

MINISTERIO DE CIENCIA, TECNOLOGÍA E INNOVACIÓN –
MINCIENCIAS

GUÍA PARA LA GESTIÓN DE DATOS DE INVESTIGACIÓN DEL MINISTERIO DE CIENCIA, TECNOLOGÍA E INNOVACIÓN

TABLA DE CONTENIDO

INTRODUCCIÓN	3
1 GESTIÓN DE DATOS DE INVESTIGACIÓN: DEFINICIONES Y CONCEPTOS BÁSICOS	3
1.1 Aclarando el concepto de dato (aplicado en el ámbito digital)	3
1.2 Big data y la explosión de los datos	4
1.3 De los datos a los Datos de Investigación.....	4
1.4 DataSets de Investigación: Objetos digitales compuestos y heterogéneos	6
1.5 Gestión de datos de investigación - GDI.....	8
1.6 Ciclo de vida de los datos de investigación	8
2 PRINCIPIOS FAIR (JUSTOS) Y RECOMENDACIONES DE APLICACIÓN	14
2.1 Principio 1. Datos Encontrables (Findable)	14
2.2 Principio 2. Datos Accesibles (Accessible).....	15
2.3 Principio 3. Datos Interoperables (Interoperable).....	16
2.4 Principio 4. Datos Reutilizables (Reusable).....	16
3 DISEÑO DE UN PLAN DE GESTIÓN DE DATOS DE INVESTIGACIÓN (PGD)	17
3.1 Herramientas para construir PGD	17
3.2 Preparar el Plan de Gestión de Datos	18
3.3 Descripción de los DataSets	18
3.4 Formato de procesamiento de los datos	18
3.5 Estándares de metadatos	19
3.6 Compartir datos	19
3.7 Acceso de los datos	19
3.8 Licencias de uso y atribución	19
3.9 Archivo y preservación	19
3.10 Campos de información del PGD	20
3.11 Plantilla Plan de Gestión de Datos.....	20
4 GLOSARIO DE TÉRMINOS:	24
5 BIBLIOGRAFÍA	26

INTRODUCCIÓN

Con el advenimiento de las tecnologías de Información y comunicación, que se encuentran potenciadas en la era de la explosión y abundancia de los datos comúnmente llamado Big Data, se establece la gestión de conocimiento científico como un elemento fundamental de la Ciencia Abierta.

Es así, como los sistemas de información académicos y científicos (portales de revista), deben asegurar a su comunidad de usuarios los recursos necesarios para el adecuado acceso confiable y seguro a sus fuentes documentales (recursos de información). Esto implica que deben velar por la calidad de la información (datos, metadatos y documentos) que ofrecen, buscando en todo momento que estos se encuentren alineados a las principales normas e iniciativas nacionales e internacionales y que simultáneamente estén ajustados a las nuevas tendencias globales de información académica y científica.

Para lograr lo anterior, la gestión de datos de investigación se advierte como un elemento indispensable a considerar de manera prioritaria por los investigadores en el desarrollo de sus proyectos de investigación para que, mediante el uso de licencias flexibles y preferentes, los datos se encuentren disponibles y se permita su uso y acceso libre, sin restricciones. De tal forma que sea posible verificar, evaluar, reproducir, distribuir, reutilizar y regenerar los datos y sus métodos subyacentes para el desarrollo de nuevos proyectos de investigación que contribuyan al avance rápido y óptimo en el descubrimiento de soluciones a las problemáticas que aquejan a nuestra sociedad.

Para lograr una adecuada gestión de los Datos de Investigación, el presente documento plantea los elementos básicos para comprender qué son los datos de investigación, cómo se generan, gestionan, utilizan, entre otros aspectos.

1 GESTIÓN DE DATOS DE INVESTIGACIÓN: DEFINICIONES Y CONCEPTOS BÁSICOS

1.1 Aclarando el concepto de dato (aplicado en el ámbito digital)

Desde un enfoque constructivista, se entienden los datos como la representación simbólica (numérica, alfabética, etc.), de un atributo o característica de una entidad (ser) que se manifiesta a través de un hecho o un proceso (Davenport & Prusak, 1998). Dicho de otra manera, los datos permiten representar un estado de la realidad asociado a un momento (tiempo) a través de la codificación (símbolos pertenecientes a un lenguaje) de esta realidad en un medio (soporte y formato) que puede ser entendido, utilizado, compartido y transformado tanto por un humano como por una máquina (Hardware y Software).

Los datos son infinitos, estos permiten representar hechos empíricos, sucesos y entidades, que se pueden generar en cualquier momento y lugar. Los datos se pueden ver como la unidad primaria y mínima de representación, los cuales por sí solos son irrelevantes y no suelen indicar nada sobre el porqué de las cosas. Un dato no tiene valor semántico (sentido) en sí mismo, solamente tienen valor sintáctico. Por ello, son óptimos para ser procesados en la realización de cálculos o en la toma de decisiones.

Un dato puede ser cualquier cosa, dependiendo de la disciplina científica, los datos se presentan desde múltiples orígenes y en distintos medios. Generalmente se asocian con elementos cuantitativos (finitos e infinitos dependiendo de la escala de medición) como: una temperatura, una representación horaria, una talla, una medida, un nombre, un número de teléfono, una zona geográfica, un color, una presión atmosférica, una dirección, entre otros. Sin embargo, en el ámbito más amplio puede integrar valores cualitativos que se puedan representar como una emoción, una idea, un sentimiento, un aroma, una expresión facial, una escala vocal, un post en una red social (Twitter, Facebook, etc.), una huella biométrica, entre otros.

Vivimos en un mundo rodeado por los datos. Los datos pueden ser tan simples o complejos tanto como se expresa digitalmente la representación de este. Los datos pueden ser abstractos como un número o un texto (alfanumérico),

pero también pueden representar elementos más amplios como una fotografía, un vídeo, un audio, una partitura, un audio/sonido, etc.

La información (datos organizados con sentido y utilidad para una entidad) se producen cuando diversos conjuntos de datos se relacionan entre sí a través de distintos componentes. “Los datos se convierten en información cuando son combinados entre ellos según un método que tenga el potencial de revelar los patrones del fenómeno estudiado” (Peset Mancebo & González, 2017).

1.2 Big data y la explosión de los datos

Big data es un término de origen inglés cuya traducción equivale a “datos masivos”. El concepto de Big Data se está utilizando actualmente a partir de la diversificación de los medios de recopilación y procesamiento de múltiples conjuntos de datos en el ámbito digital que son potenciadas por la creciente expansión de las tecnologías de la información y comunicación - TIC. Se está experimentando una evolución constante de las tecnologías, para el análisis, la explotación y el tratamiento de datos (cada día se están generando y recopilando nuevos tipos de datos tanto en el sector público como privado), las cuales propician la creación de nuevos servicios y aplicaciones, que, a su vez, promueven nuevas oportunidades económicas y sociales. Así mismo, impulsa el desarrollo de nuevas tecnologías como el aprendizaje automático, la inteligencia artificial y el internet de las cosas.

Una de las definiciones clásicas de Big Data que fue acuñada por la consultora McKinsey Global Institute, quien pone de manifiesto el uso avanzado de las nuevas tecnologías; el instituto indica que, “Big Data se refiere a los conjuntos de datos cuyo tamaño está más allá de las capacidades de las herramientas típicas de software de bases de datos para capturar, almacenar, gestionar y analizar” (Manyika et al., 2011). Esta definición establece que el Big Data va mucho más allá del uso masivo de los tradicionales motores de bases de datos relacionales, incorporando nuevas tecnologías, herramientas y modelos que permiten una adecuada gestión de grandes volúmenes de datos.

Otra definición ampliamente utilizada, se define como: “grandes conjuntos de datos que poseen tres características principales: volumen (cantidad), velocidad (velocidad de creación y utilización) y variedad (tipos de fuentes de datos no estructurados; tales como la interacción social, video, audio, cualquier cosa que se pueda clasificar en una base de datos)” (Prentice, S. 2011). Así mismo, otros autores definen un concepto más amplio de Big Data que incluye: “conjunto de arquitecturas y herramientas informáticas destinadas a la manipulación, gestión y análisis de grandes volúmenes de datos, desde todo tipo de fuentes diseñadas para extraer valor y beneficio de estos, con una amplia variedad en su naturaleza, mediante procesos que permitan capturar, descubrir y analizar información a alta velocidad y con un coste reducido” (García Herrero et al., 2018).

1.3 De los datos a los Datos de Investigación

Los datos de investigación (Data Research), son aquellos productos (recursos documentales) generados o recopilados durante el desarrollo de un proyecto de investigación; los cuales se convierten en evidencia temprana del proceso de investigación y que sirven para validar, verificar, reproducir o certificar los productos resultados finales de dicha investigación y que sólo adquieren significado en el contexto de esta.

Otras definiciones incluyen:

- “Los datos de investigación, son todo aquel material que sirve para certificar los resultados de la investigación que se realiza, que se han registrado durante ella y que se ha reconocido por la comunidad científica” (Peset Mancebo & González, 2017b).
- “Son todo aquel material que ha sido registrado durante la investigación, reconocido por la comunidad científica y que sirve para certificar los resultados de la investigación que se realiza” (Torres-Salinas et al., 2012).

- “Datos que se utilizan como fuentes primarias para respaldar la investigación técnica o científica, la investigación, las becas o la actividad artística, y que se utilizan como evidencia en el proceso de investigación y / o son comúnmente aceptados en la comunidad de investigadores como necesarios para validar los hallazgos y resultados de la investigación. Los datos de investigación pueden ser datos experimentales, datos de observación, datos operativos, datos de terceros, datos del sector público, datos de seguimiento, datos procesados o datos reutilizados” (Consortia Advancing Standards in Research Administration Information, CASRAI, 2021).

Los datos de investigación son dinámicos y van cambiando a lo largo de la investigación, por tanto, se deben gestionar durante todo el ciclo de vida del proyecto de investigación y pueden ser representados de manera alfabética, numérica o audiovisual, dependiendo de la disciplina y de la naturaleza de la investigación.

“Ejemplos de datos de investigación son los cuestionarios, los registros sonoros o visuales (fotografías, películas, videos, diapositivas) y todo tipo de imágenes, registros y medidas de cualquier tipo de prueba o experimento, muestras, modelos, cuadernos de laboratorio y de campo, series numéricas en tablas, registros de ordenador, estadísticas” (Aleixandre-Benavent et al., 2021). Además, para su correcta comprensión deben estar bien documentados e incluir los metadatos que los describan.

