

Gestión de datos de investigación en el contexto del Reglamento sobre el Espacio Europeo de Datos de Salud: desarrollo de estrategias

Laura Mafalda Carvalho Lopes

DIRECTORES

José Antonio Frías

António José de Almeida Soares

PLAN DE INVESTIGACIÓN

PROGRAMA DE DOCTORADO en Formación en la Sociedad del Conocimiento

UNIVERSIDADE DE SALAMANCA

INTRODUCCIÓN

La gestión de datos de investigación ha sido objeto de estudio y debate por parte de diversos grupos de investigación en distintos campos científicos. Se reconoce, por un lado, la creciente importancia de los datos para la actividad científica y, por otro, la preocupación por el aumento de los volúmenes de datos que necesitan ser gestionados. Hodson (2009) afirma que, a medida que el volumen de datos crece, también lo hace la necesidad de encontrar formas eficaces de gestión y conservación, lo que permite su compartición y reutilización. Muchas organizaciones están buscando soluciones que varían según los tipos de datos producidos. Por lo tanto, es fundamental que existan políticas a nivel internacional, nacional e institucional, que incluyan las acciones necesarias que los investigadores deben emprender (Hodson, 2009).

Como señala Surkis (2015), los datos de investigación van más allá de las hojas de cálculo con números: pueden incluir bioespecímenes, grabaciones de video, imágenes, software, algoritmos y cuadernos de laboratorio. Tal vez sea más útil pensar en los datos como todo lo necesario para reproducir un resultado científico concreto. Este entendimiento es fundamental, dado que los organismos de financiamiento y los gobiernos exigen cada vez más que los beneficiarios de subvenciones conserven y hagan accesibles los datos resultantes de sus investigaciones para su reutilización en futuros proyectos (Chiwere & Mathe, 2015).

Los datos de investigación constituyen la base de cualquier resultado científico y, por lo tanto, su gestión es esencial para garantizar la plena utilización y validación de los resultados (Carvalho, 2017). En consonancia con las recomendaciones de la Comisión Europea (CE) para el programa Horizonte 2020 (H2020), las políticas de datos abiertos tienen como objetivo mejorar el acceso y la reutilización de los datos científicos (Organización para la Cooperación y el Desarrollo Económicos [OCDE], 2007). Los investigadores, por tanto, están obligados a depositar los datos, incluidos los metadatos, necesarios para validar los resultados de sus publicaciones y a desarrollar planes de gestión de datos.

Ejemplos de mejores prácticas en la gestión de datos de investigación

1. **Planes de Gestión de Datos (PGD):** Un paso fundamental para una gestión eficaz de datos es la creación de un Plan de Gestión de Datos (PGD). Este documento describe cómo se recopilarán, organizarán, almacenarán, compartirán y preservarán los datos a lo largo del proyecto y después de su finalización. En el contexto del Espacio Europeo de Datos Sanitarios (EEDS), un buen PGD incluiría directrices específicas para el almacenamiento seguro de datos sensibles de salud, como datos genéticos e historiales clínicos, cumpliendo con las normas de protección de datos y privacidad (Reglamento General de Protección de Datos [RGPD]). Herramientas como el *Data Management Planning Tool* (DMPTool) y *DMPonline* son útiles para la creación de planes que cumplan con los requisitos europeos.
2. **Infraestructuras Seguras de Almacenamiento de Datos:** Los datos sanitarios incluyen con frecuencia información personal sensible, lo que exige infraestructuras seguras que garanticen la protección de dichos datos. Ejemplos de buenas prácticas incluyen el uso de plataformas como la *European Open Science Cloud (EOSC)*, que ofrece almacenamiento seguro y compartición controlada de datos de investigación dentro de la Unión Europea. Además, es esencial utilizar sistemas de seguridad robustos, como la encriptación y la autenticación multifactor, para proteger los datos de accesos no autorizados.
3. **Anonimización y Pseudonimización de los Datos:** Para garantizar el cumplimiento de las normativas de protección de datos, una de las prácticas recomendadas es la anonimización de los datos personales o, cuando esto no es posible, la pseudonimización. La anonimización elimina cualquier identificación personal de los datos, mientras que la pseudonimización sustituye los datos identificables por un identificador falso, permitiendo su posible reversión bajo control. Un ejemplo de esta aplicación es el uso de software como el *Arx Data Anonymization Tool*.
4. **Metadatos Estandarizados:** El uso de metadatos estandarizados es una práctica esencial para garantizar la reutilización de los datos y su interoperabilidad entre diferentes plataformas y sistemas. El uso de esquemas de metadatos como el *Dublin Core* o *Health Level 7 (HL7)* permite que otros investigadores encuentren, comprendan y reutilicen los datos de forma eficiente. Esto es particularmente relevante en salud, donde los datos suelen ser complejos y de diferentes formatos (imágenes médicas, datos genéticos, etc.).

