CÓMO ENFRENTARSE A UN PLAN DE GESTIÓN DE DATOS DE INVESTIGACIÓN

IV JORNADA BIBLIOMADSALUD

ALICIA FÁTIMA GÓMEZ SÁNCHEZ, PHD

9 de junio de 2022

BIBLIO MAD SALUD

Objetivos

- Conocer qué son los datos de investigación abiertos y su gestión en el marco de la ciencia abierta.
- Entender qué significan los criterios FAIR y cómo hacer que los datos cumplan dichos criterios.
- Distinguir las fases que intervienen en el ciclo de vida de los datos y conocer aspectos esenciales para gestionar y compartir datos.
- Aprender qué son los planes de gestión de datos y cómo se pueden desarrollar.
- Identificar qué servicios pueden ofrecer las bibliotecas en materia de gestión de datos.

Contenidos

- Introducción. Conceptos generales en torno a la gestión de los datos de investigación
- Qué son los planes de gestión de datos (PGDs).
- Información mínima a incluir en los PGDs
- Plantillas y herramientas (DMPonline, Argos, etc.) para elaborar PGDs.

¿Dónde ubicamos la gestión de los datos?

Qué sabemos de...

Ciencia abierta

Datos abiertos

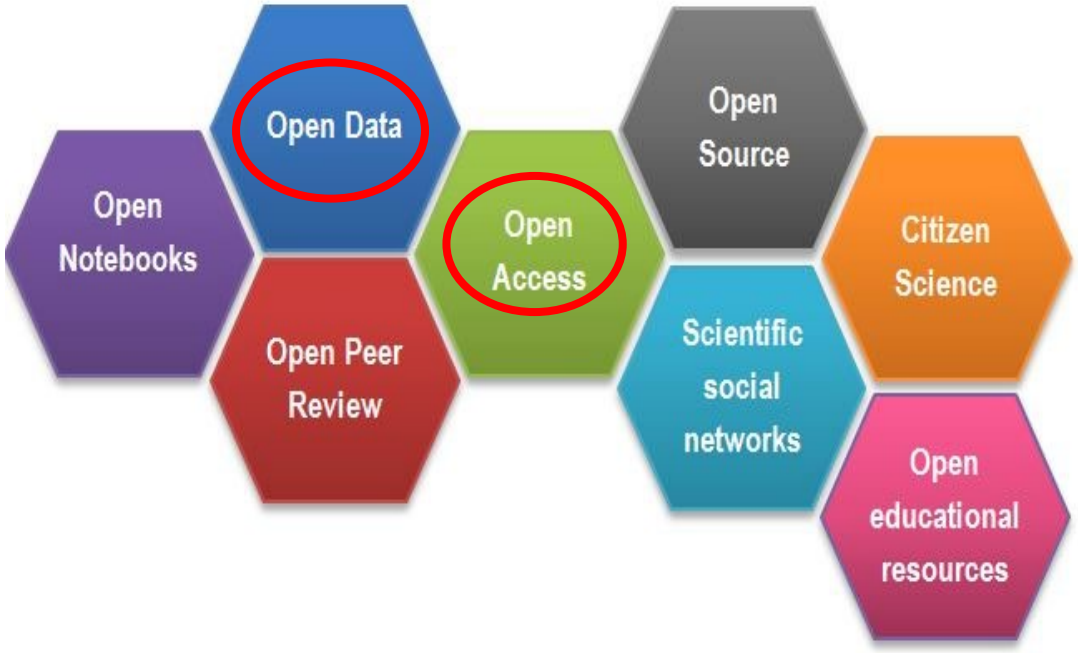
RRI

FAIR

¿Beneficios?