Los principales beneficios de los datos de investigación abiertos son, según la Biblioteca de la Universidad de Alcalá Biblioteca, (2021):

- Refuerza la investigación científica, gracias a la reutilización de investigaciones anteriores.
- Promueve y mejora la investigación y la comprobación de métodos nuevos o alternativos, y alienta la diversidad de análisis y opiniones.
- Mejora métodos de recogida y medición de datos, a través del escrutinio de los demás.
- Garantiza la transparencia e integridad de la investigación.
- Reduce los costes evitando los esfuerzos de recolección de datos duplicados.
- Permite cumplir con los requisitos de los organismos financiadores.
- Integración de los datos en las redes de repositorios a nivel internacional.
- Hace visible la inversión pública en investigación.
- Mejora el impacto y la visibilidad de un investigador y de su proyecto.

Existen diferentes modelos de clasificación de los datos de investigación, en general distintos autores proponen clasificarlos en función de su origen, su estatus/versión o de la metodología aplicada para su uso; tal como (Grupo de Trabajo de Depósito y Gestión de datos en Acceso Abierto del proyecto RECOLECTA, 2012) y Borgman, 2012; National Science Board, 2005), quienes propone los siguientes:

Clasificación de los datos de investigación según su metodología de recolección o generación:

- **Datos Experimentales:** Son datos obtenidos a partir de la realización de un experimento (desde su planificación y preparación hasta la obtención de resultados). Son datos que, estando bien documentados, son normalmente reproducibles.
- **Datos observacionales:** Corresponde con registros históricos de un fenómeno ocurrido en el tiempo. Se pueden obtener únicamente en un lugar y en un momento en el tiempo (son irrepetibles).
- **Datos de simulación:** Son aquellos datos generados durante la aplicación de modelos predefinidos (algoritmos, modelos matemáticos, o simulaciones de experimentos) y ejecutados en ambientes controlados de prueba.
- **Datos computacionales, compilados y derivados** (Melero & Hernández-San-Miguel, 2014): Son datos creados al tomar datos ya existentes y realizar algún tipo de manipulación sobre los mismos mediante la utilización

de herramientas computacionales especializadas y obtener resultados reproducibles con soporte tecnológico.

- Datos canónicos o de referencia: Conjuntos de datos tomados de fuentes predefinidas que pueden ser utilizados para validación, comparación o búsqueda de información.

Clasificación de los datos de investigación según el objetivo de su recogida (Swan & Brown, 2008):

- Datos Específicos: Datos generados para un proyecto de investigación específico y cuyos datos pueden tener poco interés fuera de dicho proyecto.
- Datos de alcance medio: Datos que pueden proveer información de interés o ser de referencia para una disciplina o grupo de disciplinas, es decir, los de los que puede beneficiarse una comunidad muy limitada de investigadores pertenecientes a una determinada especialidad.
- Datos de interés general: Son datos cuya importancia es vital para el desarrollo de la ciencia y que pueden constituir una referencia canónica en la comunidad científica.

Clasificación de los datos de investigación según el tratamiento (nivel de procesamiento) que hayan recibido (Green, 2009):

- Datos primarios (preliminares): Datos originales que han sido creados o recolectados pero que no han sido procesados o analizados ni han recibido ningún tipo de tratamiento por parte del investigador (raw data).
- Datos procesados: Datos que han sido tabulados, digitalizados, traducidos, transcritos, limpiados, validados, verificados y/o anonimizados.
- Datos finales: Corresponde con el conjunto de datos seleccionados o filtrado del conjunto de datos procesados que finalmente se utilizan para validar los resultados.
- Datos derivados (analizados): Son datos transformados y listos para su uso y consulta que han sido creados a partir de los datos finales, y que mediante su transformación (Modelos, gráficos, tablas, textos, entre otros), se pretende sean de ayuda en el descubrimiento de información útil, la presentación de conclusiones y la toma de decisiones.

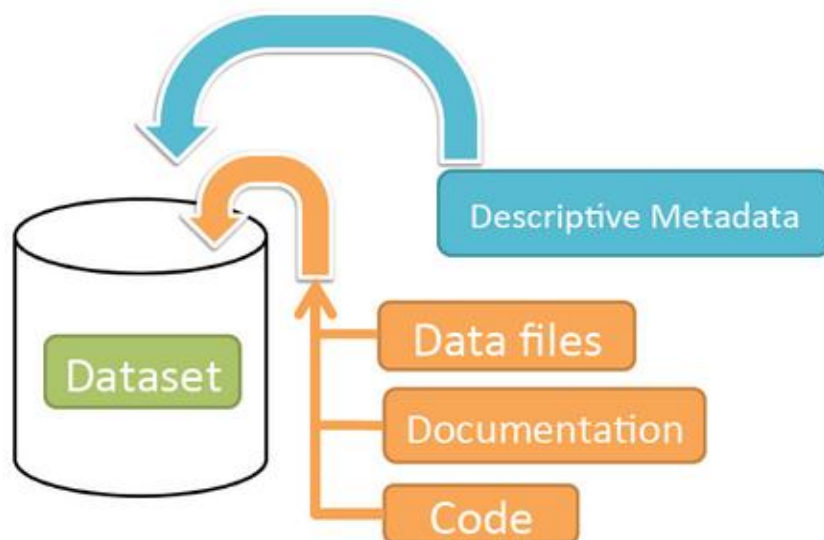
1.4 DataSets de Investigación: Objetos digitales compuestos y heterogéneos

La literatura científica comúnmente define a los DataSets como conjunto de datos de investigación generados durante la ejecución de un proyecto de investigación. Estos DataSets son los productos resultantes de la aplicación de algún instrumento de investigación (encuestas, entrevistas, fichas de observación, cuaderno de notas, diarios de campo, cuestionarios, cámara fotográfica, etc.). Generalmente un proyecto de investigación puede crear tantos DataSets como instrumentos aplique, entendiendo que su gestión se puede hacer de manera independiente o mezclada con otros DataSets.

Sin embargo, cuando se asocia la gestión de los datos de investigación, con la implementación de sistemas de información especializados para la gestión de estos, el concepto y aplicación de un DataSets se amplía radicalmente. Se denomina DataSets de investigación, a un objeto digital compuesto y heterogéneo, el cual se representa como un contenedor que agrupa: DataFiles, Documentación, Metadatos y Código Fuente (Ver Figura 1.). “Los DataSets se almacenan y gestionan en repositorios interoperables en red, integrados en una infraestructura global de investigación, desarrollados conforme a estándares internacionales” (Grupo de Trabajo de “Depósito y Gestión de datos en Acceso Abierto” del proyecto RECOLECTA, 2012b):

Figura 1.
Representación de un DataSets en un Sistema de información

Schematic Diagram of a Dataset in Dataverse 4.0



Container for your data, documentation, and code.

Nota: Tomado de (Institute for Quantitative Social Science, 2021). <https://guides.dataverse.org/en/latest/user/DataSets-management.html>

- DataFiles: Uno o varios archivos que contienen un conjunto de datos organizados según una estructura y almacenados en un formato/extensión.
- Documentación: Uno o varios archivos, generalmente creados a partir de una plantilla predefinida, que contienen documentación asociada a los datos (Conjunto de reglas que describen qué contienen los conjuntos de datos, su procesamiento, su origen, cómo se visualizan y manejan adecuadamente). La documentación de datos garantizará que sus datos sean comprendidos e interpretados por cualquier usuario. Explicará cómo se crearon sus datos, cuál es el contexto para los datos, la estructura de los datos y su contenido, y cualquier manipulación que se haya hecho a los datos. Se recomienda que la documentación que acompaña a los DataSets incluya los siguientes aspectos generales:
 - Contexto de la recopilación de datos
 - Metodología de recolección de datos
 - Estructura y organización de archivos de datos
 - Validación de datos y aseguramiento de la calidad
 - Manipulaciones de datos a través del análisis de datos a partir de datos sin procesar
 - Licencias y condiciones de confidencialidad, acceso y uso de los datos.

Así mismo, en la documentación asociada para cada archivo que contenga conjunto de datos específicos y relacionados, se recomienda incluir los siguientes aspectos:

- Nombres y descripciones de variables utilizadas
- Definición de códigos y esquemas de clasificación
- Códigos asociados, aclaraciones y procesamientos de los valores perdidos y nulos
- Definiciones de terminología y acrónimos de especialidad

- Algoritmos utilizados para transformar datos
- Formato de archivos y software utilizado
- Metadatos: El cual se entiende como información “Estructurada o Semiestructurada” (un conjunto de campos o clases) que describe, explica, localiza, administra y amplía la información acerca del conjunto de datos de forma que tenga algún sentido para las máquinas y los humanos. Adicionalmente como un elemento obligatorio asociado a los metadatos, se debe incluir el uso de identificadores persistentes (DOI, HANDLE, ORCID, ROR, ISNI, etc.)
- Código: Uno o varios archivos o scripts que contiene código asociado a un lenguaje de programación/manipulación de datos (Generalmente Stata, R, MATLAB o Python), los cuales permiten gestionar los aspectos de reproducibilidad y reutilización de los datos.

1.5 Gestión de datos de investigación - GDI

La Gestión de datos de investigación (GDI) o Research Data Management (RDM) es una práctica importante tanto para las instituciones como para los investigadores que hace parte integral en cualquier proceso de investigación, el cual engloba todos los aspectos de adquisición, organización, almacenamiento, procesamiento, transformación, visualización, uso, licenciamiento y publicación de los datos utilizados o generados durante un proyecto de investigación (gestión de los datos a lo largo de todas las fases del proyecto de investigación). “La gestión de datos de investigación no es una meta en sí misma, sino la serie de pasos y requisitos que permiten la reutilización de datos y la generación de nuevos conocimientos” (Actis & Carlino, 2017).

En resumen, la Gestión de Datos de Investigación (GDI), busca proveer los elementos, buenas prácticas y herramientas que permiten a los investigadores gestionar todo el ciclo de vida de los datos de investigación generados o adquiridos, ayudando a garantizar la calidad de la investigación. Según la Universidad de Alcalá (2021), una adecuada gestión de datos de investigación genera los siguientes beneficios:

- Mayor transparencia para la validación de los resultados de la investigación.
- Asegura que los datos sean localizables, accesibles, interoperables y reutilizables.
- Mejorar el perfil del investigador, el impacto y la visibilidad de los proyectos.
- Mejorar la protección de datos y minimizar el riesgo de pérdida de estos.
- Cumplir con los requerimientos de los organismos financiadores.
- Ahorrar tiempo y hacer uso eficiente de los recursos disponibles.