5. **Repositorios de Datos Abiertos:** Después de la finalización de un proyecto, se recomienda depositar los datos de investigación en repositorios de acceso abierto, como *Zenodo*, *OpenAIRE* o *Figshare*. Estos repositorios permiten almacenar los datos de forma segura y accesible, promoviendo la transparencia científica y la reutilización de los datos en nuevos proyectos. En el campo de la salud, repositorios especializados como el *European Genome-phenome Archive (EGA)* son una excelente opción para almacenar datos biomédicos.
6. **Formación y Concienciación de los Investigadores:** La formación continua de los investigadores en las mejores prácticas de gestión de datos es fundamental. La Comisión Europea, a través del *FOSTER Open Science*, ofrece diversos cursos y recursos educativos sobre ciencia abierta y gestión de datos. La concienciación sobre la importancia de seguir las directrices de datos abiertos y la legislación de protección de datos (como el EEDS y el RGPD) también debería ser una prioridad.

Contexto del Reglamento del Espacio Europeo de Datos Sanitarios

En este contexto, el presente trabajo explora las mejores prácticas para la gestión de datos de investigación en el ámbito de la salud, prestando especial atención al cumplimiento del Reglamento del Espacio Europeo de Datos Sanitarios (Comisión Europea, 2022). Este reglamento define directrices estrictas para la protección de datos, el consentimiento informado y la seguridad, afectando directamente la gestión de los datos de investigación. El estudio pretende identificar estrategias eficientes que mejoren la calidad de la investigación y promuevan la colaboración interdisciplinar, contribuyendo al avance de las ciencias de la salud.

El Reglamento del Espacio Europeo de Datos Sanitarios (Comisión Europea, 2022) promueve la interoperabilidad y el intercambio seguro de datos en toda la Unión Europea (García Pérez, 2023). Al introducir un nuevo paradigma de gestión de datos, el reglamento hace imprescindible el cumplimiento de sus normas, especialmente en cuanto a la protección de datos y la seguridad. Así, este estudio busca desarrollar estrategias eficientes, basadas en los principios de la Ciencia de la Información, para una gestión adecuada de los datos de investigación, contribuyendo a mejorar la investigación en salud y a avanzar en la práctica científica.

Dada la inminente aplicación del reglamento, existe una necesidad urgente de desarrollar estrategias que cumplan con las normativas europeas y fomenten la colaboración interdisciplinar. Este trabajo pretende llenar un vacío existente, proporcionando orientaciones valiosas para investigadores, profesionales de la salud y reguladores en el contexto de la gestión de datos de investigación en salud.

HIPÓTESIS DE TRABAJO Y PRINCIPALES OBJETIVOS

El objetivo principal es analizar cómo puede afectar el Reglamento sobre el Espacio Europeo de Datos Sanitarios (EHDS) a la recogida, almacenamiento, tratamiento y puesta en común de datos de investigación en salud y cómo pueden aplicarse eficazmente sus principios.

A partir de aquí, se establecen 4 objetivos generales y los consiguientes objetivos específicos

1. 1. Analizar el Reglamento sobre el Espacio Europeo de Datos de Salud (EEES).

1.1 Realizar un análisis exhaustivo del EHDS, identificando sus directrices y normativa específica relativa a la gestión de datos de salud.

