizar búsquedas en función de las siguientes palabras clave:

- IoT
- Ambiental Intelligence
- Ambient Intelligence
- User Identification
- Access Control
- Human activity recognition systems
- HARS

De entre todos los artículos que nos proporcionen dichas búsquedas se van a seleccionar varios siguiendo criterios de ubicación y temporales para ver si existen diferencias significativas. Además, se optará por publicaciones que tengan varias citaciones, entiendo así que son publicaciones de referencia.

Los criterios de ubicación van a ser seleccionar investigaciones de Asia (China principalmente), Europa y América (Estados Unidos principalmente).

Los criterios temporales van a venir definidos por las fechas de publicación de las diferentes legislaciones de protección de datos, siendo los hitos claves:

Ley	Publicación	Entrada en Vigor
Directiva 95/46/CE de la Unión Europea	1995	1995
Ley de Protección de Información Personal (APPI) de Japón	2003	2005
Ley de Protección de Datos (PIPA) de Corea del sur	2011	2011
Reglamento General de Protección de Datos (RGPD) de la Unión Europea	2016	2018
Ley de Privacidad del Consumidor de California (CCPA)	2018	2020
Ley General de Protección de Datos (LGPD) de Brasil	2018	2020
Ley de Protección de Información Personal (PIPL) de China	2021	2021
Ley de Protección de Datos Personales (PDPB) de India	2019	2023

Tabla 3 Años de publicación y entrada en vigor de las leyes de protección de datos

En función del ámbito geográfico de la publicación se seleccionarán artículos tanto previos como posteriores a la entrada en vigor de su correspondiente ley de protección de datos para comprobar si se produce algún cambio sustancial en las publicaciones.

Se llevará un registro de las búsquedas realizadas anotando donde se realiza la búsqueda, los términos de búsqueda, filtros utilizados como fechas o países.

Con los criterios descritos se hará la selección de artículos de la siguiente forma:

- Artículos de Estados Unidos anteriores a 2018 (previos a la publicación de CCPA).

- Artículos de Estados Unidos posteriores a 2020 (posteriores a la entrada en vigor de CCPA).
- Artículos de Europa anteriores a 1995 (previos a la publicación y entrada en vigor de la Directiva 95/46/CE).
- Artículos de Europa entre 1995 y 2016 (periodo entre la Directiva 95/46/CE y la publicación de RGPD).
- Artículos de Europa posteriores a 2018 (posteriores a la entrada en vigor de RGPD).
- Artículos de China anteriores a 2021 (previos a la publicación y entrada en vigor de PIPL).
- Artículos de China posteriores a 2021 (posteriores a la entrada en vigor de PIPL).

En Estados Unidos el periodo entre el 28 de junio de 2018, fecha de publicación de CCPA y el 1 de enero de 2020, fecha de entrada en vigor de CCPA, inicialmente no está contemplado ya que es un periodo de transición donde las investigaciones se están adaptando a la nueva ley y puede que empiecen a cumplir con las especificaciones de la ley. En función de los resultados de las búsquedas en la literatura se incluirá ese intervalo temporal o no.

En China no existe este desfase, ya que la ley de protección de datos se publicó el 20 de agosto de 2021 y entró en vigor el 1 de noviembre de 2021, por lo que periodo de adaptación fue de 2 meses, que es prácticamente inexistente.

Cuando una investigación este realizada por instituciones de diversos países, entre los cuales esté involucrado algún país del que estamos realizando búsqueda de investigaciones, será tenido en cuenta ya que la investigación debe cumplir la legislación de todos los países involucrados.

Si la investigación analizada está realizada por instituciones de dos o más países objeto de estudio en esta investigación, la elección de introducirlo en uno o de los apartados vendrá dada por la fecha de la publicación, si cumple o no alguna de las legislaciones necesarias, y de la cantidad de artículos que se hayan encontrado en esas secciones. Cada investigación solo se tendrá en cuenta una vez.

Con la distribución de artículos mencionada se pretende encontrar una cantidad suficiente de artículos válidos sobre los que realizar el análisis. Una vez finalizada

la búsqueda, se añadirán dos artículos iniciales propuestos por los tutores (Vegas et al. [146] y Xu et al. [147]).

Si resultara necesaria una búsqueda más amplia para obtener más artículos o detalles sobre los artículos se realizaría una búsqueda más general en Google Scholar o Consensus.app. Otra posible opción de búsqueda de artículos además de las búsquedas mencionadas, sería a través de las referencias existentes en los artículos que se seleccionen inicialmente. Para el posible uso de estas referencias, se utilizarían los mismos criterios que la selección inicial, fijándose en las palabras clave, las menciones, la fecha de publicación, el país y que el artículo esté disponible. Para las referencias se utilizaría solo un nivel, es decir referencias de los artículos iniciales, no se profundizará a más niveles.

En caso de que se descarte algún artículo de la lista inicial, se indicará el motivo por el que se ha descartado para evitar introducir un sesgo personal o de cualquier otro tipo.

6.3.2. Análisis inicial

Antes de realizar las búsquedas, se analizan las opciones de búsqueda que ofrece cada revista además de los resultados que ofrece.

- IEEE Xplore: En la búsqueda avanzada permite filtrar por varios campos, entre ellos las palabras clave (*Index Terms*), año en que se añadió o por año en el que se publicó. No permite búsqueda por cantidad de citaciones o país donde se realizó la investigación. En este caso usaremos el año en que se publicó y las palabras clave. Una vez realizada la búsqueda por cada artículo nos detalla el título, los autores, donde se ha publicado, el año de publicación, el tipo de publicación y la editorial. Dispone de filtros para ordenar los artículos, donde aparece un filtro de relevancia y de los más citados (*Most Cited by Papers*), cuando se ordenan por el número de citaciones, por cada artículo aparece el número de veces que ha sido citado. También posee una serie de filtros, de entre todos ellos se van a utilizar el de tipo de publicaciones, restringiendo a los tipos a “Conferences”, “Journals” y “Magazines”, descartando “Early Access Articles”, “Standards”,

“Courses” y “Books” ya que las publicaciones de investigaciones se publican en los tipos seleccionados. También se usará el filtro de tipo de acceso, seleccionando “Subscribed Content” para que solo nos muestre artículos a los que podemos acceder. Permite exportar los resultados de la búsqueda con un límite de 1000 artículos.

- MDPI: En la búsqueda avanzada inicialmente no permite filtrar por muchos campos, pero si está disponible el campo de palabras clave (*keywords*) que es que vamos a usar. Una vez realizada la búsqueda, por cada artículo listado nos muestra el título, los autores, donde se ha publicado, el año de publicación, el enlace al registro doi, el número de veces que se ha visto el artículo y el *abstract*. Dispone de una serie de filtros de ordenación como relevancia y número de veces citado, cuando se ordenado por número de veces citado, por cada artículo aparece el número de veces que ha sido citado. También hay filtros adicionales para selección de artículos de entre los cuales se utilizarán el filtro de año y el de países/regiones y un filtro por tipo de artículo que se usará para filtrar por “Article” y “Conference Report” para obtener los resultados de las investigaciones. Permite exportar los resultados de la búsqueda limitado a los que se ven en la hoja del navegador, por lo que el máximo son 200 artículos.
- ScienceDirect: En la búsqueda avanzada no permite buscar directamente solo por palabras clave, sino que busca en el conjunto título/resumen/palabras clave (*Title/Abstract/Author-specified Keywords*), tiene un filtro de años. Una vez realizada la búsqueda, por cada artículo listado nos muestra el título, los autores, donde se ha publicado, el mes y año de publicación, además de desplegables para permitir ver más información. Como filtros de ordenación solo dispone de relevancia y fecha. Dispone de filtros adicionales de selección de artículos, de entre todos ellos se utilizarán el filtro de “Subscribed journals” para que se muestren solo artículos que están accesibles, también se usarán el filtro de años, el de tipo de artículos limitándolo a los tipos “Research articles” y “Conference abstracts” que es donde se publican los resultados de las investigaciones, también se hará uso del filtro de áreas temáticas para limitarlo a nuestras áreas de interés que serán de forma amplia “Computer Science” y “Engineering” en caso de que estén disponibles. En este caso no hay

ninguna opción que haga referencia a las citaciones. Permite exportar los resultados de la búsqueda limitado a los que se ven en la hoja del navegador, por lo que el máximo son 100 artículos.

Con estos criterios de búsqueda, primero tendremos que realizar una búsqueda general con los términos a buscar y después aplicar los diferentes filtros que necesitemos, además de ordenar por citaciones para obtener los más citados. Después en función del título y el *abstract* habrá que seleccionar los artículos que sean relevantes para el análisis que estamos realizando.

Una vez analizados los criterios de búsqueda disponibles, se proceden a realizar las búsquedas. Por cada una de las revistas analizadas, primero se realiza una búsqueda general sin filtros y después se irán aplicando filtros.

Las búsquedas realizadas usando la palabra clave sin ningún tipo de filtro han sido las siguientes:

Revista	Fecha Búsqueda	Palabra Clave	Total	Año inicial	Año final
IEEE Xplore	30/03/2025	IoT	109.336	1991	2025
		Ambiental Intelligence	21	1997	2024
		Ambient Intelligence	6.758	1895	2025
		Aml	6.164	1914	2025
		User Identification	38.409	1960	2025
		Access Control	189.716	1896	2025
		Human activity recognition systems	14.052	1925	2025
		HARS	2.403	1936	2025
MDPI	30/03/2025	IoT	33.660	2010	2025
		Ambiental Intelligence	491	2008	2025
		Ambient Intelligence	1.587	2008	2025
		Aml	8.138	2000	2025
		User Identification	820	2009	2025
		Access Control	7.807	2002	2025
		Human activity recognition systems	174	2009	2025
		HARS	26.635	1997	2025
ScienceDirect	30/03/2025	IoT	81.202	2001	2025
		Ambiental Intelligence	32.971	2002	2025

Ambient Intelligence	32.971	2002	2025
Aml	139.758	2001	2025
User Identification	457.076	2001	2025
Access Control	2.286.060	1997	2025
Human activity recognition systems	619.794	2001	2025
HARS	70.304	2001	2025

Tabla 4 Búsquedas sin filtros

Las búsquedas realizadas usando palabra clave en el filtro palabra clave sin otros filtros han sido las siguientes:

Revista	Fecha Búsqueda	Palabra Clave	Total	Año inicial	Año final
IEEE Xplore	30/03/2025	IoT	64.567	1992	2025
		Ambiental Intelligence	0	-	-
		Ambient Intelligence	2.706	1895	2025
		Aml	811	2000	2025
		User Identification	20.631	1970	2025
		Access Control	114.825	1896	2025
		Human activity recognition systems	6.581	1958	2025
		HARS	660	2007	2025
MDPI	30/03/2025	IoT	5.196	2012	2025
		Ambiental Intelligence	117	2009	2025
		Ambient Intelligence	117	2009	2025
		Aml	152	2012	2025
		User Identification	33	2011	2024
		Access Control	427	2009	2025
		Human activity recognition systems	3	2017	2023
		HARS	201	2008	2025
ScienceDirect	30/03/2025	IoT	16.452	2001	2025
		Ambiental Intelligence	598	2000	2025
		Ambient Intelligence	598	2000	2025
		Aml	9.782	2001	2025
		User Identification	7.905	2001	2025

		Access Control	41.829	2001	2025
		Human activity recognition systems	1.703	2001	2025
		HARS	1.545	2001	2025

Tabla 5 Búsquedas con el filtro "palabra clave"

Las búsquedas realizadas por palabra clave aplicando los filtros explicados anteriormente de artículos accesibles y publicados en el contexto de artículos de investigación han sido las siguientes:

Revista	Fecha Búsqueda	Palabra Clave	Total	Año inicial	Año final
IEEE Xplore	30/03/2025	IoT	64.046	1992	2025
		Ambiental Intelligence	0	-	-
		Ambient Intelligence	2.684	1895	2025
		Aml	802	2000	2025
		User Identification	20.265	1970	2025
		Access Control	112.556	1896	2025
		Human activity recognition systems	6.545	1958	2025
		HARS	658	2007	2025
MDPI	30/03/2025	IoT	4.336	2012	2025
		Ambiental Intelligence	102	2009	2025
		Ambient Intelligence	102	2009	2025
		Aml	124	2012	2025
		User Identification	31	2011	2024
		Access Control	392	2009	2025
		Human activity recognition systems	3	2017	2023
		HARS	174	2008	2025
ScienceDirect	30/03/2025	IoT	6331	1985	2025
		Ambiental Intelligence	204	1998	2025
		Ambient Intelligence	204	1998	2025
		Aml	345	2000	2025
		User Identification	2257	2001	2025
		Access Control	6082	2001	2025
		Human activity recognition systems	371	1999	2025
		HARS	248	1980	2025

Tabla 6 Búsquedas con filtro "palabra clave", accesibilidad y tipo de artículo

Si observamos la Tabla 6 podemos ver que puede resultar complicado encontrar artículos europeos previos a 1995 ya que no hay muchos artículos y los que se encuentran puede que no estén relacionados con nuestro contexto de identificación de usuarios mediante datos biométricos.

Los motivos que se plantean para rechazar artículos serán los siguientes:

- La investigación a la que se refiere el artículo no ha creado un corpus de datos propio, ya sea porque se ha usado corpus de datos existentes o porque el artículo es una revisión de literatura.
- Los artículos no explican las características del grupo de personas utilizado para la obtención del corpus de datos.
- Los artículos no se ajustan a los parámetros establecidos anteriormente en cuanto al país o fecha de la investigación.
- El artículo no se ajusta completamente a nuestro contexto por tener una temática diferente como por ejemplo centrarse en la seguridad, la encriptación de los datos biométricos o las propuestas de esquemas de autenticación.

6.3.3. Búsqueda adicional de literatura

En este punto hay que hacer una aclaración inicial, aunque los criterios de búsqueda y selección se han consensuado entre el autor y los tutores de este proyecto de investigación, el autor es la única persona que ha realizado el análisis y cribado de los artículos.

Las búsquedas detalladas en el punto anterior ofrecieron unos resultados poco alentadores, ya que de los resultados obtenidos en las búsquedas tras analizar una media de 20 artículos por búsqueda no se encontró ningún artículo que se ajustara a los criterios de selección determinados. Por este motivo, se tuvieron que emplear otros parámetros de búsqueda.

El detalle de las siguientes búsquedas utilizadas para encontrar los artículos para el desarrollo de esta investigación ha sido:

Editorial	Fecha Búsqueda	Palabra Clave	Total artículos	Artículos analizados	Artículos válidos
ScienceDirect	24/05/2025	Biometric Identification	29040	60	1
	27/05/2025	Biometric Identification USA	9472	45	0
	31/05/2025	Biometric Identification China	7214	45	5
	01/06/2025	Biometric Identification wearable	3852	45	1
IEEE Xplore	02/05/2025	Biometric Identification USA	1863	45	1
	08/05/2025	Biometric Identification China	1679	45	1
	17/05/2025	Biometric Identification wearable	321	45	0
MDPI	07/06/2025	Biometric Identification [filtro países de la Unión Europea]	2883	45	0

Tabla 7 Búsquedas adicionales

De entre todas las búsquedas realizadas y artículos analizados, los motivos para descartar artículos han sido:

- No ajustarse al contexto de la investigación.
- El artículo es una revisión de artículos, no una investigación propia.
- La investigación no está realizada en alguno de los países determinados para el análisis.
- La investigación no crea su propio corpus de datos ya que utiliza un corpus que está accesible desde internet.
- El artículo no describe como se ha formado el corpus de datos.
- No realiza identificación de humanos, se identifican animales u objetos.