1.6 Ciclo de vida de los datos de investigación

Como se mencionó anteriormente, durante la ejecución de un proyecto de investigación, se debe establecer un plan de trabajo que permita realizar una adecuada Gestión de datos de investigación (GDI). Este plan de trabajo establece las distintas actividades y tareas asociadas durante todo el ciclo de vida de los datos de investigación (desde su creación o captura y almacenamiento, pasando por su uso en distintas actividades de procesamiento, hasta que éstos se vuelven obsoletos y se eliminan), que se debe realizar y se formalizan en la construcción de un documento (generalmente basado en una plantilla) llamado “Plan de Gestión de Datos de investigación” (PGD).

La Gestión de datos de investigación se desarrolla incluso desde antes del inicio de un proyecto de investigación y permite que los datos de investigación tengan una vida útil que trascienda más allá del proyecto de investigación de quien los crea. La misma dinámica de los datos, permiten que estos se vayan transformando a lo largo de la investigación, esta continua transformación se conoce como el “Ciclo de vida de los datos de investigación”. Los investigadores pueden continuar trabajando con distintos datos de investigación mucho después de que haya cesado

el proyecto que los generó, así mismo, los proyectos de investigación derivados también pueden reutilizar y ampliar datos de investigación previamente creados en otros proyectos. “Los datos bien organizados, bien documentados, preservados y compartidos son invaluable para promover la investigación científica y aumentar las oportunidades de aprendizaje e innovación” (UK Research and Innovation, UKRI, 2021).

El Ciclo de vida de los datos de investigación proporciona una estructura que describe cada una de las fases (proceso sistemático dividido en etapas), que se deben tener en cuenta durante la gestión de datos de investigación.

Existen diferentes modelos que representan el ciclo de vida de los datos de investigación los cuales proporcionan una base que define las muchas operaciones que deberán llevarse a cabo durante la gestión de datos de investigación. Algunos de los principales modelos que representan el ciclo de vida de los datos de investigación se muestran a continuación:

El modelo propuesto por Digital Curation Centre (DCC) del Reino Unido (Higgins, 2008), establece un ecosistema de fases y relaciones asociadas con el ciclo de vida de los datos de investigación y las tareas asociadas a los procesos de curación de estos (Curation Lifecycle Model). Este modelo está orientado principalmente a los profesionales de la información, que participan en los distintos roles de apoyo a los investigadores y que estarían a cargo de la curaduría y preservación de los datos. El modelo incluye cuatro acciones persistentes durante todo el proyecto:

- Describir/Representar DataSets,
- Planificar la preservación
- Integrar a la comunidad en Participación/Vigilancia/Monitoreo,
- Realizar Preservación Digital/Curación de Datos)

Además, incluye ocho acciones secuenciales continuas:

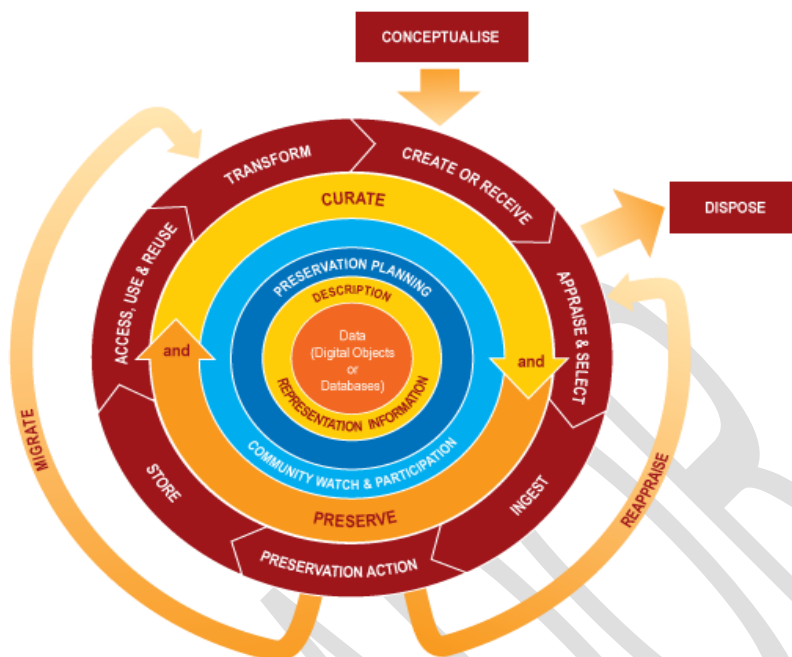
- Conceptualizar/Modelar
- Crear/Adquirir
- Valorar/Seleccionar
- Consumir
- Preservar
- Almacenar
- Acceder/Utilizar/Reutilizar,
- Transformar

Y tres acciones ocasionales:

- Disponer/Publicar
- Reevaluar
- Migrar

“El modelo expuesto por el DCC proporciona las etapas requeridas para la curaduría y la preservación de los datos desde la conceptualización inicial y puede ser utilizado para planificar las actividades de un proyecto de investigación para asegurar que se llevan a cabo todas las etapas necesarias, cada una en la secuencia correcta. Permite definir roles y responsabilidades, y construir un marco normativo. En el centro del modelo se encuentran los datos digitales, que están identificados con objetos o bases de datos simples y complejos. Las relaciones entre las etapas del ciclo de vida presentadas por el modelo señalan los principales niveles de acciones sobre la curaduría de los datos” (Couto Corrêa, 2016).

Figura 2
Modelo de Ciclo de Vida propuesto por el Digital Curation Centre - DCC



Nota: Tomado de (Curation Lifecycle Model). <https://www.dcc.ac.uk/guidance/curation-lifecycle-model>

La organización **DataONE** (Data Observation Network for Earth) propone un modelo de ciclo de vida de los datos de investigación (Data life cycle model) (Strasser et al., 2012):

Figura 3
Modelo de Ciclo de Vida propuesto por DataONE



Nota: Tomado de (Data Observation Network for Earth). https://dataoneorg.github.io/Education/bp_step/analyze/

El modelo está orientado principalmente para ser ejecutado desde la perspectiva de los investigadores y se desarrolló como una guía para fortalecer las competencias en los investigadores sobre las mejores prácticas que pueden seguir en la gestión de sus datos. El modelo propuesto por la Biblioteca de la Universidad de Valladolid (2021), define ocho fases para la gestión de datos:

- Planificar: Establece un Plan de Gestión de datos (Data Management Plan). Describir los datos que se recopilarán y cómo se gestionarán los datos y se harán accesibles a lo largo de su vida. Un PGD ayuda a planificar el proceso de investigación y a definir responsabilidades en un proyecto, en el que participan varios investigadores.
- Recopilar: En el proceso de recolección de datos se seguirá lo establecido en el Plan de Gestión de Datos. El resultado final de este proceso de decisión es un modelo de datos que describe la forma en que se estructuran los datos.
- Asegurar: Emplear procedimientos que aseguren la calidad de los datos.
- Describir: los datos se describen de forma precisa y exhaustiva utilizando los estándares de metadatos adecuados. Esto nos permitirá contar luego con información para comprender, utilizar y reutilizar.
- Conservar: los datos se envían a un archivo apropiado a largo plazo. También, es importante definir los repositorios donde se van a almacenar tus datos. Se debe tener en cuenta considerar aspectos como la privacidad, el derecho de autor y el licenciamiento.
- Descubrir: se localizan y obtienen datos potencialmente útiles para la investigación, junto con la información relevante sobre los datos (metadatos)
- Integrar: los datos de fuentes dispares se combinan para formar un conjunto homogéneo de datos que se pueden analizar fácilmente
- Analizar: se analizan los datos.

El modelo del ciclo de los datos científicos desarrollado por la Red de Bibliotecas Universitarias Españolas (REBIUN) en su estrategia LÍNEA 2, indica que, “la transformación Digital y Conocimiento Abierto, está basado en la propuesta del ciclo de vida de la gestión de datos de investigación del según el Joint Information Systems Committee (JISC), (2018) Este modelo construido mediante una infografía incluye once (11) fases agrupado en cuatro áreas de trabajo que contemplan las distintas acciones que se deben realizar desde las dos perspectivas: investigadores y profesionales de la información:

- Área 1. Establecer el plan de gestión de datos: Busca identificar los elementos estratégicos que compondrán el documento del Plan de Gestión de Datos de investigación - PGD (Data Management Plan) y realizar las tareas de adquisición de datos.
- Área 2: Asignar metadatos: Este escenario busca pasar de los datos de investigación en bruto a DataSets de Investigación (Objetos digitales compuestos y heterogéneos) que incluyen metadatos para su correcta gestión
- Área 3: Estándares, Almacenamiento y copias de seguridad: Busca aplicar los distintos estándares y buenas prácticas que permitan almacenar y preservar los datos y DataSets para que puedan ser reutilizados en cualquier momento y lugar.
- Área 4: Acceso y utilización, licencias y uso ético. Incluye los procesos asociados al licenciamiento, la aplicación de métodos y modelos de análisis y visualización de datos con el fin de obtener información relevante y pertinente en el marco del proyecto investigación.

Figura 4
Modelo de ciclo de vida de los datos científicos propuesto por REBIUN



Nota: Tomado de <https://rebiun.xercode.es/xmlui/handle/20.500.11967/69>

El modelo expone de manera explícita, el desarrollo de un trabajo colaborativo permanente durante todo el desarrollo del proyecto de investigación y da especial importancia a la presencia de un equipo especializado en datos de investigación que debería tener la unidad de información de cualquier organización. Las fases de este modelo se explican a continuación:

- **Diseñar:** Esta fase busca establecer el alcance, objetivos, tiempos, metodologías de recolección, acuerdos de intercambio y controles de acceso relacionados con la gestión de datos de investigación en el marco de las políticas, los servicios y los requisitos de la organización, el ente financiador y los organismos nacionales/internacionales públicos y privado (el diseño de un buen plan de gestión de datos debe reflejar el contexto del entorno).
- **Planificar:** La planificación de la gestión de datos de investigación es la base de cualquier proyecto. Esta fase busca establecer de manera temprana (de un proceso en constante cambio), qué datos y DataSets se van a recopilar, procesar o generar durante el proyecto de investigación. Una buena planeación ayuda a entender cómo organizar, almacenar y conservar los datos de una manera JUSTA (FAIR) (localizable, accesible, interoperable y reutilizable). El Diseño y la planificación permite generar un documento denominado "Plan de gestión de datos de investigación - PGD"
- **Crear, Recoger, Recolectar:** Esta fase se inicia una vez se ha generado un Plan de gestión de datos de investigación - PGD y se ha iniciado el proyecto de investigación. Aquí, se concentran las tareas de adquisición de datos que implica obtener datos existentes o generar nuevos datos para luego recopilarlos y organizarlos como datos de investigación en bruto (raw data). No todos los datos que se proyectan en el plan de gestión de datos de investigación - PGD, se crean/recopilan o se utilizan, depende de los investigadores y del equipo responsable por la gestión de datos; identificar qué datos deben capturarse y los mejores medios para hacerlo.