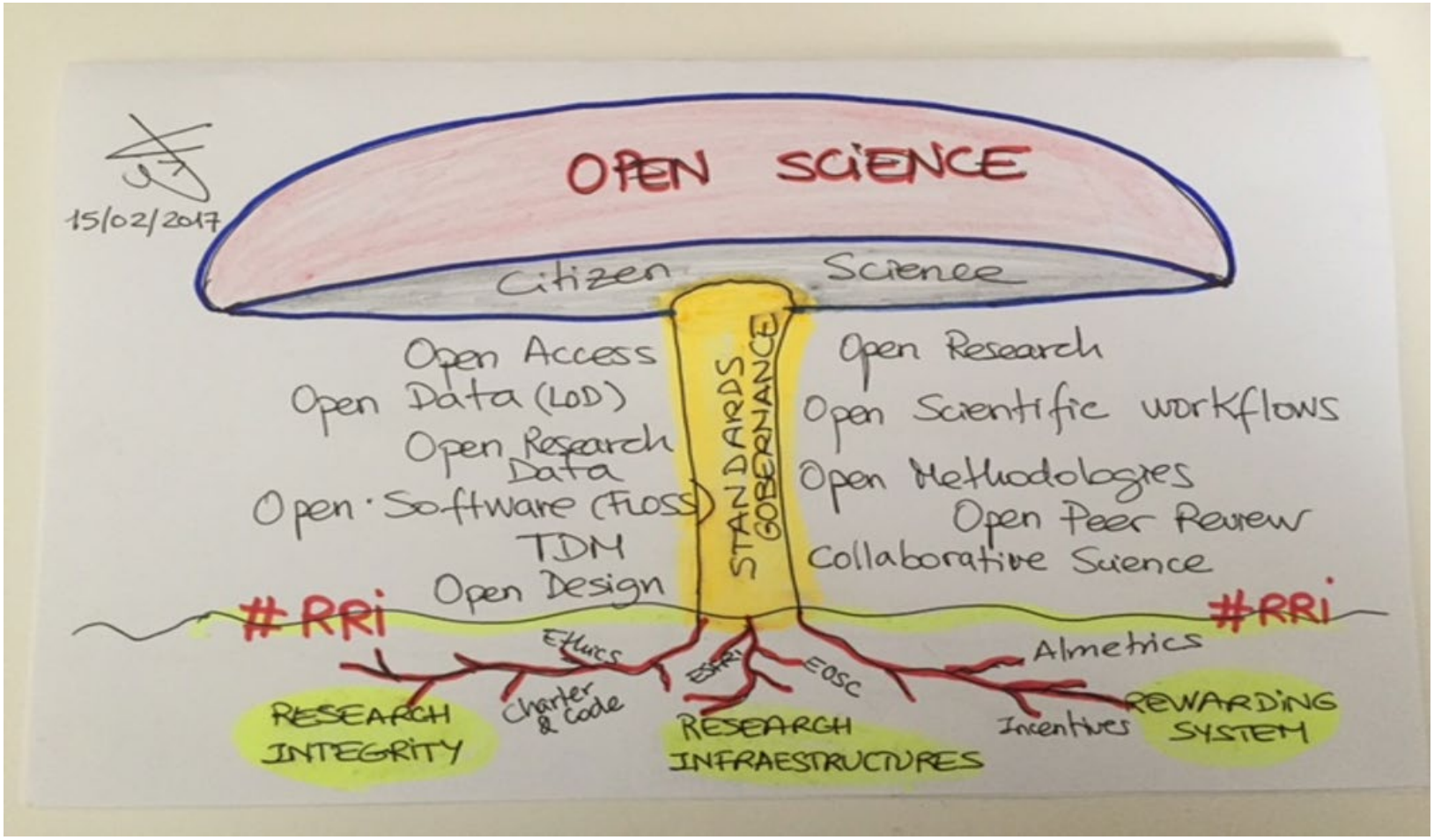
Ciencia abierta

¿Para qué?



- Mayor transparencia
- Ciencia más eficiente, con un mayor impacto
- …y mayor visibilidad!

https://www.fosteropenscience.eu/content/what-open-science-introduction



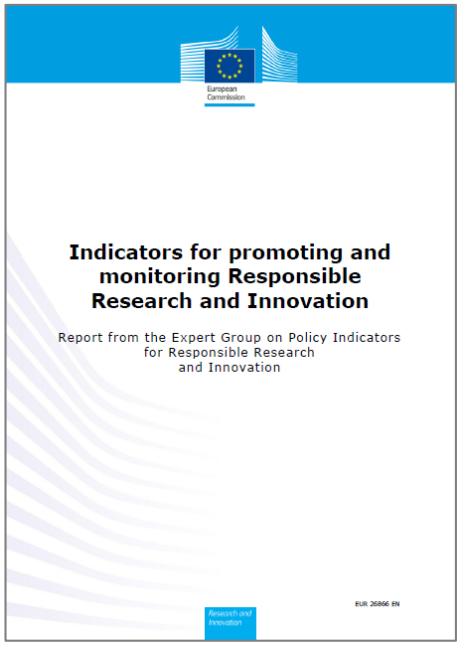
Eva Méndez (2017). CC-BY
https://twitter.com/evamen/status/844158610225664000?lang=en

Responsible Research and Innovation (RRI)

- La investigación y la innovación responsables son acciones clave del objetivo **"Ciencia por, con y para la sociedad"**.

- Incluyen acciones sobre los siguientes elementos temáticos:
  - **Participación ciudadana**
  - **Acceso abierto**
  - **Igualdad de género**
  - **Ética**
  - **Educación científica**

https://ec.europa.eu/programmes/horizon2020/en/h2020-section/responsible-research-innovation