Al realizar las búsquedas indicando los países en los términos de búsqueda, se ha tenido algún problema con el término USA en la editorial IEEE Xplore, ya que a veces si la institución a la que pertenecen los investigadores es de Estados Unidos, se identifica la institución y el estado, pero no se llega a identificar USA como país y entonces no aparece en las búsquedas.

6.3.4. Artículos seleccionados

En este apartado se va a presentar una lista con los artículos analizados que son el resultado de los artículos válidos procedentes de las búsquedas y los artículos de ejemplo proporcionados por los tutores. En dicha lista se muestra el nombre del artículo, el país donde se realizó la investigación y el año de publicación. Además, se ha agrado una primera columna con un ID para identificar el artículo y poder enlazarlo con la tabla del apéndice donde se presenta toda la información recogida de los artículos.

De los artículos de ejemplo indicados anteriormente, tras su análisis, el artículo "*A Two-Step Fusion Method of Wi-Fi FTM for Indoor Positioning*" no es seleccionable ya que no se usa para identificación de personas si no para localización en el interior mediante WiFi FTM, además, las mediciones se realizan con un smartphone y no hay mediciones realizadas con personas.

La lista de artículos es:

ID	Título	País	Año
1	Identifying Users from the Interaction with a Door Handle, Vegas et al. [146]	España	2020
2	ECG-based biometric recognition under exercise and rest situations, Cui et al. [148]	China	2021
3	Quality-Specific Hand Vein Recognition System, Wang et al. [149]	China	2017
4	3D dental biometrics: Alignment and matching of dental casts for human identification, Zhong et al. [150]	China y Singapur	2013

5	Towards online applications of EEG biometrics using visual evoked potentials, Zhao et al. [151]	China	2021
6	ECG-based biometric under different psychological stress states, Zhou et al. [152]	China	2021
7	Physiological and behavioral lip biometrics: A comprehensive study of their discriminative power, Wang y Liew [153]	China y Australia	2012
8	I Sense You by Breath: Speaker Recognition via Breath Biometrics, Lu et al. [154]	China	2017
9	TrueHeart: Continuous Authentication on Wrist-worn Wearables Using PPG-based Biometrics, Zhao et al. [155]	EEUU y China	2020
10	Mapping sweat pores for biometric identification based on a donor-acceptor hydrophilic fluorescent probe, Zhao et al. [156]	China	2024

Tabla 8 Artículos seleccionados

6.3.5. Riesgo de sesgo

Para la selección de la literatura se han usado unos criterios concretos acorde al contexto de este proyecto de uso de datos biométricos y los marcos éticos y legales de uso. Estos criterios pueden inducir a que los artículos analizados contengan algún tipo de sesgo ya que zonas geográficas que no se han tenido en cuenta a la hora de seleccionar los artículos como pueden ser los continentes de África, Oceanía o América del Sur o gran parte de Asia al incluir solo China. También se puede producir algún tipo de sesgo en cuanto al tipo de dato biométrico analizado, ya que hay muchos datos biométricos y en la literatura analizada no están reflejados todos ellos.

No se han podido evitar estos posibles sesgos ya que es imposible estudiar toda la literatura existe a la hora de seleccionar los artículos a analizar y determinar un número concreto de artículos a analizar.

6.4. Resultados obtenidos

Para la presentación de los resultados obtenidos, se van a realizar diferentes agrupaciones y análisis de datos para poder obtener una evolución y unas conclusiones.

Primero se van a exponer una serie de datos obtenidos a través del análisis de los artículos:

- Tipo de dato biométrico:

El listado de datos biométricos que se han utilizado en los artículos analizados es el que se muestra a continuación

Dato Biométrico	Número de artículos analizados
Apertura de puerta	1
Electrocardiograma (ECG)	2
Venas de la mano	1
Molde dental 3D	1
Electroencefalograma (EEG)	1
Labios	1
Respiración	1
Señal cardiaca fotopletismografía (PPG)	1
Poros sudoríparos de la mano	1

Tabla 9 Listado de datos biométricos utilizados

- Número de autores

En este apartado se presenta la distribución en función del número de autores que han publicado los artículos analizados.

Número de autores	Número de artículos
1	0
2	2
3	1
4	2
5	1
6	1

7	2
8	0
9	1

Tabla 10 Distribución del número de autores por artículo

Una visualización gráfica de la distribución es la siguiente:

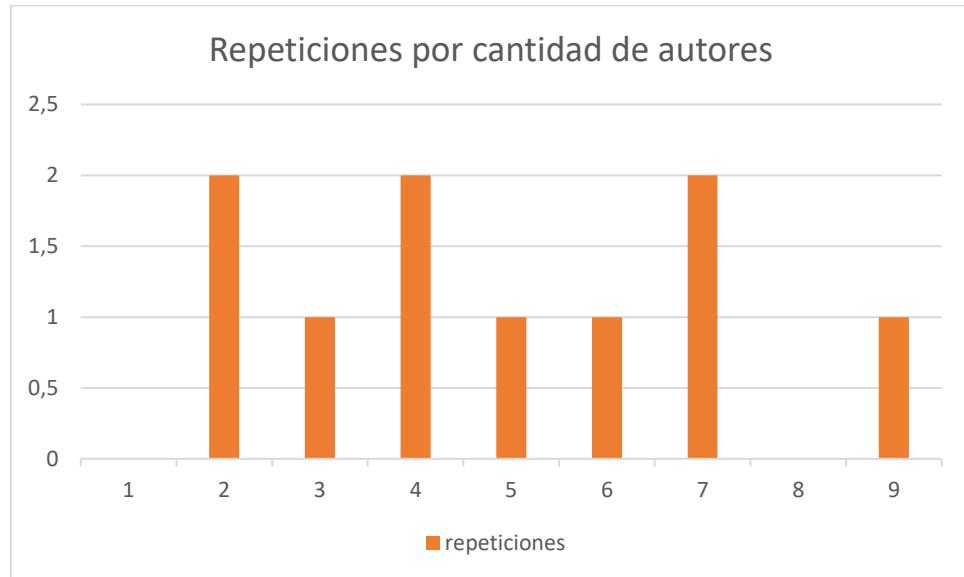


Ilustración 1 Distribución del número de autores por artículo

El número mínimo de autores es de 2 y el máximo de 9, siendo la media de 5 autores por artículo.

- Número de instituciones

En este apartado entendemos por institución el lugar al que está adscrito un autor, que puede ser una empresa, una universidad u otro tipo de organización. En los artículos donde aparece una misma institución repetida diferenciando entre áreas o departamentos, lo hemos resumido en una única aparición, por ejemplo si en un artículo apareciera Departamento de Informática, Universidad de Valladolid y Departamento de Física, Universidad de Valladolid, se cuenta como una institución por ser de la misma universidad.

La distribución del número de instituciones que intervienen en los artículos es la que se muestra a continuación.

Número de instituciones	Número de artículos
1	5
2	2
3	1
4	1
5	1

Tabla 11 Distribución del número de instituciones por artículo

El número mínimo de instituciones es de 1 y el máximo de 5, siendo la media de 2 instituciones por artículo.

Una visualización gráfica de la distribución es la siguiente:

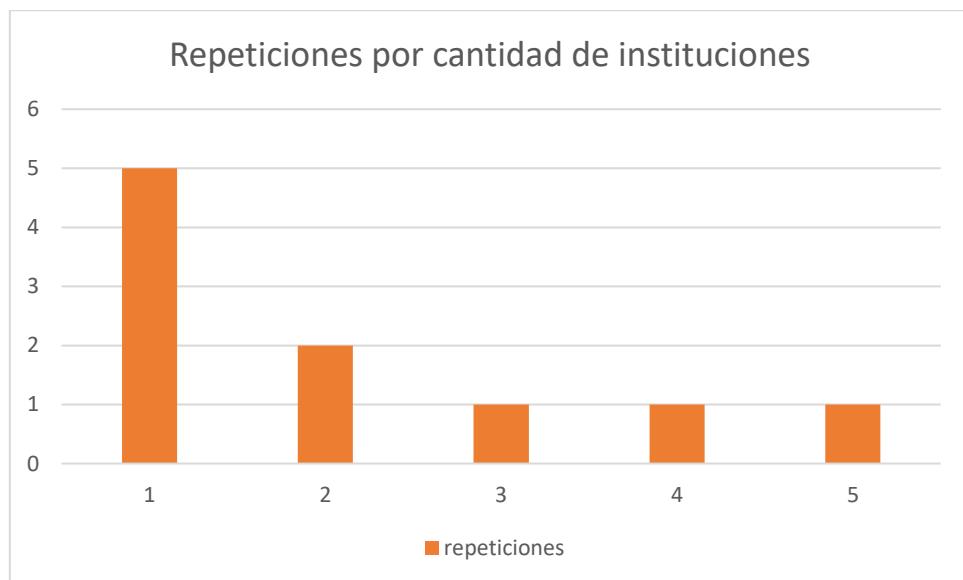


Ilustración 2 Distribución del número de instituciones por artículo

- Número de citaciones

La distribución del número de citaciones que tienen los artículos es la que se muestra a continuación, para mejorar la visibilidad se agrupan las citaciones en grupos de 10, salvo el 0 que lo dejamos de forma individual:

Número de citaciones	Número de artículos
----------------------	---------------------

0	2
1-10	0
11-20	3
21-30	1
31-40	1
41-50	2
51-60	0
61-70	1

Tabla 12 Distribución del número de citaciones por artículo

El número mínimo de instituciones es de 0 y el máximo de 63, siendo la media de 25 citaciones por artículo.

Una visualización gráfica de la distribución es la siguiente:

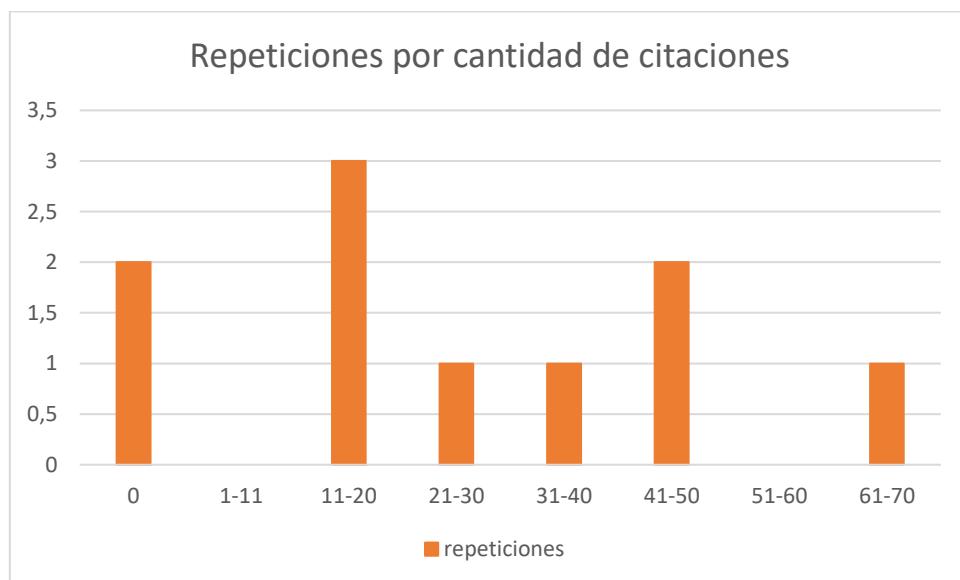


Ilustración 3 Distribución del número de citaciones por artículo

Se realiza una comparativa con la media de citaciones de los artículos en función del año de publicación y se obtiene la siguiente gráfica de distribución:

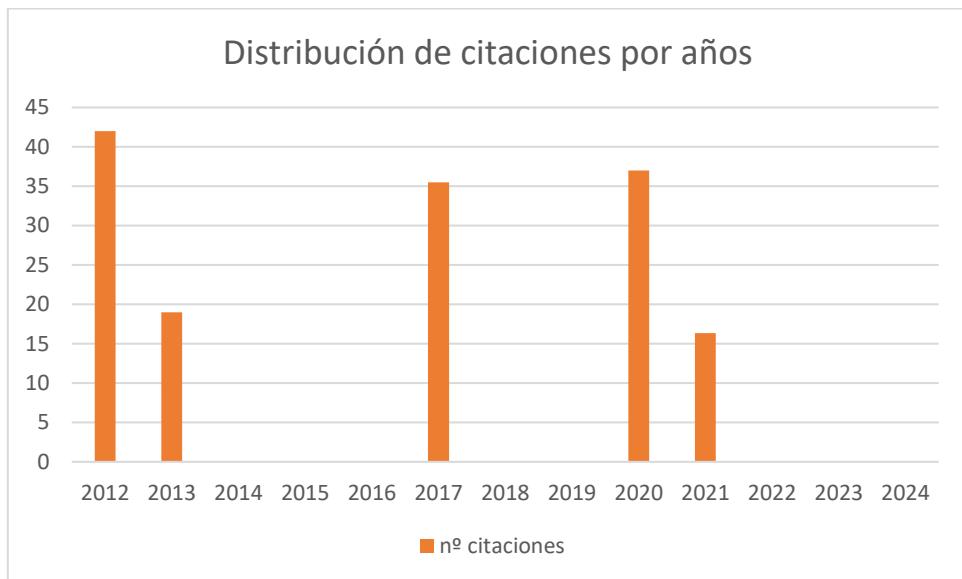


Ilustración 4 Distribución del número de citaciones por año de publicación

- Número de palabras clave

Antes del análisis de este apartado, hay que mencionar que estas palabras clave se refieren a las definidas por los autores, ya que se han encontrado casos que hay varios grupos de palabras claves, un grupo definido por el autor y otro definido por la editorial.

La distribución del número de palabras clave que tienen los artículos es la que se muestra a continuación.

Número de palabras clave	Número de artículos
0	2
1	0
2	0
3	2
4	0
5	3
6	2
7	0
8	1

Tabla 13 Distribución del número de palabras clave por artículo

El número mínimo de palabras clave es de 0 y el máximo de 8, siendo la media de 4 palabras clave por artículo.

Una visualización gráfica de la distribución es la siguiente:

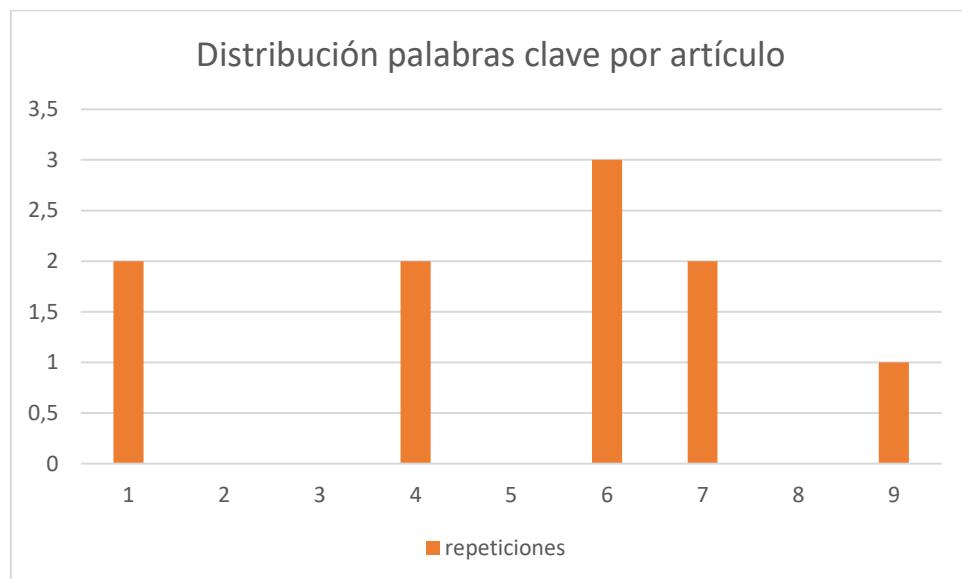


Ilustración 5 Distribución del número de palabras clave por artículo

Los resultados obtenidos en cuanto al **marco ético** son los siguientes:

- Comité de ética:

Del total de artículos analizados, el 20% mencionan la supervisión de un comité de ética.