- **Describir:** Realizar una adecuada gestión de metadatos es uno de los aspectos más importantes de la gestión de datos de investigación. Existen múltiples modelos de metadatos que incluyen campos especializados que pueden ser aplicados simultáneamente (como modelos de metadatos mixtos), a los diversos conjuntos de datos dependiendo de la representación de la naturaleza del contenido, y el área o disciplina en la cual fueron generados. Independientemente de estos campos de metadatos especializados, se recomienda utilizar como elementos base de descripción el esquema de metadatos propuesto por DataCite (2019) que también es utilizado por las directrices OpenAIRE (2015), para repositorios de datos.
- **Procesar:** En esta fase la preparación se orienta a la adecuación de los datos en bruto para que puedan ser seleccionados/transformados y utilizados posteriormente. El procesamiento de datos incluye generalmente las siguientes tareas:
 - Tareas de limpieza de datos las cuales permiten restringir datos, eliminar ruido y datos no válidos.
 - Tareas de normalización de datos que corrige el despliegue de los datos en un mismo rango.
 - Tareas de discretización las cuales realizan conversión de variables continuas a categóricas con el objetivo de que haya un número limitado de valores posibles.
 - Tareas de cambio de dimensionalidad que permiten ampliar o reducir la base de los datos a utilizar.
- **Archivar:** Como un proceso de conservación de la evidencia del proceso de investigación, se deben seleccionar los distintos datos de investigación para su depósito en sistemas de almacenamiento de datos que incluye, desde la escritura en distintos formatos de archivo organizado en directorios, la incorporación de los datos en una base de datos o en un sistema distribuido de datos o lago de datos. A su vez, estos datos deberán incorporarse en un repositorio de datos que permitan su transformación en DataSets organizados, y de esta forma incrementar su uso, preservación y reutilización.
- **Seleccionar:** En esta fase se incorporan los procesos de consolidación de variables de datos (variables nominales, ordinales y métricas) presentes en los conjuntos de DataSets, con el fin de realizar una tarea de filtrado (selección) de los datos y DataSets más relevantes, que serán utilizados durante la investigación del conjunto total generado/recopilado y también de aquellos que serán publicados y compartidos en el futuro. Así mismo, en esta fase se incluyen las tareas de anonimización de los datos sensibles presentes de acuerdo con las distintas políticas de protección de datos aplicables.
- **Preservar:** La preservación de datos tiene como objetivo, mantener seguros y accesibles a largo plazo los DataSets (datos, metadatos y contexto), por tanto, se busca implementar las mejores prácticas para que los datos y DataSets se puedan recuperar y reutilizar más allá del tiempo de ejecución del proyecto de investigación que los creó. Se incluye la selección de formatos de archivo estandarizados (ISO, NISO), copias de seguridad, cifrado y verificación de firmas de archivos (HASH), migración entre sistemas, formatos y soportes, entre otros.
- **Acceder, Usar:** Esta fase incluye los procesos de acceso a los DataSets autorizados (incorporación de tablas y niveles de acceso para acceder, copiar o editar), de transformación y de visualización mediante el uso de herramientas especializadas que facilitan el análisis e interpretación de los datos de investigación que permitan generar información relevante y pertinente para los investigadores. Adicionalmente, se incluyen el uso de las técnicas (minería de datos, Inteligencia de negocios -BI-, etc.) y modelos de manipulación de datos masivos (modelos descriptivos predictivos y prescriptivos) que mejoran la consolidación del conocimiento descubierto.
- **Compartir:** En esta fase se incorporan a los DataSets, una o varias licencias de acceso y uso a los datos de investigación, previo permiso otorgado por los investigadores a las instituciones y organizaciones que publicarán y compartirán estos recursos en sus respectivos sistemas de información. Adicionalmente se deben incorporar las distintas tareas asociadas al mercadeo y difusión de los contenidos en los distintos buscadores, cosechadores y recolectores de datos especializados y distintas redes sociales con el fin de

incrementar su buscabilidad, su consulta, citación e impacto. En el marco de la Ciencia Abierta, se recomienda la publicación de los datos de investigación en cualquiera de los siguientes medios:

- Repositorios de datos de investigación
 - Revistas de datos (artículos de datos)
 - Material complementario a las publicaciones (productos resultados de proyectos de investigación)
- Reutilizar: Esta fase incluye las guías y protocolos necesarios para que los datos de investigación se puedan descargar, abrir (por un sistema informático), entender su estructura, contexto, formas de interpretación y visualización, que permitan su reutilización para verificar y contrastar los resultados finales de proyecto que los originó, así como servir de base para nuevos proyectos de investigación.

2 PRINCIPIOS FAIR (JUSTOS) Y RECOMENDACIONES DE APLICACIÓN

FAIR, es un acrónimo acuñado por la Force11 (una comunidad de académicos, bibliotecarios, archivistas, editores y financiadores de la investigación que han surgido orgánicamente para ayudar a facilitar el cambio hacia la creación y el intercambio de conocimientos) (Peset et al., 2017), que es generado a partir de la combinación de cuatro principios esenciales que se deben tener en cuenta durante la gestión de datos de investigación (Findable, Accesible, Interoperable and Reusable):

“Los principios enfatizan la capacidad de acción de la máquina (es decir, la capacidad de los sistemas computacionales para encontrar, acceder, interoperar y reutilizar datos con ninguna o mínima intervención humana) porque los humanos dependen cada vez más del soporte computacional para manejar los datos como resultado del aumento en el volumen, complejidad y velocidad de creación de datos” (Wilkinson et al., 2016).

FAIR, propone un conjunto de principios esenciales que se deben tener en cuenta durante la gestión de datos de investigación (Findable, Accesible, Interoperable and Reusable). Cada principio incluye un conjunto de recomendaciones asociadas que son ampliamente explicadas en el portal web de la Iniciativa GO FAIR (<https://www.go-fair.org/fair-principles/>) (2021).

“El ecosistema de datos FAIR está compuesto por: las políticas que regulan y definen los datos, los investigadores que los producen o utilizan, los planes de gestión de datos, los identificadores, los estándares, los repositorios de confianza y los servicios en la nube donde se almacenan los datos. A su vez, estos componentes deben desarrollarse en un marco proactivo de cuatro elementos clave: las capacidades o competencias (skills), las métricas (metrics), el sistema de reconocimiento (rewards) y la inversión (investment)” (Melero, 2018b).

Los cuatro principios FAIR y sus recomendaciones asociadas se presentan a continuación:

2.1 Principio 1. Datos Encontrables (Findable)

- Recomendación F1 (GO FAIR, 2021a): Asignar (Utilizar) identificadores globales, únicos y persistentes a los datos y metadatos. Los identificadores globales únicos y persistentes PID (por sus siglas en inglés) surgen como una alternativa para solucionar la ambigüedad y localización de las distintas entidades asociadas a los ecosistemas de investigación (investigadores, datos, documentos/productos, instituciones, financiadores/patrocinadores, etc.). El uso de los identificadores persistentes permite eliminar la ambigüedad en el significado de los datos publicados al asignar un identificador único a cada DataSets y sus metadatos asociados (incluyendo la descripción de otros PID en los metadatos). Posibilita su localización, aunque estos hayan cambiado de ubicación en la red (cambio de URL). A su vez, permite garantizar la citación correcta de los datos, facilitar la interconexión e interoperabilidad entre sistemas y la detección de duplicidad de datos y fuentes de publicación. Se estipulan dos condiciones para su construcción:

- Debe ser globalmente único: (No se puede reutilizar para otros recursos). Se puede obtener identificadores globales y únicos mediante la suscripción de algún servicio de registro de identificadores, libre de barreras geográficas, que garantice la unicidad de los identificadores (DOI DataCite, DOI Crossref, Handle, PURL, ORCID, ROR, ISNI, etc.)
 - Debe ser persistente: Los identificadores deben trascender el tiempo, deben ser permanentes con el fin de garantizar la re-solubilidad de los identificadores en el futuro y poder seguir el enlace al recurso fuente.
- Recomendación F2 (GO FAIR, 2021b): Los metadatos que acompañan a los datos de investigación deben ser generosos y extensos. Estos deben contener campos de carácter descriptivo (descubrimiento del contexto, calidad y condición para identificar y describir el recurso), administrativo (manejo y procesamiento del recurso) y estructural (exploración, composición y representación del recurso). Adicionalmente, los datos se deben describir con metadatos enriquecidos que puedan ser procesables por máquinas, es decir, agregando declaraciones semánticas, vocabularios controlados, anotaciones, etc. Algunos de los modelos de metadatos más utilizados para describir datos de investigación son: DataCite Metadata Schema, Dublin Core Metadata Initiative y OpenAIRE Guidelines for Data Archives.
 - Recomendación F3 (GO FAIR, 2021c): Los metadatos incluyen clara y explícitamente el identificador de los datos que describen. La asociación de un conjunto de datos y sus metadatos asociados debe hacerse explícita mencionando el identificador persistente en todos los contextos, por ejemplo, al registrar e indexar datos y metadatos en un sistema de información (repositorio de datos, revista de datos, servicio de datos, etc.).
 - Recomendación F4 (GO FAIR, 2021d): Los datos y metadatos se registran o indexan en un recurso de búsqueda. Se recomienda proveer los mecanismos de registro y mercadeo de los datos de investigación para que sean visibles (indexados, navegables, buscables y recuperables) en repositorios, directorios, bancos de datos, buscadores y recolectores tanto generales como especializados en datos de investigación

2.2 Principio 2. Datos Accesibles (Accessible)

Busca que los datos, una vez localizados mediante el uso de herramientas y protocolos abiertos, se pueda acceder directa y libremente a sus metadatos para comprender su contenido, enfoque y alcance. Así mismo, tanto como sea posible, puedan descargar los DataSets completos para su reutilización.