1.2 Evaluar cómo las instituciones de investigación, los hospitales y los investigadores están aplicando y teniendo en cuenta el EHDS en la recogida, el almacenamiento, el tratamiento y la puesta en común de datos de salud.

1.3 Identificar los principales retos y oportunidades a los que se enfrentan las partes interesadas en la investigación sanitaria y la gestión de datos de salud a la hora de adherirse a la EHDS.

2. Desarrollar estrategias para el cumplimiento de la EHDS en la gestión de datos de salud.

2.1 Proponer estrategias eficaces para garantizar el cumplimiento de la EHDS en la recogida de datos de salud.

2.2 Elaborar directrices prácticas para el almacenamiento y el tratamiento seguros de datos de salud en cumplimiento de la EHDS.

2.3 Elaborar un modelo de plan de gestión de datos que cumpla las recomendaciones de la EHDS.

3. Identificar las mejores prácticas en la gestión de datos de investigación en salud de acuerdo con la EHDS.

3.1 Llevar a cabo una revisión bibliográfica exhaustiva sobre las mejores prácticas en la gestión de datos de investigación en salud en Europa, teniendo en cuenta el EHDS.

3.2 Identificar estudios de casos y ejemplos de aplicación con éxito de prácticas eficaces que se ajusten al EHDS.

3.3 Proponer directrices para la adopción de mejores prácticas en la gestión de datos de salud que cumplan con la normativa EHDS.

4. Promover la colaboración interdisciplinar en la gestión de datos de salud a la luz de la EHDS.

4.1 Investigar cómo una gestión eficiente de los datos de salud puede facilitar la colaboración interdisciplinar entre instituciones de investigación, hospitales y otras partes interesadas en la salud en Europa.

4.2 Proponer estrategias para fomentar la colaboración interdisciplinar en la gestión de datos de investigación en salud, en consonancia con las directrices de la EHDS relativas al uso secundario de datos.

4.3 Desarrollar un Plan Modelo de Gestión de Datos de Salud, crear un plan de gestión de datos que integre los resultados de los estudios y proporcione una guía práctica para los investigadores y profesionales de la salud que deseen gestionar los datos de investigación de acuerdo con las mejores prácticas y normativas.

METODOLOGÍA

Se trata de un estudio mixto. Cualitativo para el análisis documental de la legislación ya identificada. Cuantitativo para el análisis de los conocimientos y prácticas ya utilizados por los investigadores (1) y para la muestra que se incluirá en la elaboración del modelo de Plan de Gestión de Datos (2).

Como método de revisión de la literatura, se llevará a cabo una revisión sistemática de la literatura porque, aunque las preguntas de investigación son de carácter general, en el campo de la Ciencia de la Información es un tema específico que merece un estado del arte detallado. En este proceso, se utilizará la lista de verificación PRISMA para garantizar que la revisión sistemática se realiza de forma transparente, rigurosa, exhaustiva y replicable (Ferrerías-Fernández et al., 2016; García-Peñalvo, 2022).

La investigación se llevará a cabo de acuerdo con los preceptos éticos desarrollados en las Directrices Éticas para la Investigación Educativa (Asociación Británica de Investigación Educativa [BERA], 2019).

En cuanto al desarrollo de la investigación, se proponen dos estudios como parte de la tesis.

En primer lugar, un análisis cualitativo resultante del análisis de la documentación de los proyectos a incluir en este estudio mediante el mapeo de los principales conceptos del Reglamento del Espacio Europeo de Salud. La muestra, aún por definir, se situará en el contexto de la investigación CINTESIS.

Posteriormente, se elaborará un cuestionario con el objetivo de discutir los principales resultados y conclusiones obtenidos en la investigación cualitativa previa.

En segundo lugar, un análisis cuantitativo con el objetivo de informar sobre las prácticas en uso por parte de la comunidad investigadora que representa la muestra a incluir.

La muestra se seleccionará en el contexto de los investigadores que componen el CINTESIS. Los criterios de inclusión y exclusión se definirán a posteriori y se ajustarán a los requisitos formales para la realización de estudios cualitativos, a saber, la presentación de solicitudes de autorización al comité de ética para llevar a cabo el estudio, la creación de un formulario de consentimiento informado y la creación y realización de un cuestionario.