- Consentimiento informado:

Del total de artículos analizados, el 30% mencionan la firma de un consentimiento informado por parte de los participantes.

- Financiación:

Del total de artículos analizados, el 90% mencionan la fuente de financiación del proyecto de investigación.

En este apartado es conveniente mencionar que se han encontrado situaciones en las que el comentario sobre la financiación de la investigación se ha encontrado

durante el análisis del artículo en su versión online, pero en la versión en PDF descargable esa mención a la financiación no se encontraba.

- Declaración de conflicto de intereses:

Del total de artículos analizados, el 50% mencionan si existe o no algún conflicto de intereses por parte de los autores en la realización del proyecto de investigación.

Los resultados obtenidos en cuanto al **marco legal** son los siguientes:

- Leyes de protección de datos:

Ninguno de los artículos analizados menciona que leyes de protección de datos estaban vigentes durante la investigación y publicación del artículo.

- Normas o protocolos:

Ninguno de los artículos analizados menciona que normas o protocolos han seguido o utilizado durante el transcurso de la investigación realizada.

Los resultados obtenidos sobre la composición del **corpus de datos** son los siguientes:

Para esta sección, hay que hacer una aclaración previa, hay algunos artículos que en la explicación de su corpus de datos dos grupos de participantes. Cuando se ha dado esta situación, en función de los parámetros “Tomas por usuario” y “Número de repeticiones/sesiones” analizados se ha contabilizado como un único grupo, sumando la cantidad de gente de cada grupo cuando el resto de parámetros eran iguales o se ha separado en dos grupos cuando los parámetros eran diferentes.

- Número de participantes:

Para este tipo de dato, el número mínimo de participantes ha sido de 5, el número máximo ha sido de 260 y la media de participantes es de 46.

- Número de repeticiones o sesiones:

El número mínimo de repeticiones/sesiones es de 1, el máximo es de 32 y la media es de 5.

- Número de muestras por participante en cada sesión:

El número mínimo de muestras por participante en cada sesión es de 1, el máximo es de 200 y la media es de 61.

- Número total de datos obtenidos:

Este dato hace referencia al total de datos que se tienen en el corpus de datos, es el resultado de multiplicar el número de participantes por el número de sesiones y por el número de muestras por sesión. El valor mínimo es de 60, el valor máximo es de 36000 y la media es de 6150.

- Diferenciación entre participantes:
 - Edad:

En este apartado se estudia cuantos artículos han mencionado la diferenciación por edad en el corpus de datos. No hay un estándar claro a la hora de mencionar la edad de los participantes, ya que algunos artículos detallan la edad mínima y la máxima, otros mencionan la edad media y otros mencionan la edad media con un dato añadido de variación, por lo que se van a realizar un doble análisis, primero distinguiendo entre SI y NO y después dentro del SI otro análisis con la información proporcionada. En el primer análisis para los artículos que mencionan más de un grupo de datos, se va a tomar que solo hay uno y para el segundo análisis se tomara cada grupo por separado.

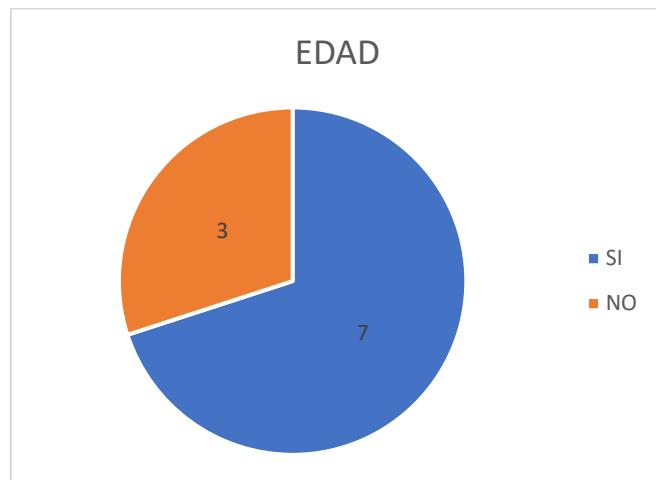


Ilustración 6 Ratio diferenciación por edad

Ahora se muestra un gráfico donde vemos la distribución de las diferentes formas en que se muestra la diferenciación por edad.

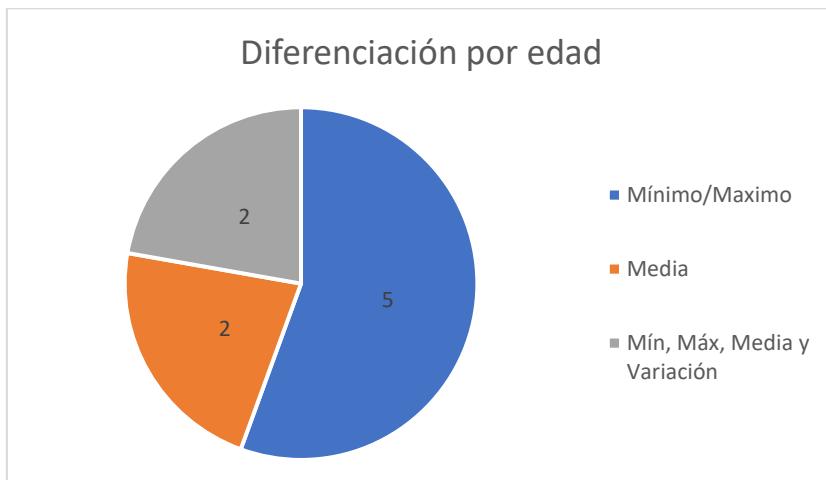


Ilustración 7 Formas de diferenciar la edad

De entre los datos analizados, se obtiene que la edad mínima es de 10 años, la máxima de 68 años y la media, calculada haciendo la media de las medias proporcionadas, es de 23 años.

Ante la existencia de menores de edad, hay que comentar que las personas menores de edad aparecen en dos investigaciones, para el resto de investigaciones la edad mínima de los participantes es de 18 años, aunque varía por cada estudio. En los artículos que mencionan la media no se puede saber la

edad mínima, aunque las medias indican que los participantes son mayores de edad.

- Género

En este apartado se muestran las estadísticas sobre la diferenciación entre género en los corpus de datos. Los artículos que mencionan la diferencia de género, lo hacen con géneros binarios distinguiendo entre hombre y mujer.

En un primer análisis se muestra el porcentaje de artículos que mencionan la diferencia de género.

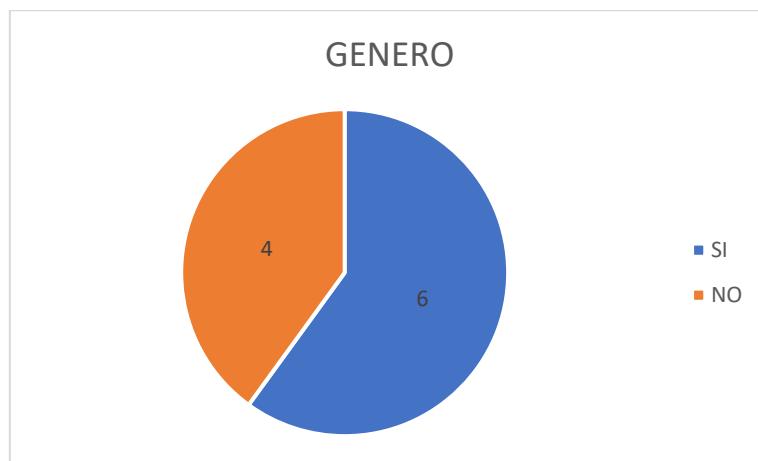


Ilustración 8 Ratio diferenciación por género

Ahora se realiza un análisis de los artículos que sí realizan diferencia por género, mostrando la ratio hombre/mujer global y por cada artículo.

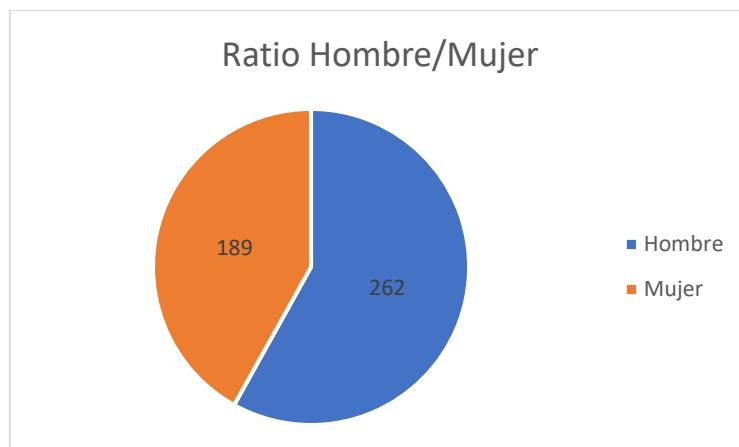


Ilustración 9 Ratio Hombre/Mujer

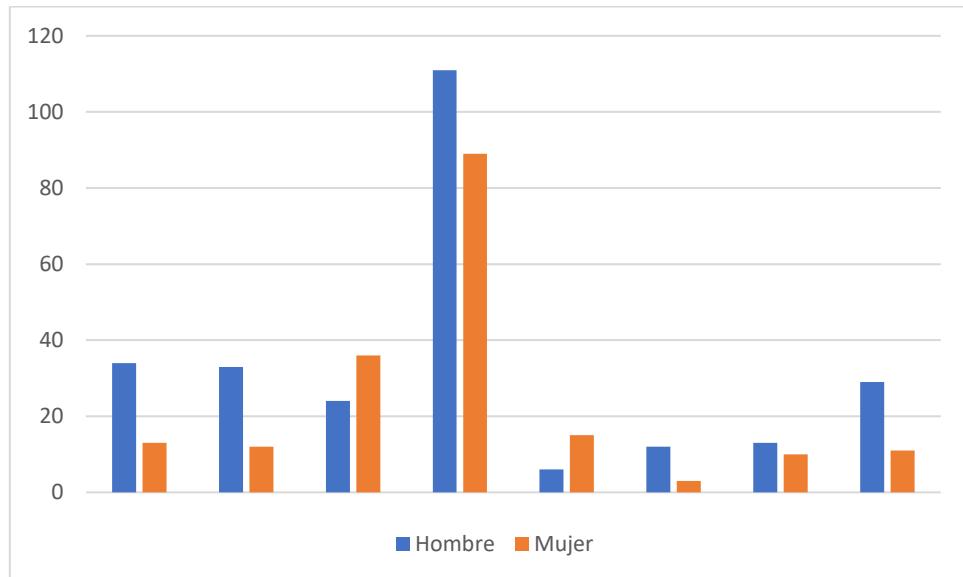


Ilustración 10 Ratio Hombre/Mujer por artículo

- Procedencia/Nacionalidad

En este apartado se muestran las estadísticas sobre la diferenciación entre procedencias en los corpus de datos. La ratio entre las investigaciones que mencionan o no la diferenciación por procedencia o nacionalidad se muestra en el siguiente gráfico.

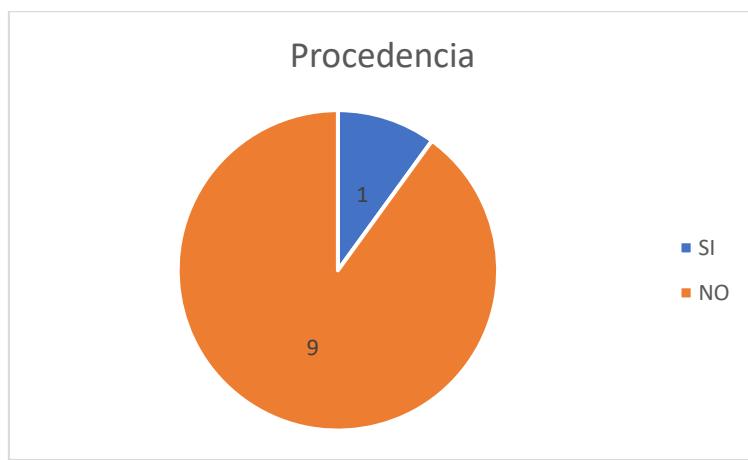


Ilustración 11 Ratio diferenciación por procedencia/nacionalidad

El único artículo que menciona la diferenciación entre nacionalidades lo hace entre las nacionalidades china, india y malaya.

- Otros

Esta sección se incluye para los casos en que los artículos hacen alguna otra distinción entre los participantes en la investigación que integran el corpus de datos.

Se realiza un doble análisis, primero la ratio entre los artículos que hacen o no alguna distinción y después se muestran las distintas diferenciaciones que se han encontrado.

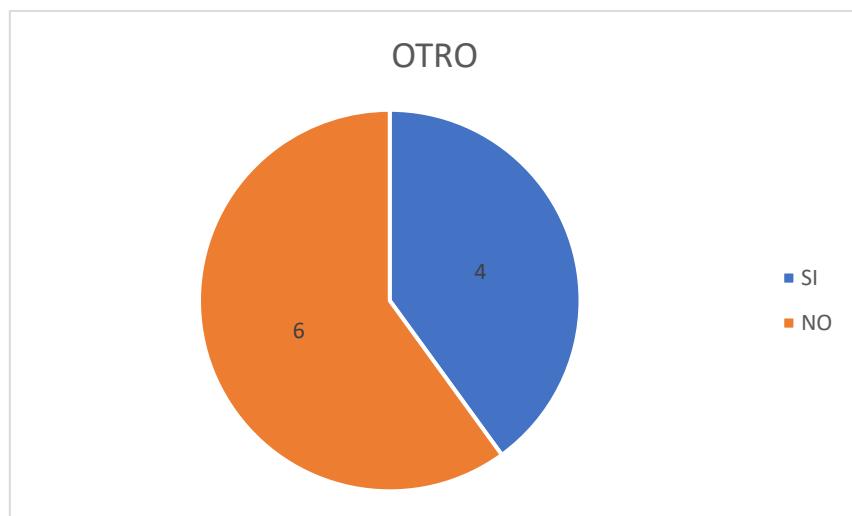


Ilustración 12 Ratio otras diferenciaciones

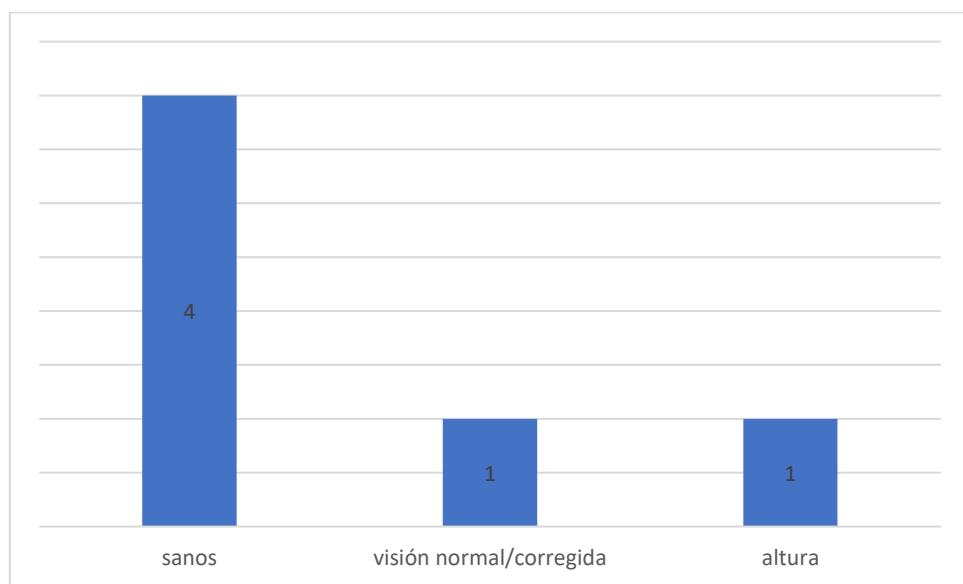


Ilustración 13 Otras diferenciaciones encontradas

Esta otra distinción se ha tomado a grandes rasgos, ya que los artículos mencionan que los participantes están sanos, que se ha anotado su altura o que los participantes tienen tanto visión normal como corregida, pero sin proporcionar más detalles.