- Recomendación A1 (GO FAIR, 2021e): Los datos y metadatos se pueden recuperar por su identificador utilizando un protocolo de comunicación estandarizado. Los identificadores persistentes, generalmente incluyen los protocolos de red asociados (http/https, ftp, etc.) que permite acceder a los datos y metadatos de investigación. Se busca que el acceso a los datos y metadatos se haga con protocolos abiertos y que se eviten las barreras de acceso o el uso de protocolos que tienen implementaciones limitadas, documentación deficiente y componentes que involucran la intervención humana o manual.
- Recomendación A1.1 (GO FAIR, 2021f): Para garantizar la accesibilidad, los protocolos de acceso a los datos y metadatos deben ser abiertos, gratuitos y de implementación universal. Se busca que cualquier persona que tenga un dispositivo electrónico conectado a Internet y como mínimo un navegador de internet pueda acceder a los contenidos completos o, al menos a los metadatos asociados. No deben existir barreras que impliquen instalar aplicativos especializados, versiones específicas de los navegadores, o la conexión mediante protocolos propietarios.
- Recomendación A1.2 (GO FAIR, 2021g): Los protocolos deben permitir, además, un procedimiento de autenticación y autorización, si fuera necesario. Que un recurso sea accesible, no significa que se encuentre necesariamente abierto. Pueden existir datos que tengan limitaciones de acceso debido a su contenido (datos

sensibles, personales) o que tengan barreras impuestas por las instituciones (patentes), los protocolos de acceso deben proporcionar las condiciones exactas bajo las cuales los datos son accesibles para que sean JUSTOS.

- Recomendación A2 (GO FAIR, 2021dh): Los metadatos deben ser accesibles incluso cuando los datos ya no estén disponibles. Mediante la aplicación de directrices de preservación a largo plazo, se debe garantizar como mínimo, el poder acceder a los metadatos, aunque los datos ya no estén disponibles (se encuentren obsoletos, cambien de ubicación, se hayan descartado o eliminado).

2.3 Principio 3. Datos Interoperables (Interoperable)

Los datos y DataSets deben ser construidos con formatos abiertos (formatos que pueden ser desplegados por múltiples herramientas y sistemas de información), así mismo deben ser almacenados en sistemas de información que integren estándares y protocolos abiertos que permitan su intercambio de manera transparente y autónoma.

- Recomendación I1 (GO FAIR, 2021i): Los datos y metadatos utilizan un lenguaje formal, accesible, compartido y de amplia aplicación para la representación del conocimiento. Se busca que el intercambio y la interpretación de los contenidos (datos y metadatos), se haga de manera transparente entre humanos y entre máquinas. Por eso, los sistemas informáticos utilizados deben incluir protocolos de interoperabilidad accesibles, compartidos y normalizados que permitan el intercambio y la interpretación automática y transparente de los recursos.
- Recomendación I2 (GO FAIR, 2021j): Describir los datos y metadatos con vocabularios (esquemas, ontologías, etc.) que también sigan los principios FAIR. Durante la organización de los datos y en la construcción de los DataSets, se busca utilizar las mejores prácticas estandarizadas de notación de campos, representación de contenidos/variables (especialmente categóricos) y la gestión de campos y contenidos de los metadatos mediante el uso de lenguajes formales y de uso estandarizado. Se recomienda utilizar los vocabularios semánticos propuestos por COAR, DataCite, y los entes reguladores nacionales. El uso de vocabularios semánticos involucra la creación de conceptos que incluyen identificadores persistentes.
- Recomendación I3 (GO FAIR, 2021k): Incluir referencias cruzadas y enlaces entre datos y metadatos. Una referencia cruzada especifica la constitución de una relación de carácter “normalizado” de un recurso/datos/metadato que enlaza mediante el uso de un identificador persistente, a otro recurso/datos/metadatos (Creación de vínculos científicos entre conjuntos de datos). Las distintas relaciones posibles entre recursos se deben tomar de vocabularios normalizados siguiendo los principios de la gestión de datos abiertos enlazados (Linked Open Data).

2.4 Principio 4. Datos Reutilizables (Reusable)

Este principio específico se debe crear DataSets completos y organizados que incluyan datos, metadatos, licencias, guías y documentos que describan todo el ciclo de vida de los datos de investigación y permitan la reutilización de los datos contenidos

- Recomendación R1 (GO FAIR, 2021l): Los datos y metadatos se describen detalladamente con una pluralidad de atributos precisos y relevantes. Esta recomendación se centra en la capacidad de un usuario (máquina o humano) para decidir si los datos son realmente ÚTILES en un contexto particular. Para ello, los datos deben incluir etiquetas y ayudas auto aclaratorias que permitan entender las variables asociadas. Así mismo, deben incluir documentación anexa que describa detalladamente el contexto en el que se generaron los datos y su entorno de medición, aplicación y limitaciones. Adicionalmente los metadatos deben ser ampliamente construidos a través de la mayor cantidad de campos de descripción normalizados utilizando

distintos esquemas de metadatos asociados a la tipología del dato, los cuales permitan su descubrimiento y reutilización.

- Recomendación R1.1 (GO FAIR, 2021m): Los datos y metadatos se publican con una licencia de uso de datos clara y accesible. Los investigadores que generaron los distintos conjuntos de datos deben proveer autorizaciones de publicación que permitan la distribución libre y abierta de los recursos mediante el uso de sistemas de información (repositorios, revistas y sistemas especializados). Dichos sistemas a su vez deben incluir en los datos publicados, de manera clara y visible, las distintas licencias de uso y acceso que especifiquen claramente las condiciones en las que se pueden reutilizar los datos y que deben ser claras para las máquinas y los seres humanos.
- Recomendación R1.2 (GO FAIR, 2021n): Los datos y metadatos incluyen una descripción detallada de la procedencia de estos. Se recomienda utilizar los criterios de procedencia que detallan los elementos de la creación, atribución, flujos de trabajo e historial de versiones de los datos durante todo su ciclo de vida. Idealmente, estos criterios de procedencia se deben describir en formatos legibles por máquina.
- Recomendación R1.3 (GO FAIR, 2021o): Los datos y metadatos cumplen con los estándares relevantes para el dominio y área de conocimiento comúnmente aceptados. Durante la generación de los datos, se deben revisar las mejores prácticas y estándares aplicables para su estructura y organización en el área de conocimiento aplicada: Utilizar los mismos tipos de datos, organizados de manera estandarizada, formatos de archivo bien establecidos y sostenibles, documentación (metadatos) siguiendo una plantilla común y utilizando un vocabulario común.

3 DISEÑO DE UN PLAN DE GESTIÓN DE DATOS DE INVESTIGACIÓN (PGD)

Un Plan de Gestión de Datos de investigación (PGD), es un documento corto y estructurado a partir de una plantilla que describe las tareas asociadas a la gestión de datos de investigación que se realizan durante todo el proyecto de investigación y que están asociadas a cada una de las fases del ciclo de vida de los datos de investigación. Dichas tareas describen de manera específica, bajo el contexto propio del proyecto de investigación (vinculados al dominio y disciplina científica del proyecto), el tratamiento que van a recibir los datos de investigación recopilados, procesados o generados por el mismo.

Un Plan de Gestión de Datos de investigación busca que los datos cumplan los Principios FAIR y se constituye como un documento vivo en continuo cambio y evolución que se va perfeccionando durante el período de vigencia del proyecto:

“El PGD debería actualizarse al menos una vez hacia la mitad del transcurso del proyecto y otra al final para hacer los ajustes necesarios sobre los datos generados y los usos identificados por el proyecto de investigación, ya que no todos los datos o los usos potenciales que se pueda hacer de ellos están claros desde el principio. Se deberán crear nuevas versiones del PGD siempre que haya cambios importantes en el proyecto debidos a la inclusión de nuevos conjuntos de datos, a cambios en las políticas o a factores externos” (Directrices para la Gestión de Datos en Horizonte 2020 Versión 1.0, 2013).

3.1 Herramientas para construir PGD

Para la construcción de planes de gestión de datos de investigación (PGD), se puede utilizar alguna de las herramientas existentes, las cuales proveen múltiples plantillas, modelos de aplicación y ejemplos de PDG de otros proyectos. Estas herramientas facilitan a los investigadores el desarrollo de este documento por medio de formularios amigables que incluyen explicaciones y aclaraciones. Las herramientas más conocidas son:

- DMP on line del Digital Curation Center (Reino Unido). <https://dmponline.dcc.ac.uk/>

- DMPTool (CDL-US) de la University of California Curation Center (UC3). <https://cdlib.org/services/uc3/dmptool/>
- PGDonline del Consorcio Madroño España. <https://pgd.consorciomadrono.es/>
- La iniciativa Argos de la Comisión Europea (Proyecto OpenAIRE). <https://argos.OpenAIRE.eu/>

3.2 Preparar el Plan de Gestión de Datos

El Plan de Gestión de Datos (PGD), generalmente establece qué datos se crearán, las herramientas y/o metodologías que se usarán para crearlos, así como la descripción de tareas para compartir, usar y preservar, señalando lo que es apropiado según la naturaleza de los datos y cualquier restricción que pueda ser necesaria aplicar a los datos según los términos de referencia asociados al desarrollo del proyecto, por ello, antes de elaborar el Plan de Gestión de Datos del proyecto, es indispensable examinar los siguientes elementos:

Si el proyecto cuenta con financiación interna o externa a su institución, es posible que, la unidad o agencia financiadora tenga requisitos más específicos para los planes de gestión de datos, por ello, es importante validar esta información en el sitio web de la agencia que proporciona los recursos.

Adicionalmente, para cada DataSets es necesario examinar las siguientes consideraciones generales como la descripción de los DataSets, estándares para datos y metadatos, criterios para compartir los datos, acceso de los datos, propiedad intelectual, archivo y preservación de los datos (Stanford University. Libraries, 2011):

3.3 Descripción de los DataSets

- Cómo se van a generar y usar los datos en el proyecto.
- Cada DataSets debe ser descrito completamente, considerando elementos como: información sobre el formato, el tamaño y volumen de los datos.
- Procesos que atravesarán los datos, por ejemplo: datos recopilados, procesados, analizados y publicados, etc.
- Metodologías que se utilizarán para cada proceso.
- Herramientas e instrumentos que se usarán.
- Cantidad estimada de datos que se generarán.
- Personas involucradas en cada proceso.
- Administración de los datos: cómo y cuándo serán almacenados los datos y en qué medio. Regularidad y persona encargada de realizar copias de seguridad de los datos.
- Costos asociados a la administración de los DataSets

3.4 Formato de procesamiento de los datos

- Existencia de estándares de datos en el área de dominio o conocimiento, por ejemplo: .XML, .CSV, .TXT, DTA, etc. (Vilches, 2020). Si no hay estándares definidos para el área de dominio debe declarar en qué formato estándar procesa los datos para que estén disponibles para otros investigadores. Se recomienda usar formatos que cumplan principios FAIR.
- Personal a cargo de garantizar que se cumpla con la aplicación y estructuración del estándar de datos declarado.

3.5 Estándares de metadatos

El estándar de metadatos “establece criterios, métodos, procesos y prácticas uniformes. Han sido creados con el propósito específico de guiar el diseño, creación e implementación de estructuras de datos, valor de los datos, contenido de los datos y el intercambio de los datos de manera consistente y eficiente” (Zeng & Qin, 2016, p. 23).