La recogida de datos incluirá entrevistas, cuestionarios, revisión de documentos, observación directa y análisis de historias clínicas, entre otros. Los cuestionarios se distribuirán digitalmente (herramienta aún por definir).

En cuanto a las limitaciones que pueden darse en el desarrollo de este proyecto, la dificultad de acceso a la población es un factor determinante. La muestra permite obtener una muestra representativa de la investigación en el ámbito estudiado, pero puede dejar fuera información relevante para la investigación. Además, se conoce el riesgo de una elevada tasa de no respuesta y habrá que prever medidas para ajustar la muestra.

Por último, se propone crear un documento rector de la política de gestión de datos en el CINTESIS y un plan modelo de gestión de los datos obtenidos en la investigación para que puedan ser reconocidos en el paradigma de la ciencia abierta. Como parte de este plan, se realizará una selección de formatos que permitan que los datos sean FAIR (recuperabilidad, accesibilidad, interoperabilidad y reutilización). (Wilkinson, 2016).

MEDIOS Y RECURSOS MATERIALES DISPONIBLES

El doctorando cuenta con los recursos materiales del CINTESIS y del Departamento de Biblioteconomía y Documentación de la Universidad de Salamanca, donde dispone de un espacio de trabajo durante la realización del doctorado.

Este trabajo se desarrolla en el marco del programa de doctorado: Formación en la Sociedad del Conocimiento (García-Peñalvo, 2014), y su portal (García-Peñalvo et al., 2019), accesible desde <http://knowledgesociety.usal.es>, es la principal herramienta de comunicación y visibilización de los avances. En él se incluirán todas las publicaciones, estancias y participaciones en congresos a lo largo del trabajo.

PLANIFICACIÓN TEMPORAL AJUSTADA A CINCO AÑOS (Tiempo parcial):

Año 1: Planificación y revisión bibliográfica

Tarea 1 - 1er trimestre

- Perfeccionar la propuesta de tesis.
- Establecer contacto con el supervisor, el co-supervisor y el comité de tesis.
- Definir objetivos de investigación más detallados.

Tarea 2 - 2º trimestre

- Realizar una revisión exhaustiva de la literatura relacionada con la gestión de datos de investigación sanitaria y la normativa europea.
- Identificar lagunas en el conocimiento.

Tarea 3 - 3er trimestre

- Finalizar el protocolo de investigación.
- Definir la metodología de recogida y análisis de datos.
- Preparar la revisión bibliográfica para la publicación de un artículo.

Año 2: Recogida de datos y desarrollo de la estrategia

Tarea 1 - 1er trimestre

- Obtener las aprobaciones éticas y normativas necesarias.
- Iniciar la recopilación de datos, incluidas entrevistas, cuestionarios y revisión de documentos.

Tarea 2 - 2º trimestre

- Analizar los datos recogidos.
- Desarrollar estrategias para el cumplimiento de la EHDS y otras normativas.

Tarea 3 - 3er trimestre

- Seguir analizando los datos y perfeccionar las estrategias según sea necesario.
- Identificar retos comunes en la gestión de datos sanitarios.

Año 3: Desarrollo de mejores prácticas y soluciones

Tarea 1 - 1er trimestre

- Proponer soluciones a los retos identificados.
- Empezar a desarrollar las mejores prácticas.

Tarea 2 - 2º trimestre

- Realizar estudios de casos para ejemplificar la aplicación de las prácticas propuestas.
- Continuar investigando soluciones innovadoras.

Tarea 3 - 3er trimestre

- Finalizar el desarrollo de las mejores prácticas.
- Preparar presentaciones y artículos relacionados con la investigación para conferencias.

Año 4: Evaluación y análisis de los resultados

Tarea 1 - 1er trimestre

- Evaluar la eficacia de las estrategias y buenas prácticas desarrolladas.
- Realizar análisis estadísticos y cualitativos de los resultados.

Tarea 2 - 2º trimestre

- Redactar capítulos relacionados con la evaluación y el análisis.
- Revisar y ajustar las prácticas y estrategias en función de los resultados.