- Publicación del corpus de datos:

Ninguno de los artículos hace mención a la publicación del corpus de dato creado para que sea accesible de forma pública.

6.4.1. Análisis de los resultados obtenidos

Tras mostrar los resultados recogidos de los diferentes artículos analizados, se puede llegar a las siguientes conclusiones:

- Hay gran variedad de datos biométricos que se pueden usar para identificar personas, esto no es una sorpresa, pero lo que puede llegar a sorprender es la precisión que pueden tener algunos datos biométricos que podrías pensar que no son tan relevantes.
- La realización de investigaciones y publicación de artículos es una tarea compleja por lo que lo habitual es colaborar entre varios investigadores, aunque parece que esa colaboración se da más entre miembros de la misma institución que formar equipo con miembros de otras instituciones.
- La calidad de la investigación influye más que el tiempo a la hora de que los artículos sean citados. Inicialmente uno podría pensar que cuanto más antiguo sea un artículo más gente lo habrá citado, pero no es el caso, ya que hay artículos recientes con más citaciones que artículos más antiguos, lo que indica que prima más la calidad que el tiempo a la hora de ser citado.
- Realizar un corpus de datos es un trabajo complicado ya que requiere juntar un grupo de participantes voluntarios, por lo que normalmente este grupo de participantes no es muy amplio y una media de alrededor de 50 participantes puede ser una cifra suficiente para el desarrollo de una investigación.

- A la hora de completar los corpus de datos con metadatos para la diferenciación de los participantes, el metadato más usado es la edad seguido por la diferenciación de género entre hombre y mujer, quedando otras posibles diferenciaciones como algo residual dependiendo del tipo de experimentos realizados durante la investigación. Cabe destacar que no hay una forma estándar en la que se facilitan los datos de la edad, ya que se han encontrado varias formas en las que se expresa ese metadato.
- Un dato sorprendente es que ningún artículo menciona que se hayan publicado los datos y que estén accesibles, lo que choca contra los objetivos de ciencia abierta que se están intentado implantar.
- En lo referente al marco legal, ningún artículo hace referencia a las leyes de privacidad o protección de datos vigentes durante la realización de la investigación, al ser un imperativo legal seguro que se han cumplido se podría considerar una buena práctica nombrarlas para que el lector tenga más confianza en la transparencia y legalidad de la investigación.
- Dentro del marco ético, lo que más importancia tiene es la financiación de la investigación por que la mayoría de los artículos la mencionan, en este punto habría que ver si por voluntad de los autores o por requisito de las editoriales. El siguiente punto importante es la declaración de que no hay conflicto de intereses por parte de los autores, quedando como temas de poca importancia la supervisión por parte de un comité de ética o el consentimiento informado que tienen que firmar los participantes. Esto puede ser debido a las diferentes normativas que han existido, donde unas veces no era necesario y otras por ser un requisito legal no se menciona al darse por sobreentendido.

7. Conclusiones y líneas de trabajo futuras

En este capítulo se van a exponer las conclusiones a las que han llegado tras todo el trabajo de análisis realizado y en qué partes se debería seguir trabajando para complementar y mantener actualizado el trabajo realizado.

7.1. Conclusiones

Se van a presentar las conclusiones de este trabajo agrupadas con los objetivos que se habían fijado junto con una conclusión inicial y conclusión general final:

- Conclusión inicial:

La realización de un trabajo de revisión sistemática de la literatura es un trabajo duro y poco grato ya que hay que tener muy claro desde el principio la metodología a seguir y ser muy meticuloso a la hora de realizar búsquedas y anotar los resultados. Además de esa parte metodológica está la parte de análisis de los artículos procedentes de las búsquedas, que muchas veces es muy desalentador ya que hay que revisar gran cantidad de trabajos para obtener los resultados deseados. En el caso particular de este trabajo, los resultados obtenidos no han sido todo lo buenos que se hubieran deseado, ya que estaba previsto obtener un número mayor de resultados de las búsquedas separadas en zonas geográficas y temporalmente, pero solo se han obtenido 9 resultados válidos, debido a ello no se han podido realizar determinados análisis para ver la evolución a lo largo del tiempo antes y después de las leyes de protección de datos y la diferencia entre zonas geográficas. En los artículos analizados había referencias a otros artículos que no se han analizado por estar fuera de la metodología utilizada, esto puede proporcionar una línea futura de trabajo, aunque no es la línea más aconsejable por no seguir una metodología clara y estructurada.

- Estado del arte:

Hay una literatura muy extensa relacionada con datos biométricos, pero al ser un campo muy extenso y abarcar muchos trabajos, los que artículos que resultan útiles para este trabajo, son una parte mínima y difícil de encontrar. Las conclusiones a las que se han llegado es que la creación de un corpus de datos y su posterior uso es muy laborioso y no es un trabajo apto para un único investigador, ya que todos los artículos están escritos por varios autores, siendo la media de 5 autores por artículo, lo que indica que la colaboración entre investigadores es lo habitual y que la colaboración más habitual es entre investigadores de la misma institución, que pueden ser del mismo o de diferentes departamentos.

Las palabras clave son algo importante para la identificación de los artículos y los autores suelen usar una media de 4 palabras clave. Muchas veces estas palabras claves no son únicas ya que además de las palabras clave de los autores, las editoriales a veces incluyen más palabras clave al artículo.

- Marco legal:

La preocupación por la privacidad de los datos y la intimidad de las personas es un objetivo legal desde hace muchos años tanto en el ámbito nacional como internacional. Este objetivo legal ha ido evolucionando con el paso del tiempo y la tecnología ya que cada vez hay mayor cantidad y variedad de datos cuyo mal uso influye de forma negativa en la calidad de vida de las personas. Esta evolución se ver reflejada en la creación de leyes concretas y específicas que regulan el uso de datos privados, estableciendo una serie de derechos sobre los datos y unas consecuencias en el caso del incumplimiento de dichas leyes.

Este marco legal es bien recibido por las personas que cada vez tienen mayor conciencia de la importancia de sus datos y la necesidad de que estén protegidos.

La legislación actual vigente en el ámbito de las investigaciones científicas en las universidades españolas es el Reglamento General de Protección de Datos (RGPD) y la Ley Orgánica 2/2023, de 22 de marzo, del Sistema Universitario (LOSU) que obligan a los investigadores de las universidades a cumplir con la protección de datos y a fomentar la ciencia abierta y los principios FAIR, por lo que las

investigaciones deberían tener un carácter abierto y publicarse en repositorios instruccionales de libre acceso.

- Marco ético

Las investigaciones científicas tienen que estar validadas por un comité de ética que certifique que dicha investigación no atenta contra los derechos y valores de las personas. Además, dentro los aspectos éticos de la investigación se encuentran que no haya conflictos de intereses de los autores a la hora de hacer la investigación o que la investigación no contenga ningún sesgo creado voluntariamente, por ejemplo, para favorecer a las entidades que han financiado la investigación, debido a esto, es importante que las investigaciones contengan una declaración de conflicto de intereses y publiquen de donde vienen la financiación de la investigación.

- Repositorios de datos:

En la actualidad hay muchos repositorios de datos en línea disponibles donde muchos de ellos tienen características similares, por ejemplo, la compatibilidad con DOI, ORCID o funcionalidades similares. Las grandes diferencias se encuentran en el tamaño de los datos que admiten, el precio y la posibilidad de publicar los corpus de datos de forma pública.

Algo en lo que todos los repositorios coinciden es en la importancia de los principios FAIR y por consiguiente de los metadatos. Dichos metadatos son de vital importancia para que un corpus de datos sea de calidad y además pueda ser encontrado para que otros investigadores lo puedan utilizar y referenciar cobrando el trabajo de los creadores del corpus de datos más importancia.

- Análisis de la literatura de creación de corpus de datos

De los artículos analizados podemos obtener dos grupos de conclusiones, la primera relacionada con los participantes donde se ha obtenido que la media de participantes es de 46, que se recogen una media de 61 muestras por participante

en una única sesión y que los metadatos más usados para la diferenciación de participantes son la edad y género, aunque pueden existir otra en función del tipo de datos como pueden ser la altura, la procedencia, si tienen o no corrección visual o si son participantes sanos. Aunque no existe una forma estándar o consensuada de mostrar esa información.

El segundo grupo de conclusiones son las relacionadas con la información más general que podríamos considerar como información de transparencia. Estas conclusiones son que el dato de la procedencia de los fondos de financiación es importante ya que aparece en la mayoría de los artículos, pero sin embargo otros datos como informar de la ley de protección de datos vigente en el momento de realizar la investigación o si se ha utilizado alguna norma o protocolo no se consideran tan importantes ya que no aparecen en ningún artículo. Otra información que tampoco aparece en muchos artículos es la información del comité de ética que ha validado la investigación o si los participantes han firmado un consentimiento informado. Otro dato a comentar es que no hay un estándar a la hora de informar sobre los medios o dispositivos de captura, ya que en algunos artículos se explica de forma muy breve y en otros se explica de forma más extensa y detallada.

Se considera que para dotar de transparencia y permitir la replicabilidad de la investigación todos estos apartados se deberían de mencionar en los artículos, además de facilitar el acceso al corpus de datos mediante una referencia al corpus alojado en un servidor de datos.

Como resumen de las conclusiones, se puede afirmar que lo más importante para publicar un corpus de datos desde una investigación de la universidad, es que tiene que cumplir los siguientes puntos:

- Tener la aprobación de un comité de ética.
- Los participantes tienen que firmar un consentimiento informado que contenga la finalidad de la investigación y cuánto tiempo se van a almacenar los datos.
- Los datos se tienen que publicar anonimizados o seudonimizados.

- Tiene que estar publicado de forma abierta y accesible en el repositorio de la universidad.
- Los datos tienen que ir acompañados de metadatos.
- Un buen corpus de datos tiene que contar con una media de 46 participantes, una media de 61 muestras por sesión y una sesión. Estos datos los tomamos como medias de los datos analizados, pero no son obligatorios, y sería recomendable realizar 2 o más sesiones separadas en el tiempo para ver la evolución con el tiempo. De esta forma podríamos tomar como recomendable 50 participantes, 2 sesiones separadas en el tiempo, que no sean el mismo día y 30 muestras por sesión obteniendo un resultado de 3000 muestras.
- La diferenciación que se tendría que hacer de los participantes sería como mínimo por edad y por género, pero en función de los datos y de la calidad que queramos en los datos, se pueden añadir más diferenciaciones para que los metadatos sean más ricos.
- La publicación debería tener secciones dedicadas a los conflictos de intereses, la procedencia de la financiación y la disponibilidad y referencia a donde está publicado el corpus de datos.

7.2. Líneas de trabajo futuras

Este trabajo teórico tiene varias líneas de trabajo futuras posibles, entre las que se destacan las tres siguientes:

- Proseguir con la búsqueda de artículos usando la misma metodología empleada hasta conseguir el objetivo de artículos fijado inicialmente y poder comparar entre zonas geográficas y a lo largo del tiempo teniendo en cuenta el antes y el después de la entrada en vigor de las leyes de protección de datos.
- Mantener el trabajo actualizado según vayan analizándose nuevos estudios publicados, se creen nuevos repositorios de datos o cambien sus

condiciones de uso y se actualicen las leyes referentes a la protección de datos.

- Realizar un análisis de los corpus de datos biométricos de acceso público más utilizados de los que se han encontrado en los artículos analizados y descartados por no crear su propio corpus de datos.
- Realizar un trabajo práctico consistente en crear un corpus de datos que cumplan con las conclusiones a las que se han llegado con la realización de este proyecto.

Bibliografía

- [1] A. Jain, L. Hong, y S. Pankanti, Biometric identification, *Commun. ACM*, vol. 43, n.º 2, pp. 90-98, feb. 2000, doi: 10.1145/328236.328110.
- [2] Ética en la investigación | Consejo Superior de Investigaciones Científicas. Accedido: 30 de julio de 2024. [En línea]. Disponible en: <https://www.csic.es/es/el-csic/etica/etica-en-la-investigacion>
- [3] Repositorio (contenido digital), *Wikipedia, la enciclopedia libre*. 4 de julio de 2024. Accedido: 30 de julio de 2024. [En línea]. Disponible en: [https://es.wikipedia.org/w/index.php?title=Repositorio_\(contenido_digital\)&oldid=161119535](https://es.wikipedia.org/w/index.php?title=Repositorio_(contenido_digital)&oldid=161119535)
- [4] C. Ortega, Repositorio de datos: Qué es, ventajas y tipos, QuestionPro. Accedido: 26 de mayo de 2024. [En línea]. Disponible en: <https://www.questionpro.com/blog/es/repositorio-de-datos/>
- [5] Ciencia abierta, *Wikipedia, la enciclopedia libre*. 27 de enero de 2024. Accedido: 7 de junio de 2024. [En línea]. Disponible en: https://es.wikipedia.org/w/index.php?title=Ciencia_abierta&oldid=157778353
- [6] F. Rodriguez-Sánchez, A. J. Pérez-Luque, I. Bartomeus, y S. Varela, Ciencia reproducible: qué, por qué, cómo:, *Ecosistemas*, vol. 25, n.º 2, Art. n.º 2, jul. 2016, doi: 10.7818/ECOS.2016.25-2.11.
- [7] BOE.es - DOUE-L-1998-81549 Directiva 98/48/CE del Parlamento Europeo y del Consejo, de 20 de julio de 1998, que modifica la Directiva 98/34/CE por la que se establece un procedimiento de información en materia de las normas y reglamentaciones técnicas. Accedido: 17 de febrero de 2025. [En línea]. Disponible en: <https://www.boe.es/buscar/doc.php?id=DOUE-L-1998-81549>
- [8] BOE.es - DOUE-L-2016-80807 Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos). Accedido: 10 de junio de 2024. [En línea]. Disponible en: <https://www.boe.es/buscar/doc.php?id=DOUE-L-2016-80807>
- [9] Zotero | Your personal research assistant. Accedido: 17 de julio de 2025. [En línea]. Disponible en: <https://www.zotero.org/>

- [10] Search - Consensus: AI Search Engine for Research. Accedido: 30 de marzo de 2025. [En línea]. Disponible en: <https://consensus.app/search/>
- [11] Google Académico. Accedido: 30 de marzo de 2025. [En línea]. Disponible en: <https://scholar.google.com/>
- [12] ScienceDirect.com | Science, health and medical journals, full text articles and books. Accedido: 6 de junio de 2024. [En línea]. Disponible en: <https://www.sciencedirect.com/>
- [13] MDPI - Publisher of Open Access Journals, MDPI. Accedido: 7 de junio de 2024. [En línea]. Disponible en: <https://www.mdpi.com/>
- [14] IEEE Xplore. Accedido: 10 de junio de 2024. [En línea]. Disponible en: <https://ieeexplore.ieee.org/Xplore/home.jsp>
- [15] PRISMA statement, PRISMA statement. Accedido: 24 de mayo de 2025. [En línea]. Disponible en: <https://www.prisma-statement.org>
- [16] IEEE Code of Ethics. Accedido: 30 de julio de 2024. [En línea]. Disponible en: <https://www.ieee.org/about/corporate/governance/p7-8.html>
- [17] Normativa Comunicaciones. Accedido: 30 de julio de 2024. [En línea]. Disponible en: <https://congresocomunicacionysalud.semfyc.es/normativa-comunicaciones>
- [18] Declaración de Helsinki de la AMM - Principios éticos para las investigaciones médicas en seres humanos. Accedido: 26 de marzo de 2025. [En línea]. Disponible en: <https://web.archive.org/web/20170308101607/http://www.wma.net/es/30publications/10policies/b3/>
- [19] Comité Ético de Investigación – UVainvestiga – Investigación en la Universidad de Valladolid. Accedido: 31 de julio de 2024. [En línea]. Disponible en: <https://investiga.uva.es/buenas-practicas/comites-de-etica/comite-etico-de-investigacion/>
- [20] Código de Buenas Prácticas en Investigación (CBPI) – UVainvestiga – Investigación en la Universidad de Valladolid. Accedido: 31 de julio de 2024. [En línea]. Disponible en: <https://investiga.uva.es/buenas-practicas/politicas-de-investigacion/codigo-de-buenas-practicas-en-investigacion/>
- [21] Comité de ética CEISH – UVainvestiga – Investigación en la Universidad de Valladolid. Accedido: 11 de mayo de 2025. [En línea]. Disponible en: <https://investiga.uva.es/buenas-practicas/comites-de-etica/comite-de-etica-ceish/>
- [22] FORMULARIO-de-Evaluacion-para-Proyectos-de-Investigacion_CEISH_UVa.docx. Accedido: 11 de mayo de 2025. [En línea].