En este aspecto debe considerar:

- Cómo serán generados y capturados los metadatos para cada uno de los DataSets
- El estándar de estructura de metadatos que va a usar el DataSets. Algunas áreas de conocimiento utilizan las siguientes estructuras: .RDF, Dublin Core, Ecological Metadata Language, EML, EAD, MODS, TEI, NeXus, entre otros (RDA Metadata Standards Directory Working Group, s/f).
- La elección del estándar de metadatos deberá garantizar que estos datos sean encontrables, accesibles, interoperables y reutilizables.
- Se debe delegar a una persona que garantice el cumplimiento del estándar de metadatos declarado.

3.6 Compartir datos

En este aspecto es importante alinearse al marco de referencia normativo al que se acoja el proyecto. Algunos financiadores solicitan como resultado la posibilidad de compartir y divulgar los datos, siempre y cuando se respeten las normas relacionadas con el derecho a la privacidad, datos sensibles, así como otras consideraciones éticas que requieran ser declaradas para poder compartir los datos. Aquí es importante enfocarse en la audiencia objetivo de los DataSets, tales como: investigadores dentro o fuera de su área de conocimiento, oficinas del gobierno, etc. y cómo podrán usar los datos que genere su proyecto.

Debe definir en qué momento podrá poner a disposición de los otros los DataSets de su proyecto. Por ejemplo: inmediatamente después de haber sido recolectados, luego de ser publicados, etc.

3.7 Acceso de los datos

Identificar repositorios que soporten la publicación y disponibilidad de sus datos, así como los requisitos que debe cumplir para poder publicar sus datos en el repositorio seleccionado. Si no encuentra repositorios para sus datos, debe indicar el sistema o mecanismo a través del cual compartirá sus datos con los demás investigadores.

3.8 Licencias de uso y atribución

Si es un proyecto financiado, es necesario revisar las cláusulas relacionadas con los derechos de uso de los datos y establecer la licencia de acuerdo con esta revisión. De lo contrario debe establecer en el PGD quién tiene la propiedad de los datos, las condiciones de uso, redistribución, creación de obras derivadas y si se permitirá el uso de los datos con fines comerciales.

3.9 Archivo y preservación

La preservación digital se refiere a las actividades y estrategias para asegurar acceso a los contenidos reformateados o nativos digitales en consideración a los retos planteados por la obsolescencia tecnológica. Es necesario declarar los valores a largo plazo que puedan tener los DataSets del proyecto, garantizar el acceso a los datos después de la culminación del proyecto, información adicional que deba conservarse con los datos.

3.10 Campos de información del PGD

Existe una plantilla del Digital Curation Center con todos los campos que requiere:

- a) Información administrativa: identificador del centro, financiador, número de identificación de la financiación, nombre del proyecto, descripción, IP, identificador Orcid del IP, fecha de la primera versión, etc.
- b) Descripción de los datos: tipología, volumen, formatos, estructura de los archivos y gestión de las versiones
- c) Documentación y metadatos que acompañarán a los datos
- d) Requisitos legales y éticos: consentimiento de las personas implicadas, protección de la identidad, tratamiento de los datos sensibles y gestión de derechos de propiedad intelectual
- e) Almacenamiento y copias de seguridad: control de acceso a los colaboradores y gestión de las copias de seguridad
- f) Selección y preservación: qué datos y cómo serán conservados, dónde se preservarán y restricciones en el intercambio de datos
- g) Compartir los datos: en qué condiciones se compartirán, cómo encontrarán los usuarios los datos, cuándo estarán disponibles, obtendrán un identificador persistente, etc.
- h) Responsabilidades y recursos: asignación de responsabilidades, costes asociados al acceso abierto de los datos, ya que son elegibles para su reembolso durante el periodo del proyecto

Aunque los planes de gestión de datos puedan adoptar muchas formas (Digital Curation Centre, DCC, 2021; Consorcio de Universidades de la Comunidad de Madrid y de la UNED para la Cooperación Bibliotecaria, Madroño, 2018) según el área de conocimiento, la entidad financiadora, entre otros aspectos, deben abordarse ciertos componentes principales en todos ellos (Universidad de Navarra. Biblioteca, 2021):

- Descripción de los datos y metadatos
- Actualizar (metadatos, documentación)
- Organización
- Adquisición
- Procesamiento
- Análisis
- Preservación
- Publicación
- Identificadores
- Citación de datos
- Copia de seguridad
- Ética
- Propiedad intelectual
- Acceso y reutilización
- Almacenamiento a corto plazo de gestión y preservación
- Recursos
- Personal
- Consideraciones para compartir datos
- Formas de compartir los datos.

3.11 Plantilla Plan de Gestión de Datos

Existen diversidad de plantillas para el PGD, cada una delimitada por los objetivos, principios y requisitos de las entidades apoyan o financian proyectos de investigación, en su mayoría del norte global, así como plantillas específicas por área de conocimiento dada la especificidad con que se requieren los datos para algunas áreas. En el siguiente enlace podrá ver algunos ejemplos en inglés de PGD de diferentes instituciones y áreas de conocimiento: <https://www.dcc.ac.uk/resources/data-management-plans/guidance-examples>

A continuación se sugiere la siguiente plantilla para el PGD adaptada de las propuestas planteadas por Digital Curation Centre, DCC (2013), el Consorcio de Universidades de la Comunidad de Madrid y de la UNED para la Cooperación Bibliotecaria, Madroño (2020) y la Universidad de Stanford (2011):

Información administrativa	
Identificador	Una identificación determinada por el financiador y/o la institución.
Financiador	Nombre del financiador
Número de identificación de la financiación	Código numérico o alfanumérico que identifica la fuente y recursos.
Nombre del proyecto	Nombre exacto del proyecto tal y como fue aprobado por la instancia respectiva.
Descripción del proyecto	Realice una descripción breve del objetivo y naturaleza del proyecto, la metodología, alcance, las preguntas que orientan la investigación y el objetivo por el cual los datos van a ser recopilados o creados.
Investigador	Nombre del o los investigadores(es) principal(es)
ID del investigador	Identificador persistente del o los investigadores principales: ejemplo ORCID, ISNI, etc.
Datos de contacto del investigador principal	Teléfono y correo electrónico de los investigadores principales
Fecha de la primera versión	Fecha en la que el Plan de Gestión de Datos fue completado
Fecha de la última actualización	Fecha en la que el PGD fue actualizado o cambiado
Marco normativo relacionado	Debe listar en este espacio el marco normativo o políticas con las cuales los datos estén relacionados. Esto incluye los requisitos del financiador (si aplica), las políticas establecidas por su institución para la gestión de datos, las políticas a nivel nacional e internacional que le apliquen.
Creación o recolección de datos	
Tipo de datos a crear o recolectar	<p>Incluya una breve descripción de los datos, fuentes de donde se van a recolectar, tipo de datos y cobertura de la fuente donde se encuentran los datos existentes.</p> <p>Descripción de los DataSets. Describa el tipo, formato y volumen de los datos a recolectar o crear.</p> <p>Estándares de metadatos. Especifique si el tipo de dato, formato y software seleccionados permiten compartir y acceder a los datos a largo plazo.</p> <p>Justifique su decisión de formato y considere las implicaciones de almacenamiento, copias de seguridad y acceso de los datos.</p> <p>Justifique la metodología de recolección de datos (estándares, extensiones de</p>

	archivo, control de versiones,
Metodología de creación o recolección de los datos	<p>Explique la metodología en que los datos van a ser recolectados</p> <p>Establezca una estructura en la cual serán almacenados los datos</p> <p>Incluya los criterios para garantizar la calidad y consistencia de los datos (ejemplo: captura de datos estandarizada, validación de datos, revisión por pares, uso de vocabularios controlados, etc.)</p>
Documentación de los datos	
Documentos de soporte	<p>Describa los documentos que acompañaran a los datos y que ayudan a otros investigadores a entender los datos. Los documentos soporte deben incluir detalles que ayuden con la identificación de los datos, como quién creó o contribuyó con la recolección de los datos, título, fecha de creación y las condiciones bajo las cuales pueden ser accedidos.</p> <p>También deben incluir detalles sobre la metodología utilizada para crear o recolectar los datos, análisis y procedimientos usados, definición de variables, unidades de medida y en cualquier otra característica que contribuya con la descripción y entendimiento de los DataSets.</p>
Ética y cumplimiento de requisitos legales	
Consideraciones éticas datos sensibles / datos humanos	<p>Dado que las consideraciones éticas pueden afectar el almacenamiento, acceso, uso y tiempo de preservación de los datos, es importante describir las acciones que se realizarán para cumplir con los requisitos legales asociados a datos sensibles. Dentro de estas actividades se encuentra la anonimización de los datos, información sobre el comité de ética, acuerdos de consentimiento informado, entre otros.</p> <p>Es imperativo mostrar que el PGD está al tanto de las consideraciones éticas derivadas de la recolección/creación de datos a través de la planeación y gestión de las actividades pertinentes.</p>
Gestión de propiedad intelectual	<p>Debe declarar quién tendrá la propiedad intelectual de los datos creados/recopilados del proyecto, junto con la licencia de uso que será asignada a los datos. Es importante que la propiedad y licencia estén alineados con los términos de referencia del proyecto, las cláusulas dispuestas por los financiadores en esta materia, así como de los datos recolectados de otras fuentes.</p> <p>La licencia debe especificar claramente los términos de acceso, uso, divulgación o construcción de obras derivadas a partir de los datos.</p>
Almacenamiento y copias de seguridad (Backup)	
Backups	<p>Indique la periodicidad con la que se realizarán copias de seguridad de sus datos, así como el lugar de almacenamiento de las mismas. Dado que el almacenamiento en computadores, discos duros y otras unidades propias puede ser riesgoso, se recomienda gestionar el almacenamiento en repositorios robustos proveídos por su institución; también es recomendable gestionar las copias de seguridad de forma automática.</p>