Año 5: Conclusión, debate y redacción final

Tarea 1 - 1er trimestre

- Redactar los capítulos de discusión y conclusión.
- Identificar las implicaciones prácticas y teóricas de los resultados.

Tarea 2 - 2º trimestre

- Redactar la introducción y revisar el trabajo en su totalidad.

Tarea 3 - 3er trimestre

- Defensa de la tesis y presentación final.

PLAN DE FORMACIÓN PERSONAL

Declaración de intenciones: Participación en actividades formativas y régimen de publicaciones durante el Programa de Doctorado

Como parte de mi programa de doctorado en la Universidad de Salamanca, dentro del PROGRAMA DE DOCTORADO FORMACIÓN EN LA SOCIEDAD DEL CONOCIMIENTO, me comprometo a asistir y participar activamente en diversas actividades formativas durante los próximos años.

Mi objetivo es ampliar mis conocimientos, intercambiar experiencias con otros investigadores y fortalecer mis competencias académicas y profesionales. En este contexto, tengo la firme intención de participar en una amplia gama de seminarios, conferencias y cursos de formación. Estos eventos serán fundamentales para mi desarrollo en el doctorado por las siguientes razones:

- Actualización constante en mi área de investigación: La asistencia a seminarios y conferencias me permitirá estar al día de los últimos avances en los campos de la Ciencia de la Información, la Ciencia Abierta y la gestión de datos de investigación. La exposición a las últimas investigaciones me proporcionará una base sólida para integrar estos conocimientos en mi propio proyecto de tesis.

- Intercambio de conocimientos y creación de redes: Las conferencias y seminarios me darán la oportunidad de interactuar con académicos de renombre y otros investigadores que trabajan en temas similares. Estos eventos son espacios cruciales para el intercambio de ideas, lo que será inestimable para enriquecer mis planteamientos y generar nuevas perspectivas que contribuyan al desarrollo de mi investigación.

- Formación especializada: Los cursos de formación específica me permitirán adquirir competencias técnicas y metodológicas relacionadas con la gestión de datos, el uso de tecnologías digitales y las herramientas de análisis de datos en la investigación científica. Esto será especialmente importante dado el enfoque de mi investigación en el ámbito del Reglamento sobre el Espacio Europeo de Datos Sanitarios y la Ciencia de Datos.

Plan de publicaciones

En cuanto a la difusión de los resultados de mi investigación, me propongo seguir un riguroso plan de publicaciones que refleje el desarrollo progresivo de mi trabajo doctoral. Mi objetivo es publicar artículos en revistas científicas que cumplan los siguientes criterios:

- Revisión por pares y acceso abierto: las revistas a las que pretendo enviar mis artículos deben ser revisadas por pares, garantizando la calidad y el rigor de las publicaciones. Además, buscaré revistas de acceso abierto, de acuerdo con los principios de Open Science, que permitan que los resultados de mi investigación sean accesibles a la comunidad científica mundial.

- Indexación en bases de datos internacionales: tendré en cuenta las revistas indexadas en bases de datos de alto impacto, como Scopus o Web of Science, para garantizar que mi trabajo tenga una amplia visibilidad y se integre en el debate científico mundial.

- Centrarme en temas de gestión de datos, ciencia abierta y salud: me concentraré en revistas que cubran áreas específicas de la gestión de datos de investigación, ciencia abierta y la intersección entre la ciencia de la información y la salud. Esto garantizará que mi trabajo sea revisado por pares en mi campo y contribuya al avance de estas áreas.

- Colaboración interdisciplinar: Algunas de las publicaciones también se dirigirán a revistas que promuevan la colaboración interdisciplinar, dado el carácter interdisciplinar de mi investigación, que abarca las ciencias de la información y las ciencias de la salud.

Así pues, este plan de publicaciones me permitirá difundir eficazmente los resultados de mi investigación y contribuir al avance de la gestión de datos sanitarios en el marco normativo europeo.