- Disponible en:
https://view.officeapps.live.com/op/view.aspx?src=https%3A%2F%2Finvestiga.uva.es%2Ffiles%2F2025%2F03%2FFORMULARIO-de-Evaluacion-para-Proyectos-de-Investigacion_CEISH_UVa.docx&wdOrigin=BROWSELINK
- [23] Protección de datos | Secretaría General. Accedido: 11 de mayo de 2025. [En línea]. Disponible en: <https://secretariageneral.uva.es/competencias/proteccion-de-datos/>
- [24] Jefatura del Estado, *Instrumento de Ratificación del Convenio para la Protección de los Derechos Humanos y de las Libertades Fundamentales, hecho en Roma el 4 de noviembre de 1950, y enmendado por los Protocolos adicionales números 3 y 5, de 6 de mayo de 1963 y 20 de enero de 1966, respectivamente*, vol. BOE-A-1979-24010. 1979, pp. 23564-23570. Accedido: 12 de febrero de 2025. [En línea]. Disponible en: [https://www.boe.es/eli/es/ai/1950/11/04/\(1\)](https://www.boe.es/eli/es/ai/1950/11/04/(1))
- [25] BOE-A-1978-31229 Constitución Española. Accedido: 14 de enero de 2025. [En línea]. Disponible en: <https://www.boe.es/buscar/act.php?id=BOE-A-1978-31229>
- [26] Jefatura del Estado, *Convenio para la protección de las personas con respecto al tratamiento automatizado de datos de carácter personal, hecho en Estrasburgo el 28 de enero de 1981*, vol. BOE-A-1985-23447. 1985, pp. 36000-36004. Accedido: 14 de enero de 2025. [En línea]. Disponible en: [https://www.boe.es/eli/es/ai/1981/01/28/\(1\)](https://www.boe.es/eli/es/ai/1981/01/28/(1))
- [27] La protección de los datos personales | Fichas temáticas sobre la Unión Europea | Parlamento Europeo. Accedido: 14 de enero de 2025. [En línea]. Disponible en: <https://www.europarl.europa.eu/factsheets/es/sheet/157/la-proteccion-de-los-datos-personales>
- [28] Jefatura del Estado, *Ley Orgánica 5/1992, de 29 de octubre, de regulación del tratamiento automatizado de los datos de carácter personal*, vol. BOE-A-1992-24189. 1992, pp. 37037-37045. Accedido: 14 de enero de 2025. [En línea]. Disponible en: <https://www.boe.es/eli/es/lo/1992/10/29/5>
- [29] Ministerio de Justicia e Interior, *Real Decreto 1332/1994, de 20 de junio, por el que se desarrolla determinados aspectos de la Ley Orgánica 5/1992, de 29 de octubre, de regulación del tratamiento automatizado de los datos de carácter personal*, vol. BOE-A-1994-14121. 1994, pp. 19199-19203. Accedido: 17 de febrero de 2025. [En línea]. Disponible en: <https://www.boe.es/eli/es/rd/1994/06/20/1332>

- [30] BOE.es - DOUE-L-1995-81678 Directiva 95/46/CE del Parlamento Europeo y del Consejo, de 24 de octubre de 1995, relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos. Accedido: 12 de febrero de 2025. [En línea]. Disponible en: <https://www.boe.es/buscar/doc.php?id=DOUE-L-1995-81678>
- [31] Jefatura del Estado, *Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal*, vol. BOE-A-1999-23750. 1999, pp. 43088-43099. Accedido: 14 de enero de 2025. [En línea]. Disponible en: <https://www.boe.es/eli/es/lo/1999/12/13/15>
- [32] Ministerio de Justicia, *Real Decreto 994/1999, de 11 de junio, por el que se aprueba el Reglamento de medidas de seguridad de los ficheros automatizados que contengan datos de carácter personal*, vol. BOE-A-1999-13967. 1999, pp. 24241-24245. Accedido: 17 de febrero de 2025. [En línea]. Disponible en: <https://www.boe.es/eli/es/rd/1999/06/11/994>
- [33] BOE.es - DOUE-L-2000-81295 Directiva 2000/31/CE del Parlamento Europeo y del Consejo, de 8 de junio de 2000, relativa a determinados aspectos jurídicos de los servicios de la sociedad de la información, en particular el comercio electrónico en el mercado interior (Directiva sobre el comercio electrónico). Accedido: 17 de febrero de 2025. [En línea]. Disponible en: <https://www.boe.es/buscar/doc.php?id=DOUE-L-2000-81295>
- [34] BOE-A-2002-13758 Ley 34/2002, de 11 de julio, de servicios de la sociedad de la información y de comercio electrónico. Accedido: 17 de febrero de 2025. [En línea]. Disponible en: <https://www.boe.es/buscar/act.php?id=BOE-A-2002-13758>
- [35] BOE-A-2002-22188 Ley 41/2002, de 14 de noviembre, básica reguladora de la autonomía del paciente y de derechos y obligaciones en materia de información y documentación clínica. Accedido: 17 de febrero de 2025. [En línea]. Disponible en: <https://www.boe.es/buscar/act.php?id=BOE-A-2002-22188>
- [36] BOE-A-2011-9617 Ley 14/2011, de 1 de junio, de la Ciencia, la Tecnología y la Innovación. Accedido: 13 de mayo de 2025. [En línea]. Disponible en: <https://www.boe.es/buscar/act.php?id=BOE-A-2011-9617>
- [37] BOE.es - DOUE-L-2016-80808 Directiva (UE) 2016/680 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales por parte de las autoridades competentes para fines de prevención, investigación, detección o enjuiciamiento de infracciones penales o de ejecución de

- sanciones penales, y a la libre circulación de dichos datos y por la que se deroga la Decisión Marco 2008/977/JAI del Consejo. Accedido: 18 de febrero de 2025. [En línea]. Disponible en: <https://www.boe.es/buscar/doc.php?id=DOUE-L-2016-80808>
- [38] BOE-A-2018-16673 Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales. Accedido: 14 de enero de 2025. [En línea]. Disponible en: <https://www.boe.es/buscar/act.php?id=BOE-A-2018-16673>
- [39] BOE.es - DOUE-L-2022-80835 Reglamento (UE) 2022/868 del Parlamento Europeo y del Consejo de 30 de mayo de 2022 relativo a la gobernanza europea de datos y por el que se modifica el Reglamento (UE) 2018/1724 (Reglamento de Gobernanza de Datos). Accedido: 5 de mayo de 2025. [En línea]. Disponible en: <https://www.boe.es/buscar/doc.php?id=DOUE-L-2022-80835>
- [40] BOE-A-2023-7500 Ley Orgánica 2/2023, de 22 de marzo, del Sistema Universitario. Accedido: 10 de mayo de 2025. [En línea]. Disponible en: <https://www.boe.es/buscar/act.php?id=BOE-A-2023-7500>
- [41] Jefatura del Estado, *Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales*, vol. BOE-A-2018-16673. 2018, pp. 119788-119857. Accedido: 10 de junio de 2024. [En línea]. Disponible en: <https://www.boe.es/eli/es/lo/2018/12/05/3>
- [42] guia-control-presencia-biometrico.pdf. Accedido: 11 de junio de 2024. [En línea]. Disponible en: <https://www.aepd.es/guias/guia-control-presencia-biometrico.pdf>
- [43] Agencia Española de Protección de Datos | AEPD. Accedido: 11 de junio de 2024. [En línea]. Disponible en: <https://www.aepd.es/>
- [44] U. Nations, La Declaración Universal de los Derechos Humanos | Naciones Unidas, United Nations. Accedido: 25 de marzo de 2025. [En línea]. Disponible en: <https://www.un.org/es/about-us/universal-declaration-of-human-rights>
- [45] International Covenant on Civil and Political Rights, OHCHR. Accedido: 25 de marzo de 2025. [En línea]. Disponible en: <https://www.ohchr.org/en/instruments-mechanisms/instruments/international-covenant-civil-and-political-rights>
- [46] Pacto Internacional de Derechos Económicos, Sociales y Culturales, OHCHR. Accedido: 25 de marzo de 2025. [En línea]. Disponible en: <https://www.ohchr.org/es/instruments-mechanisms/instruments/international-covenant-civil-and-political-rights>

mechanisms/instruments/international-covenant-economic-social-and-cultural-rights

- [47] California Consumer Privacy Act (CCPA), State of California - Department of Justice - Office of the Attorney General. Accedido: 23 de marzo de 2025. [En línea]. Disponible en: <https://oag.ca.gov/privacy/ccpa>
- [48] W. G. Voss, The CCPA and the GDPR Are Not the Same: Why You Should Understand Both, 18 de enero de 2021, *Social Science Research Network, Rochester, NY*. 3769825. Accedido: 23 de marzo de 2025. [En línea]. Disponible en: <https://papers.ssrn.com/abstract=3769825>
- [49] Personal Information Protection Law of the People's Republic of China. Accedido: 23 de marzo de 2025. [En línea]. Disponible en: http://en.npc.gov.cn.cdurl.cn/2021-12/29/c_694559.htm
- [50] W. Li y J. Chen, From brussels effect to gravity assists: Understanding the evolution of the GDPR-inspired personal information protection law in China, *Computer Law & Security Review*, vol. 54, p. 105994, sep. 2024, doi: 10.1016/j.clsr.2024.105994.
- [51] The Personal Information Protection Law: China's Version of the GDPR?, *Columbia Journal of Transnational Law*. Accedido: 24 de marzo de 2025. [En línea]. Disponible en: <https://www.jtl.columbia.edu/bulletin-blog/the-personal-information-protection-law-chinas-version-of-the-gdpr>
- [52] Act on the Protection of Personal Information - English - Japanese Law Translation. Accedido: 24 de marzo de 2025. [En línea]. Disponible en: <https://www.japaneselawtranslation.go.jp/en/laws/view/4241/en>
- [53] Lei Geral de Proteção de Dados Pessoais (LGPD). Accedido: 25 de marzo de 2025. [En línea]. Disponible en: https://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2018/Lei/L13709.htm
- [54] 2bf1f0e9f04e6fb4f8fef35e82c42aa5.pdf. Accedido: 25 de marzo de 2025. [En línea]. Disponible en: <https://www.meity.gov.in/static/uploads/2024/06/2bf1f0e9f04e6fb4f8fef35e82c42aa5.pdf>
- [55] 개인정보 보호법. Accedido: 26 de marzo de 2025. [En línea]. Disponible en: https://elaw.klri.re.kr/kor_service/lawView.do?hseq=53044&lang=ENG
- [56] V. K. S. Anil y A. Babatope, Data privacy, security, and governance: A global comparative analysis of regulatory compliance and technological innovation, *Global Journal of Engineering and Technology Advances*, vol. 21, n.º 3, pp. 190-202, 2024, doi: 10.30574/gjeta.2024.21.3.0246.

- [57] S. Lim y J. Oh, Navigating Privacy: A Global Comparative Analysis of Data Protection Laws, *IET Information Security*, vol. 2025, n.º 1, p. 5536763, 2025, doi: 10.1049/ise2/5536763.
- [58] ISO/IEC 24745:2022, ISO. Accedido: 30 de marzo de 2025. [En línea]. Disponible en: <https://www.iso.org/es/contents/data/standard/07/53/75302.html>
- [59] UNE-EN ISO/IEC 27001:2023 | Normas AENOR. Accedido: 30 de marzo de 2025. [En línea]. Disponible en: <https://tienda.aenor.com/norma-une-en-iso-iec-27001-2023-n0071764>
- [60] ISO/IEC 29794-1:2024, ISO. Accedido: 19 de abril de 2025. [En línea]. Disponible en: <https://www.iso.org/standard/79519.html>
- [61] ISO/IEC JTC 1 - Information technology, ISO. Accedido: 19 de abril de 2025. [En línea]. Disponible en: <https://www.iso.org/committee/45020.html>
- [62] Red de Bibliotecas Universitarias y Científicas Españolas. Accedido: 12 de mayo de 2025. [En línea]. Disponible en: <https://www.rebiun.org/>
- [63] Inicio, CRUE. Accedido: 12 de mayo de 2025. [En línea]. Disponible en: <https://www.crue.org/>
- [64] Consejo Superior de Investigaciones Científicas. Accedido: 12 de mayo de 2025. [En línea]. Disponible en: <https://www.csic.es/es>
- [65] Welcome :: Greenstone Digital Library Software. Accedido: 13 de mayo de 2025. [En línea]. Disponible en: <https://www.greenstone.org/>
- [66] P. Vierkant, DataCite – Connecting Research, Advancing Knowledge, DataCite. Accedido: 7 de junio de 2024. [En línea]. Disponible en: <https://datacite.org/>
- [67] Home Page. Accedido: 7 de junio de 2024. [En línea]. Disponible en: <https://www.doi.org/>
- [68] El portal oficial de datos europeos | data.europa.eu. Accedido: 7 de junio de 2024. [En línea]. Disponible en: <https://data.europa.eu/es>
- [69] FAIRsharing | Home. Accedido: 7 de junio de 2024. [En línea]. Disponible en: <https://fairsharing.org/>
- [70] Dataset Search. Accedido: 7 de junio de 2024. [En línea]. Disponible en: <https://datasetsearch.research.google.com/>
- [71] Hugging Face – The AI community building the future. Accedido: 7 de junio de 2024. [En línea]. Disponible en: <https://huggingface.co/>
- [72] Kaggle: Your Machine Learning and Data Science Community. Accedido: 7 de junio de 2024. [En línea]. Disponible en: <https://www.kaggle.com/>