Seguridad y acceso de los datos	Si los datos creados/recolectados contienen información sensible (ejemplo: datos de personas, patentes, etc.) debe indicar las medidas de seguridad que aplicará el proyecto.
Preservación de los datos	
Selección	Indique qué datos deben conservarse de acuerdo con su valor como validadores de los resultados del proyecto, soporte para una nueva investigación, etc. Este proceso también puede basarse en los términos de referencia u obligaciones vinculadas a los datos, viabilidad económica de su preservación, soporte tecnológico de los formatos de archivo usados para los datos, etc.
Plan de preservación	Considere y explique cuál será el plan de preservación y curaduría a largo plazo de los DataSets, esto incluye la preparación y la documentación de los datos para compartirlos y archivarlos. Se recomienda el uso de repositorios de datos estables para este propósito o demostrar los recursos y sistemas que dispondrán los datos para ser preservados y curados después de la finalización del proyecto.
Acceder y compartir datos	
Compartir y divulgar los datos	<p>Declare cómo, cuándo y a través de qué mecanismos los datos de valor a largo plazo serán puestos a disposición de los demás investigadores o público en general, esto dependerá de factores tales como el tipo, tamaño, complejidad y sensibilidad de los datos.</p> <p>Considere los potenciales usuarios de sus datos, los repositorios o mecanismos que permitirán compartirlos, la fecha de publicación de los datos, así como que estos cuenten con un identificador persistente.</p> <p>Identificar repositorios que soporten la publicación y disponibilidad de sus datos, así como los requisitos que debe cumplir para poder publicar sus datos en el repositorio seleccionado.</p>
Restricciones	Describa las potenciales restricciones al compartir sus datos con las causas y posibles soluciones. Las restricciones pueden derivarse de acuerdos de confidencialidad del proyecto, falta de acuerdos de consentimiento, derechos de propiedad intelectual, entre otros.
Responsabilidades y recursos necesarios	
Gestión de datos	<p>Describa los roles y responsabilidades asignados a los individuos y/o organizaciones vinculadas al proyecto relacionados con la captura de datos, la producción de metadatos, la calidad de los datos, almacenamiento y copias de seguridad.</p> <p>Indique con nombres y apellidos quién será el responsable de asegurar que se cumplan todas las políticas que deban aplicarse a los datos según su naturaleza.</p>
Recursos necesarios	Realice una descripción detallada de los recursos necesarios tales como software, hardware, experiencia técnica, entre otros que se requieran para llevar a cabo el plan de gestión de datos. Cada recurso que se incluya debe estar debidamente justificado.
Divulgación de los datos	

Dirigidos a:	A quién van dirigidos los datos, cuándo se pondrán a disposición. Se tiene identificadores persistentes, qué requisitos de los financiadores se tiene presentes para que sean abiertos. Tiempo de exclusividad de los datos. Se tiene previsto acuerdos de intercambios de datos.
--------------	---

4 GLOSARIO DE TÉRMINOS:

Administración De Datos: es la implementación de políticas y procedimientos de control de datos para establecer la precisión, confiabilidad, integridad y seguridad de los datos. Las personas asignadas con responsabilidades de administración de datos gestionan y supervisan los procedimientos y herramientas usados para gestionar, almacenar y proteger los datos.

Archivo: La transferencia de material a una instalación que evalúa, conserva y proporciona acceso a ese material a largo plazo o de forma permanente.

Big Data: es un término que describe conjuntos de datos extremadamente grandes de datos estructurados, no estructurados y semiestructurados. Big Data se caracteriza a menudo por las cinco "V": el gran volumen de datos recopilados, la variedad de tipos de datos, la velocidad a la que se generan los datos, su veracidad y valor. Con sistemas y analíticas para gestión de Big Data, las empresas pueden realizar minería de Big Data para obtener información estratégica profunda que guíe la toma de decisiones y las acciones.

Calidad De Datos: describe la idoneidad y fiabilidad de los datos. Que los datos sean de buena calidad simplemente significa que son precisos (verdaderamente representativos de lo que describen), confiables (consistentes, auditables, protegidos y adecuadamente gestionados) y completos en la medida necesaria para usuarios y aplicaciones. La calidad de los datos solo puede garantizarse mediante una estrategia de datos debidamente elaborada y ejecutada que se lleve a cabo con herramientas y sistemas con solidez en la industria, junto con políticas y procedimientos de gestión de datos escrupulosamente aplicados.

Código Informático O Script: En el contexto de la gestión de datos, esto puede incluir el código utilizado en la recopilación, manipulación, procesamiento, análisis o visualización de datos, pero también puede incluir software desarrollado para otros fines.

Conjunto de Datos: Una unidad mínima de información sujeta a carga, publicación, transformación y descarga en la plataforma.

Copyright: Un conjunto de derechos legales extendidos a los propietarios de derechos de autoría (la persona autora o creadora u otra parte a la que se han asignado los derechos) que rigen actividades tales como reproducir, distribuir, adaptar o exhibir obras originales fijadas en forma tangible. Los derechos de autoría no se aplican a la información fáctica; como resultado, no se aplica a los datos.

Ciclo de vida de los datos: Toda investigación tiene un ciclo vital. El modelo del ciclo de vida de los datos de investigación describe e identifica los pasos que se deben seguir en las diferentes etapas del ciclo de investigación para garantizar el éxito en la conservación de datos.

FAIR: (de acuerdo con FORCE11 principales y según lo publicado en Nature Scientific Data) son datos Localizables, Accesibles, Interoperables y Reutilizables, con el fin de facilitar el descubrimiento de conocimiento con o sin ayuda de operadores humanos, así como el acceso, integración y análisis de datos científicos relevantes, y de los algoritmos y flujos de trabajo asociados a los mismos.

Datos Abiertos de Investigación: es uno de los pilares de Open Science, movimiento que promueve el acceso libre a los datos de investigación abiertos para facilitar que se puedan usar, reutilizar y distribuir. Además, su publicación garantiza el acceso libre, y su correcta preservación, reproducción, difusión, visibilidad e impacto.

DOI: Digital Object Identifier. Es un identificador numérico que detalla la pertenencia o asociación de un artículo o revista a una determinada asociación científica y profesional.

Formatos Abiertos: Son formatos de archivo que se pueden crear y manipular por cualquier software, libre de restricciones legales.

Formatos de Archivo Persistente/Preferido: Formatos no propietarios que siguen estándares internacionales documentados, y que son comúnmente utilizados por la comunidad de investigación, utilizan la codificación de caracteres estándar (por ejemplo, ASCII, UTF-8), y la compresión, si se usa, no ocasiona pérdidas.

Financiador de la Investigación / Research Funder Un instituto, corporación o entidad gubernamental que proporciona asistencia financiera para la investigación.

Gestión de Datos de Investigación: La Gestión de Datos de Investigación es el proceso activo del manejo de los datos generados o recopilados en una investigación. La gestión de datos debe planificarse desde el principio de la investigación y cubre todas las decisiones relacionadas con la gestión de los datos a lo largo de su ciclo vital. La gestión de datos comienza en la etapa de planificación de la investigación, continúa durante la ejecución del proyecto hasta la difusión de los resultados y la preservación de los conjuntos de datos, de forma que estos sean precisos, completos, fiables, accesibles y reutilizables a lo largo del tiempo. La gestión de datos abarca: Planificar la investigación, Recopilar o generar datos, Procesar y analizar los datos, Preservarlos y finalmente publicar y compartir los datos de investigación.

Identificador: Una referencia única y duradera que permite el acceso continuo a un objeto digital. Ejemplos de sistemas de identificación persistente son: identificadores de objetos digitales (DOI) o identificadores y claves de recursos de archivo (ARK). Los identificadores persistentes ayudan a la interoperabilidad y la citación fiable de contenido digital.

Identificador Persistente (PID): Un identificador persistente (también PID) es una denominación única y estable (referencia) de un recurso digital (por ejemplo, datos de investigación) mediante la asignación de un código al que se puede hacer referencia de forma persistente y explícita en Internet.

Integridad de los Datos: se refiere a la veracidad de los datos a largo plazo. Una vez que los datos se ingresan o importan, manipulan, validan, depuran y almacenan, su integridad es la declaración de que su calidad se mantiene, y los usuarios pueden estar seguros de que los datos que ingresaron no han cambiado ni cambiarán. Los datos que se recuperan son los mismos que se almacenaron originalmente. A veces usada como sinónimo de calidad de datos, la integridad de los datos se centra más en la confiabilidad.

Metadato: Documentación o información sobre un conjunto de datos. Puede estar incrustado en los datos en sí o existir por separado de los datos. Los metadatos pueden describir el propietario, el propósito, los métodos, la organización y las condiciones de uso de los datos, información técnica sobre los datos y otra información. Existen muchos estándares de metadatos en una amplia gama de disciplinas y aplicaciones.

ORCID: Open Research and Contributor ID. Es un identificador persistente y unívoco de autor estandarizado a nivel internacional, que facilita la recuperación bibliográfica asociada a un mismo autor en sistemas de información científica.

Plan de Gestión de Datos (Data Management Plan-Dmp): es un documento formal y "vivo" en el que se detalla y describe cómo se va a gestionar y tratar los datos de investigación recopilados o generados durante y después del proyecto de investigación. Describe el ciclo de vida de la gestión del conjunto de datos que son recopilados, procesados o generados en el transcurso del proyecto de investigación e incluso tras su conclusión. Además, describe

qué metodología y estándares se utilizarán, cómo se compartirán y serán abiertos, y por último cómo se conservarán y preservarán.

Privacidad de Datos: hace referencia a las políticas y prácticas para manejar los datos de manera que estén protegidos del acceso o la divulgación no autorizados. Las políticas y prácticas de privacidad de datos cubren cómo se recopila y almacena la información según la estrategia de datos de la organización, cómo se puede compartir o no con terceros y cómo cumplir con las restricciones regulatorias. La privacidad de datos es un imperativo de negocio que satisface las expectativas del cliente al tiempo que protege la integridad y seguridad de la información almacenada.

Preservación de Datos: Garantizar que los datos permanezcan intactos, accesibles y comprensibles a lo largo del tiempo. Esto requiere preservar la integridad de los archivos digitales y puede ser considerablemente más complicado. Las operaciones de preservación pueden incluir preservar el software requerido para interactuar con los datos o emular sistemas más antiguos, migrar datos a nuevos formatos y medios, para asegurar que haya suficientes metadatos para comprender, interpretar, administrar y preservar los datos.

Reutilización de Datos: Producto que se elabora a partir de los datos públicos, puede ser una visualización, una aplicación web, un servicio, un cuadro de mandos, una noticia o una información, un dibujo, una gráfica dinámica, entre otras cosas.

RoR: Research Organization Registry. Es un identificador persistente de filiación institucional, que facilita la correspondencia de las asociaciones científicas en diversos sistemas de información.

Usabilidad: Significa facilidad de uso, atributo de calidad, que identifica el grado en que un producto puede ser usado por determinados usuarios para lograr sus propósitos con eficacia, eficiencia y satisfacción en un contexto de uso específico.