Contribución a la comunidad científica

Además de participar en estas actividades de formación y desarrollar mi plan de publicación, también tengo la intención de presentar mi propio trabajo en conferencias relevantes. Esta experiencia me permitirá recibir valiosos comentarios y contribuir al debate científico en mi área de especialización. Así pues, estoy convencida de que participar en estas actividades será clave para enriquecer mi experiencia doctoral y contribuir al avance de las ciencias de la salud y de la información. Además, estas experiencias formativas me ayudarán a alcanzar los más altos estándares de calidad y rigor científico en mi investigación.

REFERENCIAS:

- Asociación Británica de Investigación Educativa [BERA] (2019) Guía Ética para la Investigación Educativa (4.a ed.) (L. Rivera Otero and R. Casado-Muñoz, Trads.), Londres.
<https://www.bera.ac.uk/publication/guia-etica-para-la-investigacion-educativa>.
- Carvalho, J., Furtado, F., Príncipe, P. (2017). WP4 – D25 – KIT sobre dados de investigação. RCAAP.
<https://hdl.handle.net/1822/46351>
- Chiware, E.C., Mathe, Z.M. (2015). Academic libraries' role in Research Data Management Services: a South African perspective. *South African Journal of Libraries & Information Science* 81 (2), 1–10.
<https://doi.org/10.7553/81-2-1563>.
- Comissão Europeia (2022). Proposta de Regulamento do Parlamento Europeu e do Conselho relativo ao Espaço Europeu de Dados de Saúde. Bruxelas: Comissão Europeia. <https://eur-lex.europa.eu/legal-content/PT/TXT/PDF/?uri=CELEX:52022PC0197>
- European Commission (2017). H2020 Programme: Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020. European Research Council. 2017.
- Eynden, V., Corti, L., Woollard, M., Bishop, L., & Horton, L. (2011). *Managing and sharing data : a best practice guide for researchers* (3rd ed.). U. K. Data Archive. <http://www.data-archive.ac.uk/media/2894/managingsharing.pdf>
- Ferreras-Fernández, T., Martín-Rodero, H., García-Peñalvo, F. J., & Merlo-Vega, J. A. (2016). *The Systematic Review of Literature in LIS: An approach*. In F. J. García-Peñalvo (Ed.), Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM'16) (Salamanca, Spain, November 2-4, 2016) (pp. 291-298). ACM. <https://doi.org/10.1145/3012430.3012531>.
- García Pérez, F. J. (2023). Introducción al espacio europeo de datos sanitarios: un nuevo horizonte en la gobernanza de datos sanitarios en la Unión Europea. *European Health Data Space: a new horizon in the governance of health data in the European Union*. (61), 183-196.
- García-Peñalvo, F. J. (2014). Formación en la sociedad del conocimiento, un programa de doctorado con una perspectiva interdisciplinar. *Education in the Knowledge Society*, 15(1), 4-9.
<https://doi.org/10.14201/eks.11641>.
- García-Peñalvo, F. J. (2022). Developing robust state-of-the-art reports: Systematic Literature Reviews. *Education in the Knowledge Society*, 23, Article e28600. <https://doi.org/10.14201/eks.28600>
- García-Peñalvo, F. J., Rodríguez-Conde, M. J., Verdugo-Castro, S., & García-Holgado, A. (2019). Portal del Programa de Doctorado Formación en la Sociedad del Conocimiento. Reconocida con el I Premio de Buena Práctica en Calidad en la modalidad de Gestión. In A. Durán Ayago, N. Franco Pardo, & C. Frade Martínez (Eds.), Buenas Prácticas en Calidad de la Universidad de Salamanca: Recopilación de las I Jornadas. Repositorio de buenas prácticas (pp. 39-40). Ediciones Universidad de Salamanca. <https://doi.org/10.14201/OAQ02843940>
- Hodson, S. (2009). *Meeting the research data challenge: managing research data programme 2009–11*. JISC briefing paper.
- OECD (2007). OECD Principles and Guidelines for Access to Research Data from Public Funding, OECD Publishing, Paris, <https://doi.org/10.1787/9789264034020-en-fr>.
- Surkis, A., Read, K. (2015). "Research data management." *Journal of the Medical Library Association: JMLA* 103 (3):154–156. <https://doi.org/10.3163/1536-5050.103.3.011>.
- Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. <https://doi.org/10.1038/sdata.2016.18>.