- [73] Welcome to OpenDOAR - Sherpa Services. Accedido: 7 de junio de 2024. [En línea]. Disponible en: <https://v2.sherpa.ac.uk/opendoar/>
- [74] Papers with Code - The latest in Machine Learning. Accedido: 7 de junio de 2024. [En línea]. Disponible en: <https://paperswithcode.com/>
- [75] GFZ German Research Centre For Geosciences *et al.*, Registry of Research Data Repositories. Accedido: 7 de junio de 2024. [En línea]. Disponible en: <https://www.re3data.org/>
- [76] Zenodo. Accedido: 7 de junio de 2024. [En línea]. Disponible en: <https://zenodo.org/>
- [77] Open science | UNESCO. Accedido: 7 de junio de 2024. [En línea]. Disponible en: <https://www.unesco.org/en/open-science>
- [78] Estrategia Nacional de Ciencia Abierta. Accedido: 7 de junio de 2024. [En línea]. Disponible en: <https://www.ciencia.gob.es/Estrategias-y-Planes/Estrategias/ENCA.html>
- [79] GitHub · Build and ship software on a single, collaborative platform, GitHub. Accedido: 5 de mayo de 2025. [En línea]. Disponible en: <https://github.com/>
- [80] Mendeley Data. [En línea]. Disponible en: <https://data.mendeley.com/>
- [81] National Center for Biotechnology Information. Accedido: 5 de mayo de 2025. [En línea]. Disponible en: <https://www.ncbi.nlm.nih.gov/>
- [82] Vivli - Center for Global Clinical Research Data, Vivli. Accedido: 5 de mayo de 2025. [En línea]. Disponible en: <https://vivli.org/>
- [83] Welcome to IEEE DataPort | IEEE DataPort. Accedido: 5 de mayo de 2025. [En línea]. Disponible en: <https://ieee-dataport.org/>
- [84] figshare - credit for all your research. Accedido: 5 de mayo de 2025. [En línea]. Disponible en: <https://figshare.com/>
- [85] Dataverse : Instalación y Soporte | Repositorio de datos abiertos para la ciencia y la investigación, <https://opengeekservice.cl/>. Accedido: 5 de mayo de 2025. [En línea]. Disponible en: <https://opengeekservice.cl/dataverse/>
- [86] Dryad | Publish and preserve your data. Accedido: 5 de mayo de 2025. [En línea]. Disponible en: <https://datadryad.org/>
- [87] OSF. Accedido: 5 de mayo de 2025. [En línea]. Disponible en: <https://osf.io/>
- [88] 4TU.ResearchData. Accedido: 5 de mayo de 2025. [En línea]. Disponible en: <https://data.4tu.nl/>
- [89] UVaDOC Principal. Accedido: 5 de mayo de 2025. [En línea]. Disponible en: <https://uvadoc.uva.es/>
- [90] Datasets. Accedido: 5 de mayo de 2025. [En línea]. Disponible en: <https://uvadoc.uva.es/handle/10324/39911>

- [91] Datos abiertos del Gobierno de España – Datos.gob.es. Accedido: 7 de junio de 2024. [En línea]. Disponible en: <https://datos.gob.es/es/>
- [92] Inicio — Análisis de datos abiertos JCyL. Accedido: 7 de junio de 2024. [En línea]. Disponible en: <https://analisis.datosabiertos.jcyl.es/pages/home/?flg=es-es>
- [93] Datos de Investigación. Accedido: 7 de junio de 2024. [En línea]. Disponible en: <https://uvadoc.uva.es/handle/10324/39910>
- [94] Master Data Repository List, Web of Science Group. Accedido: 7 de junio de 2024. [En línea]. Disponible en: <https://clarivate.com/webofsciencegroup/master-data-repository-list/>
- [95] Data repositories - Open Access Directory. Accedido: 7 de junio de 2024. [En línea]. Disponible en: https://oad.simmons.edu/oadwiki/Data_repositories
- [96] Home, DCMI. Accedido: 21 de mayo de 2025. [En línea]. Disponible en: <https://www.dublincore.org/>
- [97] Catálogo Colectivo del Consorcio de Bibliotecas Universitarias de Castilla y León - Consorcio BUCLE. Accedido: 20 de mayo de 2025. [En línea]. Disponible en: <https://consorciobucle.es/>
- [98] M. de Cultura, Hispana. Acceso en línea al Patrimonio Cultural Digital Español. Accedido: 21 de mayo de 2025. [En línea]. Disponible en: <https://hispana.mcu.es/>
- [99] Recolecta | Recolector de Ciencia Abierta. Accedido: 21 de mayo de 2025. [En línea]. Disponible en: <https://recolecta.fecyt.es/>
- [100] Descubre el patrimonio cultural digital europeo. Accedido: 21 de mayo de 2025. [En línea]. Disponible en: <https://www.europeana.eu/es>
- [101] K. Iatropoulou, OpenAIRE, OpenAIRE. Accedido: 14 de mayo de 2025. [En línea]. Disponible en: <https://www.openaire.eu/>
- [102] E. R. O. M. A. D. Masters (EROMM), EROMM : European Register Of Microform And Digital Masters. Accedido: 21 de mayo de 2025. [En línea]. Disponible en: <https://www.eromm.org/>
- [103] Welcome to the Registry of Open Access Repositories - Registry of Open Access Repositories. Accedido: 21 de mayo de 2025. [En línea]. Disponible en: <https://roar.eprints.org/>
- [104] Bielefeld Academic Search Engine - BASE. Accedido: 21 de mayo de 2025. [En línea]. Disponible en: <http://base.ub.uni-bielefeld.de/de/index.php>
- [105] WorldCat.org. Accedido: 21 de mayo de 2025. [En línea]. Disponible en: <https://search.worldcat.org>

- [106] SWORD v2 – SWORD. Accedido: 21 de mayo de 2025. [En línea]. Disponible en: <https://swordapp.org/category/sword2/>
- [107] Formatos soportados UVaDOC. [En línea]. Disponible en: <https://uvadoc.blogs.uva.es/files/2021/09/formatos-UVaDOC.pdf>
- [108] File Formats - Mendeley Data. Accedido: 14 de mayo de 2025. [En línea]. Disponible en: <https://data.mendeley.com/file-formats>
- [109] API Docs - Mendeley Data. Accedido: 14 de mayo de 2025. [En línea]. Disponible en: <https://data.mendeley.com/api/docs/>
- [110] Generalist Repository Ecosystem Initiative | Data Science at NIH. [En línea]. Disponible en: <https://datascience.nih.gov/data-ecosystem/generalist-repository-ecosystem-initiative>
- [111] Home | CERN. Accedido: 16 de mayo de 2025. [En línea]. Disponible en: <https://home.cern/>
- [112] Digital Science: tools for open, collaborative and inclusive research, Digital Science. Accedido: 17 de mayo de 2025. [En línea]. Disponible en: <https://www.digital-science.com/>
- [113] Figshare Plus, Figshare. Accedido: 18 de mayo de 2025. [En línea]. Disponible en: <https://info.figshare.com/figshare-plus/>
- [114] File formats supported for in-browser viewing - a help article for using figshare. Accedido: 18 de mayo de 2025. [En línea]. Disponible en: <https://help.figshare.com/article/file-formats-supported-for-in-browser-viewing>
- [115] How to choose the most appropriate license - a help article for using figshare. Accedido: 18 de mayo de 2025. [En línea]. Disponible en: <https://help.figshare.com/article/how-to-choose-the-most-appropriate-license>
- [116] How to upload and publish your research - a help article for using figshare. Accedido: 18 de mayo de 2025. [En línea]. Disponible en: <https://help.figshare.com/article/how-to-upload-and-publish-your-research>
- [117] The Principles of Open Scholarly Infrastructure (v1.1, 2023), The Principles of Open Scholarly Infrastructure. Accedido: 19 de mayo de 2025. [En línea]. Disponible en: <https://openscholarlyinfrastructure.org/>
- [118] Research Organization Registry (ROR), Research Organization Registry (ROR). Accedido: 19 de mayo de 2025. [En línea]. Disponible en: <https://ror.org/>
- [119] C. for O. Science, Center for Open Science. Accedido: 19 de mayo de 2025. [En línea]. Disponible en: <https://www.cos.io>
- [120] Home, Metadata Center. Accedido: 20 de mayo de 2025. [En línea]. Disponible en: <https://metadatacenter.org/>

- [121] M. Fairhurst, C. Li, y M. Da Costa-Abreu, Predictive biometrics: a review and analysis of predicting personal characteristics from biometric data, *IET Biometrics*, vol. 6, n.º 6, pp. 369-378, 2017, doi: 10.1049/iet-bmt.2016.0169.
- [122] S. Arman, T. Yang, S. Shahed, A. Mazroa, A. Attiah, y L. Mohaisen, A Comprehensive Survey for Privacy-Preserving Biometrics: Recent Approaches, Challenges, and Future Directions, *CMC*, vol. 78, n.º 2, pp. 2087-2110, 2024, doi: 10.32604/cmc.2024.047870.
- [123] S. Dargan y M. Kumar, A comprehensive survey on the biometric recognition systems based on physiological and behavioral modalities, *Expert Systems with Applications*, vol. 143, p. 113114, abr. 2020, doi: 10.1016/j.eswa.2019.113114.
- [124] H. Mandalapu *et al.*, Audio-Visual Biometric Recognition and Presentation Attack Detection: A Comprehensive Survey, *IEEE Access*, vol. 9, pp. 37431-37455, 2021, doi: 10.1109/ACCESS.2021.3063031.
- [125] R. Alrawili, A. A. S. AlQahtani, y M. K. Khan, Comprehensive survey: Biometric user authentication application, evaluation, and discussion, *Computers and Electrical Engineering*, vol. 119, p. 109485, oct. 2024, doi: 10.1016/j.compeleceng.2024.109485.
- [126] M. Adil, A. Farouk, A. Ali, H. Song, y Z. Jin, Securing Tomorrow of Next-Generation Technologies with Biometrics, State-of-The-Art Techniques, Open Challenges, and Future Research Directions, *Computer Science Review*, vol. 57, p. 100750, ago. 2025, doi: 10.1016/j.cosrev.2025.100750.
- [127] A. Makrushin, V. S. Mannam, y J. Dittmann, Privacy-Friendly Datasets of Synthetic Fingerprints for Evaluation of Biometric Algorithms, *Applied Sciences*, vol. 13, n.º 18, Art. n.º 18, ene. 2023, doi: 10.3390/app131810000.
- [128] I. Natgunanathan, A. Mehmood, Y. Xiang, G. Beliakov, y J. Yearwood, Protection of Privacy in Biometric Data, *IEEE Access*, vol. 4, pp. 880-892, 2016, doi: 10.1109/ACCESS.2016.2535120.
- [129] C. Jasserand, Research, the GDPR, and Mega Biometric Training Datasets: Opening the Pandora Box 1, en *2022 International Conference of the Biometrics Special Interest Group (BIOSIG)*, sep. 2022, pp. 1-6. doi: 10.1109/BIOSIG55365.2022.9897040.
- [130] D. Migdal y C. Rosenberger, Statistical modeling of keystroke dynamics samples for the generation of synthetic datasets, *Future Generation Computer Systems*, vol. 100, pp. 907-920, nov. 2019, doi: 10.1016/j.future.2019.03.056.
- [131] N. Whiskerd, J. Dittmann, y C. Vielhauer, A Requirement Analysis for Privacy Preserving Biometrics in View of Universal Human Rights and Data Protection

- Regulation, en *2018 26th European Signal Processing Conference (EUSIPCO)*, sep. 2018, pp. 548-552. doi: 10.23919/EUSIPCO.2018.8553045.
- [132] J.-C. Klie, R. Eckart de Castilho, y I. Gurevych, Analyzing Dataset Annotation Quality Management in the Wild, *Computational Linguistics*, vol. 50, n.º 3, pp. 817-866, sep. 2024, doi: 10.1162/coli_a_00516.
- [133] M. Fallahi, R. Ramesh, P. P. Ramasamy, P. A. Cabarcos, T. Strufe, y P. Terhörst, On the Reliability of Biometric Datasets: How Much Test Data Ensures Reliability?, 11 de enero de 2025, *arXiv*. arXiv:2501.06504. doi: 10.48550/arXiv.2501.06504.
- [134] P. Kasprowski y I. Rigas, The influence of dataset quality on the results of behavioral biometric experiments, en *2013 International Conference of the BIOSIG Special Interest Group (BIOSIG)*, sep. 2013, pp. 1-8. Accedido: 2 de abril de 2025. [En línea]. Disponible en: <https://ieeexplore.ieee.org/abstract/document/6617159>
- [135] H. Ren, L. Sun, J. Guo, y C. Han, A Dataset and Benchmark for Multimodal Biometric Recognition Based on Fingerprint and Finger Vein, *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 2030-2043, 2022, doi: 10.1109/TIFS.2022.3175599.
- [136] N. Nnamoko, J. Barrowclough, M. Liptrott, y I. Korkontzelos, A behaviour biometrics dataset for user identification and authentication, *Data in Brief*, vol. 45, p. 108728, dic. 2022, doi: 10.1016/j.dib.2022.108728.
- [137] Home | European Open Science Cloud - EU Node. Accedido: 13 de mayo de 2025. [En línea]. Disponible en: <https://open-science-cloud.ec.europa.eu/>
- [138] esferize, Crue aprueba el compromiso de las universidades españolas para implantar la Open Science, CRUE. Accedido: 13 de mayo de 2025. [En línea]. Disponible en: <https://www.crue.org/2019/02/crue-aprueba-compromiso-open-science/>
- [139] Guía para la evaluación de repositorios institucionales.
- [140] M. T. Malo de Molina Martín-Montalvo *et al.*, Medición del Acceso Abierto en las universidades españolas y el CSIC (2014-2018). e-cienciaDatos, pp. 7014, 20254591, 20858652, 22167712, 23063382, 23217083, 7014, 7026, 2019. doi: 10.21950/UQAK4C.
- [141] B. Aguilera Caballero *et al.*, Medición del Acceso Abierto en las universidades españolas y el CSIC (2019-2023). e-cienciaDatos, pp. 24197, 111562462, 7077, 2025. doi: 10.21950/BOMYUU.
- [142] National Institutes of Health (NIH) | Turning Discovery Into Health. Accedido: 13 de mayo de 2025. [En línea]. Disponible en: <https://www.nih.gov/>

- [143] S. Stall *et al.*, Generalist Repository Comparison Chart, may 2023, doi: 10.5281/zenodo.7946938.
- [144] Data Repositories | Data Management. Accedido: 7 de abril de 2025. [En línea]. Disponible en: <https://datamanagement.hms.harvard.edu/share-publish/data-repositories>
- [145] PRISMA 2020 statement, PRISMA statement. Accedido: 24 de mayo de 2025. [En línea]. Disponible en: <https://www.prisma-statement.org/prisma-2020>
- [146] J. Vegas, C. Llamas, M. A. González, y C. Hernández, Identifying users from the interaction with a door handle, *Pervasive and Mobile Computing*, vol. 70, p. 101293, ene. 2021, doi: 10.1016/j.pmcj.2020.101293.
- [147] S. Xu, Y. Wang, y M. Si, A Two-Step Fusion Method of Wi-Fi FTM for Indoor Positioning, *Sensors*, vol. 22, n.º 9, Art. n.º 9, ene. 2022, doi: 10.3390/s22093593.
- [148] W. Cui, Z. Wang, y Y. Li, ECG-based biometric recognition under exercise and rest situations, *Biomedical Engineering Advances*, vol. 2, p. 100008, dic. 2021, doi: 10.1016/j.bea.2021.100008.
- [149] J. Wang y G. Wang, Quality-Specific Hand Vein Recognition System, *IEEE Transactions on Information Forensics and Security*, vol. 12, n.º 11, pp. 2599-2610, nov. 2017, doi: 10.1109/TIFS.2017.2713340.
- [150] X. Zhong *et al.*, 3D dental biometrics: Alignment and matching of dental casts for human identification, *Computers in Industry*, vol. 64, n.º 9, pp. 1355-1370, dic. 2013, doi: 10.1016/j.compind.2013.06.005.
- [151] H. Zhao, Y. Chen, W. Pei, H. Chen, y Y. Wang, Towards online applications of EEG biometrics using visual evoked potentials, *Expert Systems with Applications*, vol. 177, p. 114961, sep. 2021, doi: 10.1016/j.eswa.2021.114961.
- [152] R. Zhou *et al.*, ECG-based biometric under different psychological stress states, *Computer Methods and Programs in Biomedicine*, vol. 202, p. 106005, abr. 2021, doi: 10.1016/j.cmpb.2021.106005.
- [153] S.-L. Wang y A. W.-C. Liew, Physiological and behavioral lip biometrics: A comprehensive study of their discriminative power, *Pattern Recognition*, vol. 45, n.º 9, pp. 3328-3335, sep. 2012, doi: 10.1016/j.patcog.2012.02.016.
- [154] L. Lu, L. Liu, M. J. Hussain, y Y. Liu, I Sense You by Breath: Speaker Recognition via Breath Biometrics, *IEEE Transactions on Dependable and Secure Computing*, vol. 17, n.º 2, pp. 306-319, mar. 2020, doi: 10.1109/TDSC.2017.2767587.
- [155] T. Zhao, Y. Wang, J. Liu, Y. Chen, J. Cheng, y J. Yu, TrueHeart: Continuous Authentication on Wrist-worn Wearables Using PPG-based Biometrics, en *IEEE*