5 BIBLIOGRAFÍA

Actis, G., & Carlino, L. (2017). Actas BIREDIAL-ISTEC 2017: VII Conferencia Internacional sobre Bibliotecas y Repositorios Digitales de América Latina (BIREDIAL-ISTEC'17): XII Simposio Internacional de Bibliotecas Digitales (SIBD'17) (M. R. De Giusti, Ed.). Universidad Nacional de La Plata; ISTE. <https://libros.unlp.edu.ar/index.php/unlp/catalog/download/802/794/2650-1>

Aleixandre-Benavent, R., Ferrer Sapena, A., & Peset, F. (2021). Compartir los recursos útiles para la investigación: Datos abiertos (open data). *Educación Médica*, 22, 208–215. <https://doi.org/10.1016/j.edumed.2019.07.004>

Borgman, C. (2012). Why are the attribution and citation of scientific data important? En P. E. Uhlir (Ed.), *For Attribution: Developing Data Attribution and Citation Practices and Standards: Summary of an International Workshop* (pp. 1–10). The National Academies Press. <https://doi.org/10.17226/13564>

Biblioteca Stanford Bussines. (2011). Research Databases & Datasets. <https://www.gsb.stanford.edu/library/research-resources/databases>

Consortia Advancing Standards in Research Administration Information, CASRAI. (2021). Research Data Management Glossary. CASRAI. <https://casrai.org/rdm-glossary/>

Consortio de Universidades de la Comunidad de Madrid y de la UNED para la Cooperación Bibliotecaria, Madroño. (2018). Crear su PGD. Consortio Madroño. <http://www.consorcioamadrono.es/investigam/crear-su-pgd/>

Consortio de Universidades de la Comunidad de Madrid y de la UNED para la Cooperación Bibliotecaria, Madroño. (2020). European Commission (H2020): Horizon 2020 DMP - Initial DMP. https://pgd.consorciomadrono.es/public_templates

Couto Corrêa, F. (2016). Gestión de datos de investigación. Editorial UOC.

DataCite Schema. (2019). [Website]. DataCite Schema. <https://schema.datacite.org/>

Davenport, T. H., & Prusak, L. (1998). Working knowledge: How organizations manage what they know. Harvard Business School Press. <http://search.ebscohost.com/login.aspx?direct=true&scope=site&db=nlebk&db=nlabk&AN=7259>

Directrices para la Gestión de Datos en Horizonte 2020. (2013). https://www.consorciomadrono.es/wp-content/uploads/2017/05/directrices_gestion_datos_horizon_2020_es.pdf

Digital Curation Centre, DCC. (2013). Checklist for a Data Management Plan. V.4.0. Digital Curation Centre. <http://www.dcc.ac.uk/resources/data-management-plans>

Digital Curation Centre, DCC. (2021). Example DMPs and guidance. <https://www.dcc.ac.uk/resources/data-management-plans/guidance-examples>

García Herrero, J., Molina López, J. M., Berlanga de Jesús, A., Patricio Guisado, M. Á., Bustamante, Á. L., & Padilla R., W. (2018). Ciencia de datos: Técnicas analíticas y aprendizaje estadístico. Un enfoque práctico. Alfaomega.

GO FAIR. (2021). R1: (Meta)data are richly described with a plurality of accurate and relevant attributes. GO FAIR. <https://www.go-fair.org/fair-principles/r1-metadata-richly-described-plurality-accurate-relevant-attributes/>

GO FAIR. (2021a). A1: (Meta)data are retrievable by their identifier using a standardised communication protocol. GO FAIR. <https://www.go-fair.org/fair-principles/metadata-retrievable-identifier-standardised-communication-protocol/>

GO FAIR. (2021b). A1.1: The protocol is open, free and universally implementable. GO FAIR. <https://www.go-fair.org/fair-principles/a1-1-protocol-open-free-universally-implementable/>

GO FAIR. (2021c). A1.2: The protocol allows for an authentication and authorisation where necessary. GO FAIR. <https://www.go-fair.org/fair-principles/a1-2-protocol-allows-authentication-authorisation-required/>

GO FAIR. (2021d). A2: Metadata should be accessible even when the data is no longer available. GO FAIR. <https://www.go-fair.org/fair-principles/a2-metadata-accessible-even-data-no-longer-available/>

GO FAIR. (2021e). F1: (Meta) data are assigned globally unique and persistent identifiers. GO FAIR. <https://www.go-fair.org/fair-principles/f1-meta-data-assigned-globally-unique-persistent-identifiers/>

GO FAIR. (2021f). F2: Data are described with rich metadata. GO FAIR. <https://www.go-fair.org/fair-principles/f2-data-described-rich-metadata/>

GO FAIR. (2021g). F3: Metadata clearly and explicitly include the identifier of the data they describe. GO FAIR. <https://www.go-fair.org/fair-principles/f3-metadata-clearly-explicitly-include-identifier-data-describe/>

GO FAIR. (2021h). F4: (Meta)data are registered or indexed in a searchable resource. GO FAIR. <https://www.go-fair.org/fair-principles/f4-metadata-registered-indexed-searchable-resource/>

GO FAIR. (2021i). FAIR Principles. GO FAIR. <https://www.go-fair.org/fair-principles/>

- GO FAIR. (2021j). I1: (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation. GO FAIR. <https://www.go-fair.org/fair-principles/i1-metadata-use-formal-accessible-shared-broadly-applicable-language-knowledge-representation/>
- GO FAIR. (2021k). I2: (Meta)data use vocabularies that follow the FAIR principles. GO FAIR. <https://www.go-fair.org/fair-principles/i2-metadata-use-vocabularies-follow-fair-principles/>
- GO FAIR. (2021l). I3: (Meta)data include qualified references to other (meta)data. GO FAIR. <https://www.go-fair.org/fair-principles/i3-metadata-include-qualified-references-metadata/>
- GO FAIR. (2021m). R1.1: (Meta)data are released with a clear and accessible data usage license. GO FAIR. <https://www.go-fair.org/fair-principles/r1-1-metadata-released-clear-accessible-data-usage-license/>
- GO FAIR. (2021n). R1.2: (Meta)data are associated with detailed provenance. GO FAIR. <https://www.go-fair.org/fair-principles/r1-2-metadata-associated-detailed-provenance/>
- GO FAIR. (2021o). R1.3: (Meta)data meet domain-relevant community standards. GO FAIR. <https://www.go-fair.org/fair-principles/r1-3-metadata-meet-domain-relevant-community-standards/>
- Green, T. (2009). We Need Publishing Standards for DataSets and Data Tables. OECD Publishing white papers. <https://unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.40/2010/wp.8.e.pdf>
- Grupo de Trabajo de “Depósito y Gestión de datos en Acceso Abierto” del proyecto RECOLECTA. (2012). La conservación y reutilización de los datos científicos en España [Informe del grupo de trabajo de buenas prácticas]. Fundación Española para la Ciencia y la Tecnología, FECYT. https://recolecta.fecyt.es/sites/default/files/contenido/documentos/informe_datos_cientificos_en_esp.pdf
- Higgins, S. (2008). The dcc curation lifecycle model. Proceedings of the 8th ACM/IEEE-CS Joint Conference on Digital Libraries - JCDL '08, 453. <https://doi.org/10.1145/1378889.1378998>
- Institute for Quantitative Social Science. (2021). DataSets + File Management. <https://guides.dataverse.org/en/latest/user/DataSets-management.html>
- Joint Information Systems Committee, JISC. (2018). Research data management toolkit. Jisc. <https://www.jisc.ac.uk/full-guide/rdm-toolkit>
- Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Hung Byers, A. (2011). Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute. <https://www.mckinsey.com/business-functions/mckinsey-digital/our-insights/big-data-the-next-frontier-for-innovation>
- Melero, R. (2018). Recomendaciones para la gestión de datos de investigación dirigidas a investigadores [Proyecto CSO2015-71867-REDT]. Maredata. Red Española sobre Datos de Investigación en Abierto. <http://digital.csic.es/bitstream/10261/173801/1/Maredata-recomendaciones-ESP.pdf>
- Melero, R., & Hernández-San-Miguel, J. (2014). Acceso abierto a los datos de investigación, una vía hacia la colaboración científica. Revista española de Documentación Científica, 37(4), e066. <https://doi.org/10.3989/redc.2014.4.1154>
- National Science Board. (2005). US NSF - NSB-05-40, Long-Lived Digital Data Collections Enabling Research and Education in the 21st Century. NSF. <https://www.nsf.gov/pubs/2005/nsb0540/>
- OpenAIRE. (2015). OpenAIRE Guidelines for Data Archives—OpenAIRE Guidelines documentation. <https://guidelines.OpenAIRE.eu/en/latest/data/index.html>

- Peset Mancebo, F., & González, L. (2017). ¿¿Ciencia Abierta y gestión de datos de investigación (RDM)?? Trea.
- Peset, F., Alexandre-Benavent, R., Blasco-Gil, Y., & Ferrer-Sapena, A. (2017). Datos abiertos de investigación. Camino recorrido y cuestiones pendientes. *Anales de Documentación*, 20(1). <https://doi.org/10.6018/analesdoc.20.1.272101>
- Prentice, S. (2011). CEO Advisory: "Big Data" Equals Big Opportunity. <https://www.gartner.com/en/documents/1614215/ceo-advisory-big-data-equals-big-opportunity>
- Strasser, C., Cook, R., Michener, W., & Budden, A. (2012). Primer on Data Management: What you always wanted to know. UC Office of the President: California Digital Library. <https://doi.org/10.5060/D2251G48>
- Swan, A., & Brown, S. (2008). To share or not to share: Publication and quality assurance of research data outputs. A report commissioned by the Research Information Network [Project Report]. s.n. <https://eprints.soton.ac.uk/266742/>
- Torres-Salinas, D., Robinson-García, N., & Cabezas-Clavijo, Á. (2012). Compartir los datos de investigación en ciencia: Introducción a la data sharing. *El Profesional de la Información*, 21(2), 173–184. <https://doi.org/10.3145/epi.2012.mar.08>
- UK Research and Innovation, UKRI. (2021). The importance of managing and sharing data [Research data management]. UK Data Service. <https://ukdataservice.ac.uk/learning-hub/research-data-management/>
- Universidad de Alcalá. Biblioteca. (2021). Biblioguías: Datos de investigación: La gestión de datos y los PGD. https://uah-es.libguides.com/datos_investigacion/gestion
- Universidad de Valladolid. Biblioteca. (2021). Biblioguías: Datos de Investigación: Ciclo de vida de los datos. <https://biblioguias.uva.es/datos-investigacion/ciclo-vida>
- Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., ... Mons, B. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3(1), 160018. <https://doi.org/10.1038/sdata.2016.18>
- Zeng, M. L., & Qin, J. (2016). *Metadata* (2nd edition). Neal-Schuman.