- INFOCOM 2020 - IEEE Conference on Computer Communications*, jul. 2020, pp. 30-39. doi: 10.1109/INFOCOM41043.2020.9155526.
- [156] X. Zhao *et al.*, Mapping sweat pores for biometric identification based on a donor-acceptor hydrophilic fluorescent probe, *Chinese Chemical Letters*, vol. 36, n.º 8, p. 110617, ago. 2025, doi: 10.1016/j.cclet.2024.110617.
- [157] U.-U. A. de Barcelona, Consentimientos informados, UAB Barcelona. Accedido: 18 de julio de 2025. [En línea]. Disponible en: <https://www.uab.cat/es/etica-investigacion/consentimientos-informados>
- [158] U.-U. A. de Barcelona, Consentimientos informados - Modelo de consentimiento informado (mayores de 18 años). Accedido: 18 de julio de 2025. [En línea]. Disponible en: <https://www.uab.cat/es/etica-investigacion/doc/consentimiento-informado-mayor-18.docx>
- [159] U.-U. A. de Barcelona, Consentimientos informados - Modelo de consentimiento informado genérico biomedicina. Accedido: 18 de julio de 2025. [En línea]. Disponible en: <https://www.uab.cat/es/etica-investigacion/doc/consentimiento-informado-generico-biomedicina.docx>

Apéndices

Apéndice I: Plantilla genérica de consentimiento informado

En este apéndice se va a mostrar los datos que deben aparecer y una plantilla de cómo podría ser el documento del consentimiento informado que tengan que firmar los participantes a la investigación con el objetivo de ceder sus datos de forma anónima para la investigación.

Los datos que deben aparecer son:

- Organización que recoge los datos.
- Finalidad y uso de los datos recogidos, detallados uno a uno.
- Tiempo que la organización va a mantener los datos.
- Consentimiento para transferencias internacionales, ya que se publicarán los datos en un repositorio de alcance global.

Tomando como referencia los modelos de consentimiento informado de la Universidad de Barcelona [157], en concreto el modelo para mayores de 18 años [158] y el modelo genérico biomedicina [159], se ha redactado la plantilla genérica para el consentimiento informado que deben firmar los participantes en la investigación es como se detalla a continuación.



Universidad de Valladolid



Consentimiento Informado

Título del proyecto: _____

Número de referencia: _____

Le pedimos que lea cuidadosamente este documento de consentimiento antes de decidirte a participar en esta investigación.

Objetivo de la investigación

El propósito de esta investigación es _____ con el objetivo _____.

Implicaciones de la participación

En primer lugar, le pediremos algunos datos básicos para realizar estadísticas sobre los participantes. Para terminar, le solicitaremos que realice XX repeticiones de la prueba _____, durante cada repetición recogeremos los datos recogidos por los diferentes sensores instalados.

La lista de sensores utilizados es:

- Sensor A recoge señales de _____.
- Sensor B recoge señales de _____.

Duración

La duración total aproximada de la participación es de 90 minutos, incluyéndose la sesión informativa, la recogida de datos estadísticos y la realización de las repeticiones de la prueba durante las cuales se recogerán los datos.

Riesgos y beneficios

La participación en esta investigación no conlleva ningún tipo de riesgo.

Compensación

La participación en esta investigación no conlleva ninguna compensación por participar.

Confidencialidad

Si decide participar, su identidad se mantendrá confidencial y sólo los miembros del equipo de investigación tendrán acceso a los datos de la investigación. Para la publicación de la investigación, se publicarán las estadísticas globales de los participantes y los datos recogidos por los sensores, sin relacionarlos con una persona concreta.

El equipo investigador mantendrá este consentimiento informado en un lugar seguro y lo destruirá al cabo de 5 años de finalizarse la investigación. Cuando la investigación se haya completado y se hayan analizado los datos, toda la base de datos será anonimizada y publicada en un repositorio de datos para que esté accesible para otros investigadores o personas interesadas.

Voluntariedad de la participación

La participación en esta investigación es completamente voluntaria. No hay ninguna penalización por no participar.

Tratamiento de datos personales

La realización de la investigación implica el tratamiento de sus datos personales por parte del equipo investigador de la Universidad de Valladolid para las finalidades descritas en este documento y basado en su consentimiento.

El equipo responsable de la investigación no cederá sus datos personales a terceras personas ni a otros destinatarios, ni tampoco serán objeto de transferencia internacional, salvo los datos recogidos por los sensores objeto de esta investigación, que serán publicados anonimizados sin posibilidad de identificar unívocamente al participante.

El tratamiento de sus datos personales no implica la toma de decisiones automatizadas ni la elaboración de perfiles predictivos de preferencias personales, comportamientos o actitudes.

El equipo investigador conservará sus datos personales durante el tiempo necesario para la realización el proyecto.

Tiene derecho a retirarse del estudio en cualquier momento sin dar explicaciones y sin consecuencias negativas; sólo nos lo tiene que comunicar por cualquier medio. Además, si así lo desea, puede ejercer sus derechos reconocidos por el Reglamento Europeo de Protección de Datos Personales dirigiéndote a la persona responsable del tratamiento _____ (indicar el nombre y apellido del personal investigador principal o el/la tutor/a o director/a de la tesis, y su correo electrónico) con la solicitud y una fotocopia del DNI o de cualquier otro documento identificativo.

Puede obtener más información sobre el tratamiento de datos personales y el ejercicio de derechos y reclamaciones en la dirección <https://secretariageneral.uva.es/competencias/proteccion-de-datos/> o en el correo electrónico de contacto responsable.privacidad@uva.es.

También puedes contactar con el Delegado de Protección de Datos de la Universidad de Valladolid mediante el correo electrónico delegado.proteccion.datos@uva.es.

En cualquier caso, recibirá una respuesta de la acción realizada en el plazo establecido legalmente.

Publicación/reutilización/otros procesamientos de los datos básicos y periodo de retención

Los datos de la investigación serán puestos a disposición pública de manera anonimizada una vez transcurridos *[año]* años a partir de la finalización del proyecto. Los identificadores personales serán _____ (destruidos/mantenidos confidencialmente de forma segura hasta FECHA/mantenidos confidencialmente de forma segura hasta que se alcancen los objetivos del proyecto de investigación).

Grabaciones y uso de testimonios

- Estoy de acuerdo en que se grabe (audio/vídeo) la entrevista con objetivos de investigación.
- Autorizo que se hagan citaciones literales de mis intervenciones sin mencionar mi nombre y que se reproduzcan en audio y vídeo intervenciones mías sin mencionar mi nombre.
- Autorizo el uso de mis intervenciones de audio y vídeo para fines de divulgación científica, siempre que se articulen mecanismos para preservar mi privacidad.

Persona de contacto

En caso de duda o consulta, puede contactar con el responsable de la investigación _____ mediante el correo electrónico nombre@uva.es, en el teléfono *[número]* o en la dirección postal P.º de Belén, 15, 47011 Valladolid.

Consentimiento

- He leído la información sobre el proyecto de investigación y he tenido la oportunidad de hacer preguntas, que me han sido contestadas satisfactoriamente.

- Entiendo que la información anonimizada (sin identificadores personales) de este proyecto será puesta a disposición de otros equipos investigadores un tiempo después de haber finalizado el proyecto.
- Estoy de acuerdo en participar y he recibido una copia de este consentimiento.

Consentimiento

D./D^a. (nombre y dos apellidos) _____ con NIF número _____.

DECLARO que he leído la información del proyecto de investigación que se me ha entregado, que he podido hacer preguntas al respecto las cuales han sido respondidas satisfactoriamente y que he recibido suficiente información por parte del Investigador/a _____, quien me ha explicado todos los detalles, en particular:

1. La finalidad de la investigación o de la línea de investigación para la que doy mi consentimiento.
2. Los resultados que se esperan obtener.
3. Los posibles inconvenientes asociados a la investigación, incluida la posibilidad de que se pongan en contacto conmigo posteriormente con la finalidad de recoger nuevos datos u obtener otras muestras.
4. La identidad de la persona responsable de esta investigación, _____, doctor/a de la Universidad de Valladolid.
5. El derecho que me asiste de revocación de este consentimiento y sus efectos, incluida la posibilidad de la destrucción de los datos recogidos, y el hecho de que estos efectos no se extenderán a los datos resultantes de las investigaciones que ya se hayan llevado a cabo.
6. La información anonimizada (sin identificadores personales) de esta investigación será puesta a disposición general un tiempo después de haber finalizado el proyecto

7. La garantía de confidencialidad que se me ha dado sobre la información obtenida y mis datos de carácter personal.

Declaro, además, que comprendo que mi participación es voluntaria, por lo que puedo retirarme de la investigación cuando lo desee, sin tener que dar explicaciones.

Firma del/de la participante:

Firma del investigador:

Nombre:

Nombre:

Fecha:

Fecha:

Este documento se firmará por duplicado; el investigador se quedará una copia y el participante, la otra.

Apéndice II: Tabla de datos obtenidos de los artículos revisados

En este apéndice se van a presentar los datos recogidos de cada artículo. Por motivos de visibilidad, se van a presentar varias tablas, una para los datos estadísticos de los artículos, otra para los datos relativos al corpus de datos y una última para los datos relativos a los marcos legal y ético. Además, los artículos están identificados con un campo ID que hace referencia al ID asignado en la Tabla 8 del apartado 6.3.4 Artículos seleccionados para que no poner el nombre entero y que dificulte la correcta visualización.

ID Artículo	Número Citaciones	Número Autores	Número Instituciones	Número Palabras Clave	Tipo de dato
1	11	4	1	6	Apertura de puerta
2	0	3	1	5	Electrocardiograma (EGC)
3	41	2	1	8	Venas de la mano
4	19	7	2	3	Molde dental 3D
5	32	5	3	5	Electroencefalograma (EEG)
6	17	9	4	0	Electrocardiograma (EGC)
7	42	2	2	5	Labios
8	30	4	1	3	Respiración
9	63	6	5	0	Señal cardiaca fotopletismografía PPG
10	0	7	1	6	Poros sudoríparos de la mano

Tabla 14 Datos estadísticos

ID Artículo	Número Participantes	Tomas por participante	Número repeticiones / sesiones	Total Muestras	Diferenciación			
					Edad (años)	Género	Procedencia	Otro
1	47	20	1	960	18-68	34H/13M	NO	Altura
2	45	2	1	90	18-22	33H/12M	NO	NO
3	50	5	2	500	NO	NO	NO	NO
4	60 post mortem (PM) 200 ante mortem (AM)	1 1	1	260	PM: 12-34, media 24,4 ± 4,2 AM: 12-35, media 23,6 ± 5,4	PM: 24H/36M AM: 111H/89 M	Chinos, Indios y Malayos	NO
5	21 presencial 15 online	100	2	4200 3000	Media 23 Media 24	6H/15M 12H/3M	NO	Sanos y visión norm al/corr egida
6	23	200	3	13800	20-37	13H/10M	NO	Sanos
7	40	90	10	36000	NO	29H/11M	NO	NO
8	50 1ª muestra 20 2ª muestra	25 1	1 32	1250 640	10-60	NO	NO	NO
9	20	1	20 personas durante 10 minutos estáticos y 5 personas durante 2 minutos	15000 segmentos estáticos 4200 segmentos con	20-40	NO	NO	Sanos

			moviéndose y 3 minutos sentados	movimient os				
10	6	1	10	60	NO	NO	NO	NO

Tabla 15 Datos sobre participantes

ID Artículo	Leyes	Normas / Protocolos	Comité ético	Consentimiento Informado	Financiación	Conflicto intereses	Publica corpus
1	NO	NO	NO	SI	SI	SI	NO
2	NO	NO	NO	NO	SI	SI	NO
3	NO	NO	NO	NO	SI	NO	NO
4	NO	NO	NO	NO	NO	NO	NO
5	NO	NO	SI	SI	SI	SI	NO
6	NO	NO	NO	NO	SI	SI	NO
7	NO	NO	NO	NO	SI	NO	NO
8	NO	NO	NO	NO	SI	NO	NO
9	NO	NO	NO	NO	SI	NO	NO
10	NO	NO	SI	SI	SI	SI	NO

Tabla 16 Marco legal y Marco ético

Apéndice III: PRISMA 2020

Este trabajo al ser una revisión sistemática de la literatura existente, se va a mostrar el proceso seguido en función de la guía PRISMA [15] (Preferred Reporting Items for Systematic reviews and Meta-Analyses o Elementos de notificación preferidos para revisiones sistemáticas y metaanálisis en español) en su versión 2020 [145]. La primera versión de la guía PRISMA surgió en 2009, su aparición se debió debido a la gran proliferación de revisiones sistemáticas que se estaban realizando y la utilidad que proveen al concentrar diversas investigaciones en un solo punto. La guía PRISMA 2009 y su posterior actualización PRISMA 2020 dotan a los autores de herramientas que facilitan el objetivo de que sus publicaciones de revisiones sistemáticas sean transparentes, completas y precisas, describiendo por qué se ha realizado la investigación, cómo se han identificado y seleccionado las fuentes, explicando las fuentes descartadas y los motivos o los resultados de los análisis.

La forma de usar PRISMA 2020 es a través de su lista de verificación, que está compuesta de siete secciones con 27 ítems, algunos de los cuales están compuestos por subítems, existe una versión resumida de 12 ítems para resúmenes estructurados y del uso de un diagrama de flujo para mostrar el proceso de selección de las fuentes.

Aunque PRIMSA 2020 se ha diseñado fundamentalmente para revisiones sistemáticas de intervenciones sanitarias, los ítems de su lista de verificación o parte de sus ítems son válidos para otros contextos, por lo que esta guía es ampliamente utilizada.

Se van a mostrar la tabla de verificación y el diagrama de selección de artículos aplicados a este trabajo.

Sección / Tema	Ítem n.º	Ítem de la lista de verificación	Localización del ítem en la memoria del TFG
TÍTULO			

Sección / Tema	Ítem	Ítem de la lista de verificación n.º	Localización del ítem en la memoria del TFG
Título	1	Identifique el reporte como una revisión sistemática.	Portada
RESUMEN			
Resumen estructurado	2	Vea la lista de verificación para resúmenes estructurados de la declaración PRISMA 2020.	Apartado Resumen
INTRODUCCIÓN			
Justificación	3	Describa la justificación de la revisión en el contexto del conocimiento existente.	Apartado 1.1 Motivación
Objetivos	4	Proporcione una declaración explícita de los objetivos o las preguntas que aborda la revisión.	Apartado 1.2 Objetivos del proyecto
MÉTODOS			
Criterios de elegibilidad	5	Especifique los criterios de inclusión y exclusión de la revisión y cómo se agruparon los estudios para la síntesis.	Apartado 6.3.1 Criterios de elección de la literatura
Fuentes de información	6	Especifique todas las bases de datos, registros, sitios web, organizaciones, listas de referencias y otros recursos de búsqueda o consulta para identificar los estudios. Especifique la fecha en la que cada recurso se buscó o consultó por última vez.	Apartado 6.3.1 Criterios de elección de la literatura, Apartado 6.3.2 Análisis inicial y Apartado 6.3.3 Búsqueda adicional de literatura

Sección / Tema	Ítem n.º	Ítem de la lista de verificación	Localización del ítem en la memoria del TFG
Estrategia de búsqueda	7	Presente las estrategias de búsqueda completas de todas las bases de datos, registros y sitios web, incluyendo cualquier filtro y los límites utilizados.	Apartado 6.3.1 Criterios de elección de la literatura, Apartado 6.3.2 Análisis inicial y Apartado 6.3.3 Búsqueda adicional de literatura
Proceso de selección de los estudios	8	Especifique los métodos utilizados para decidir si un estudio cumple con los criterios de inclusión de la revisión, incluyendo cuántos autores de la revisión cribaron cada registro y cada reporte recuperado, si trabajaron de manera independiente y, si procede, los detalles de las herramientas de automatización utilizadas en el proceso.	Apartado 6.3.1 Criterios de elección de la literatura, Apartado 6.3.2 Análisis inicial y Apartado 6.3.3 Búsqueda adicional de literatura
Proceso de extracción de los datos	9	Indique los métodos utilizados para extraer los datos de los informes o reportes, incluyendo cuántos revisores recopilaron datos de cada reporte, si trabajaron de manera independiente, los procesos para obtener o confirmar los datos por parte de los investigadores del estudio y, si procede, los detalles de las herramientas de automatización utilizadas en el proceso.	Apartado 6.3.1 Criterios de elección de la literatura, Apartado 6.3.2 Análisis inicial y Apartado 6.3.3 Búsqueda adicional de literatura

Sección / Tema	Ítem n.º	Ítem de la lista de verificación	Localización del ítem en la memoria del TFG
Lista de los datos	10a	Enumere y defina todos los desenlaces para los que se buscaron los datos. Especifique si se buscaron todos los resultados compatibles con cada dominio del desenlace (por ejemplo, para todas las escalas de medida, puntos temporales, análisis) y, de no ser así, los métodos utilizados para decidir los resultados que se debían recoger.	Apartado 6.3.1 Criterios de elección de la literatura, Apartado 6.3.2 Análisis inicial y Apartado 6.3.3 Búsqueda adicional de literatura
	10b	Enumere y defina todas las demás variables para las que se buscaron datos (por ejemplo, características de los participantes y de la intervención, fuentes de financiación). Describa todos los supuestos formulados sobre cualquier información ausente (<i>missing</i>) o incierta.	
Evaluación del riesgo de sesgo de los estudios individuales	11	Especifique los métodos utilizados para evaluar el riesgo de sesgo de los estudios incluidos, incluyendo detalles de las herramientas utilizadas, cuántos autores de la revisión evaluaron cada estudio y si trabajaron de manera independiente y, si procede, los detalles de las herramientas de automatización utilizadas en el proceso.	Apartado 6.3.5 Riesgo de sesgo
Medidas del efecto	12	Especifique, para cada desenlace, las medidas del efecto (por ejemplo,	Apartado 6.3.5 Riesgo de sesgo

Sección / Tema	Ítem n.º	Ítem de la lista de verificación	Localización del ítem en la memoria del TFG
		razón de riesgos, diferencia de medias) utilizadas en la síntesis o presentación de los resultados.	
Métodos de síntesis	13a	Describa el proceso utilizado para decidir qué estudios eran elegibles para cada síntesis (por ejemplo, tabulando las características de los estudios de intervención y comparándolas con los grupos previstos para cada síntesis (ítem n.º 5).	Apartado Criterios de elección de la literatura, Apartado 6.3.2 Análisis inicial y Apartado 6.3.3 Búsqueda adicional de literatura
	13b	Describa cualquier método requerido para preparar los datos para su presentación o síntesis, tales como el manejo de los datos perdidos en los estadísticos de resumen o las conversiones de datos.	Apartado Resultados obtenidos
	13c	Describa los métodos utilizados para tabular o presentar visualmente los resultados de los estudios individuales y su síntesis.	Apartado Resultados obtenidos
	13d	Describa los métodos utilizados para sintetizar los resultados y justifique sus elecciones. Si se ha realizado un metaanálisis, describa los modelos, los métodos para identificar la presencia y el alcance de la heterogeneidad estadística, y los programas informáticos utilizados.	Apartado Resultados obtenidos

Sección / Tema	Ítem n.º	Ítem de la lista de verificación	Localización del ítem en la memoria del TFG
	13e	Describa los métodos utilizados para explorar las posibles causas de heterogeneidad entre los resultados de los estudios (por ejemplo, análisis de subgrupos, metarregresión).	Apartado 6.4 Resultados obtenidos
	13f	Describa los análisis de sensibilidad que se hayan realizado para evaluar la robustez de los resultados de la síntesis.	Apartado 6.4 Resultados obtenidos
Evaluación del sesgo en el reporte	14	Describa los métodos utilizados para evaluar el riesgo de sesgo debido a resultados faltantes en una síntesis (derivados de los sesgos en los reportes).	Apartado 6.3.5 Riesgo de sesgo
Evaluación de la certeza de la evidencia	15	Describa los métodos utilizados para evaluar la certeza (o confianza) en el cuerpo de la evidencia para cada desenlace.	Apartado 6.4 Resultados obtenidos
RESULTADOS			
Selección de los estudios	16a	Describa los resultados de los procesos de búsqueda y selección, desde el número de registros identificados en la búsqueda hasta el número de estudios incluidos en la revisión, idealmente utilizando un diagrama de flujo.	Apéndice III: PRISMA 2020 - Diagrama de flujo de PRISMA 2020
	16b	Cite los estudios que aparentemente cumplían con los criterios de inclusión, pero que fueron excluidos, y explique por qué fueron excluidos.	Apéndice III: PRISMA 2020 - Diagrama de flujo de PRISMA 2020

Sección / Tema	Ítem n.º	Ítem de la lista de verificación	Localización del ítem en la memoria del TFG
Características de los estudios	17	Cite cada estudio incluido y presente sus características.	Apartado Artículos seleccionados 6.3.4
Riesgo de sesgo de los estudios individuales	18	Presente las evaluaciones del riesgo de sesgo para cada uno de los estudios incluidos.	Apartado Riesgo de sesgo 6.3.5
Resultados de los estudios individuales	19	Presente, para todos los desenlaces y para cada estudio: a) los estadísticos de resumen para cada grupo (si procede) y b) la estimación del efecto y su precisión (por ejemplo, intervalo de credibilidad o de confianza), idealmente utilizando tablas estructuradas o gráficos.	Apartado Artículos seleccionados 6.3.4
Resultados de la síntesis	20a	Para cada síntesis, resuma brevemente las características y el riesgo de sesgo entre los estudios contribuyentes.	Apartado Resultados obtenidos 6.4
	20b	Presente los resultados de todas las síntesis estadísticas realizadas. Si se ha realizado un metaanálisis, presente para cada uno de ellos el estimador de resumen y su precisión (por ejemplo, intervalo de credibilidad o de confianza) y las medidas de heterogeneidad estadística. Si se comparan grupos, describa la dirección del efecto.	Apartado Resultados obtenidos 6.4
	20c	Presente los resultados de todas las investigaciones sobre las posibles	Apartado Resultados obtenidos 6.4

Sección / Tema	Ítem	Ítem de la lista de verificación n.º	Localización del ítem en la memoria del TFG
		causas de heterogeneidad entre los resultados de los estudios.	
	20d	Presente los resultados de todos los análisis de sensibilidad realizados para evaluar la robustez de los resultados sintetizados.	Apartado 6.4 Resultados obtenidos
Sesgos en el reporte	21	Presente las evaluaciones del riesgo de sesgo debido a resultados faltantes (derivados de los sesgos del reporte) para cada síntesis evaluada.	Apartado 6.3.5 Riesgo de sesgo
Certeza de la evidencia	22	Presente las evaluaciones de la certeza (o confianza) en el cuerpo de la evidencia para cada desenlace evaluado.	Apartado 6.3.1 Criterios de elección de la literatura, Apartado 6.3.2 Análisis inicial y Apartado 6.3.3 Búsqueda adicional de literatura
DISCUSIÓN			
Discusión	23a	Proporcione una interpretación general de los resultados en el contexto de otras evidencias.	Apartado 6.4 Resultados obtenidos
	23b	Argumente las limitaciones de la evidencia incluida en la revisión.	Apartado 6.3.1 Criterios de elección de la literatura, Apartado 6.3.2 Análisis inicial y Apartado 6.3.3

Sección / Tema	Ítem n.º	Ítem de la lista de verificación	Localización del ítem en la memoria del TFG
			Búsqueda adicional de literatura
	23c	Argumente las limitaciones de los procesos de revisión utilizados.	Apartado 6.3.1 Criterios de elección de la literatura, Apartado 6.3.2 Análisis inicial y Apartado 6.3.3 Búsqueda adicional de literatura
OTRA INFORMACIÓN			
Registro y protocolo	24a	Proporcione la información del registro de la revisión, incluyendo el nombre y el número de registro, o declare que la revisión no ha sido registrada.	NO PROCEDE - No está registrado hasta su presentación
	24b	Indique dónde se puede acceder al protocolo, o declare que no se ha redactado ningún protocolo.	NO PROCEDE
	24c	Describa y explique cualquier enmienda a la información proporcionada en el registro o en el protocolo.	NO PROCEDE
Financiación	25	Describa las fuentes de apoyo financiero o no financiero para la revisión y el papel de los	NO PROCEDE

Sección / Tema	Ítem n.º	Ítem de la lista de verificación	Localización del ítem en la memoria del TFG
		financiadores o patrocinadores en la revisión.	
Conflicto de intereses	26	Declare los conflictos de intereses de los autores de la revisión.	Apéndice IV: Información adicional - Declaración de conflicto de intereses
Disponibilidad de datos, códigos y otros materiales	27	Especifique qué elementos de los que se indican a continuación están disponibles al público y dónde se pueden encontrar: plantillas de formularios de extracción de datos, datos extraídos de los estudios incluidos, datos utilizados para todos los análisis, código de análisis, cualquier otro material utilizado en la revisión.	Apéndice IV: Información adicional - Disponibilidad de datos

Tabla 17 Lista de verificación PRISMA 2020

Diagrama de flujo de PRISMA 2020

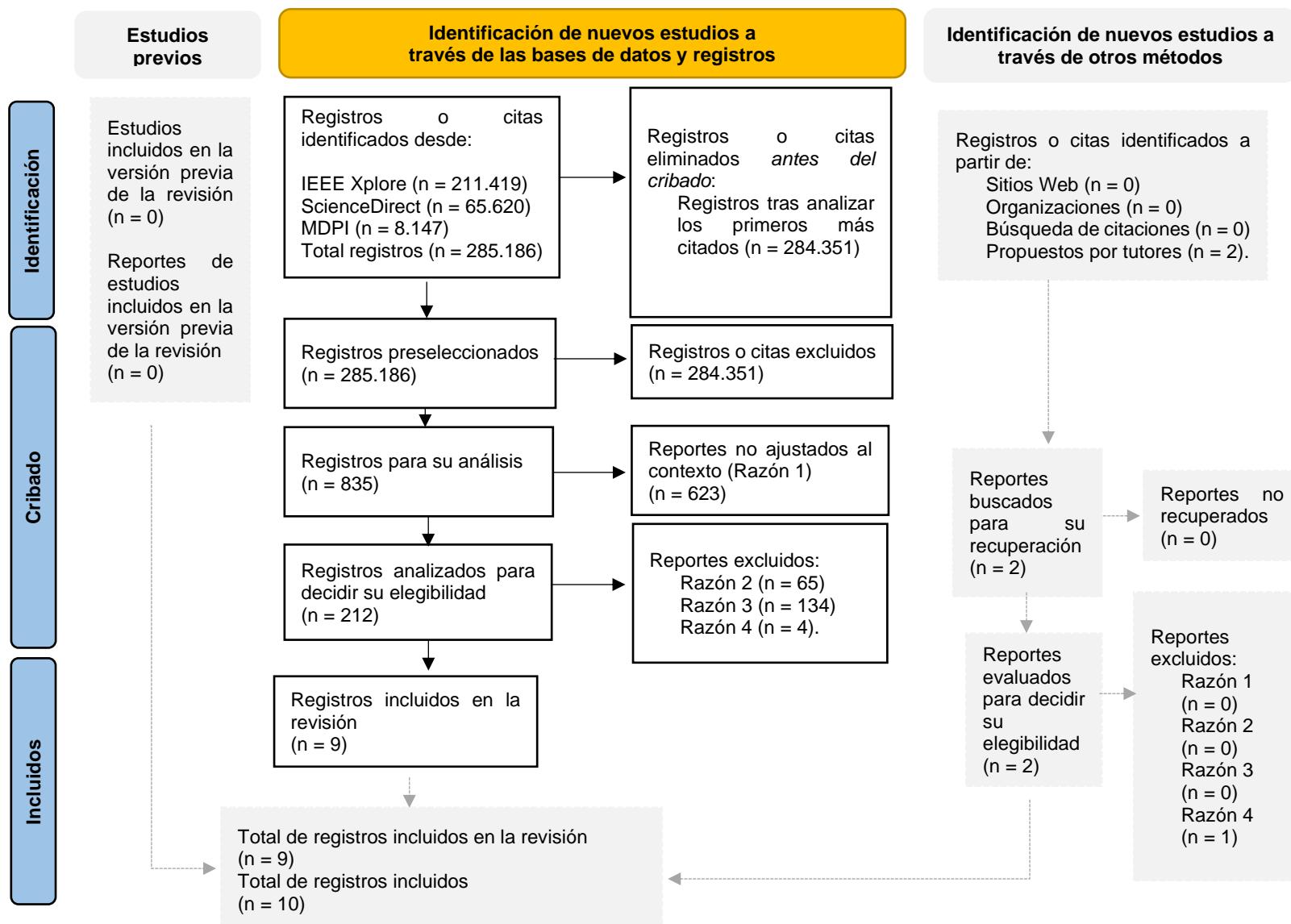
Antes de exponer el diagrama de flujo, se van a recordar las razones de exclusión:

Razón 1: No ajustarse al contexto de la investigación, esta razón se incluyen los artículos que son revisiones sistemáticas de literatura.

Razón 2: No ajustarse a alguno de los países determinados.

Razón 3: No crea su propio corpus de datos, utiliza uno público.

Razón 4: No describe como se ha formado el corpus de datos o el corpus no es válido por no ser referente a personas.



Apéndice IV: Información adicional

Una de las conclusiones a las que se ha llegado en este trabajo es que en los artículos es conveniente que aparezcan las secciones de declaración de conflicto de intereses y disponibilidad de los datos. Debido a esto, se ha considerado oportuno que dicha información aparezca en este trabajo y se refleja a continuación.

Declaración de conflicto de intereses

El autor notifica que no existe ningún conflicto de intereses en la realización de este trabajo.

Disponibilidad de datos

No se utilizaron datos para la investigación descrita en el artículo.