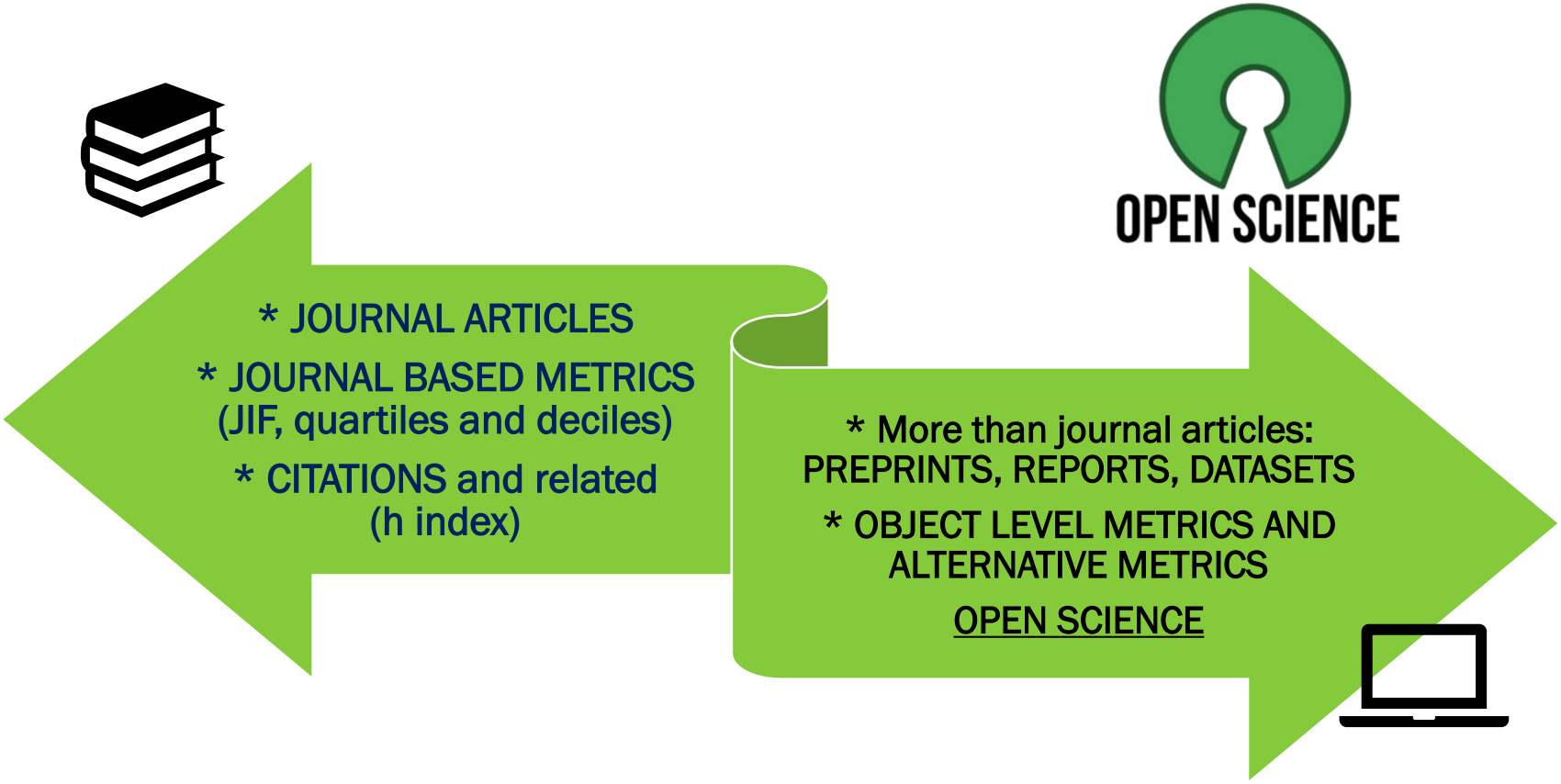


https://op.europa.eu/s/owD3

Qué es la ciencia abierta

- Es el movimiento que busca la **apertura de la investigación científica (métodos, instrumentos, datos, etc.) para beneficio de toda la sociedad**.
- Se basa en hacer que la investigación científica en su conjunto (no solo las publicaciones, los datos, las muestras físicas y el software, sino también los procesos que generan la ciencia) y también su difusión sean accesibles a todos los niveles de una sociedad, sean investigadores, aficionados o profesionales.
- Según la Comisión Europea (2016), la ciencia abierta representa un **nuevo enfoque del proceso científico basado en el trabajo cooperativo y nuevas formas de difundir el conocimiento** mediante el uso de tecnologías digitales y nuevas herramientas colaborativas.

Ciencia abierta y el cambio de paradigma



Cambio de paradigma

Ciencia Cerrada	Ciencia Abierta
• Basada en **publicar artículos** científicos	• Basada en **publicar cualquier resultado de investigación**
• Ciencia **individualista**	• Ciencia **colaborativa**
• Solo unos **pocos tienen acceso** a los resultados de investigación	• **Toda la sociedad** tiene acceso a los resultados de investigación
• Ciencia **vertical / especializada**	• Ciencia horizontal, **interdisciplinar**
• **Sin participación** ciudadana	• Ciencia Ciudadana **con y para la sociedad**
• Sistema de crédito basado en el **impacto de las publicaciones**	• Sistema de crédito basado en el **impacto del investigador**
• **Factor de Impacto** (FI – JCR)	• **Nuevas métricas** (Next Generation Metrics). Nuevos indicadores

(Adaptado de Méndez, 2019)

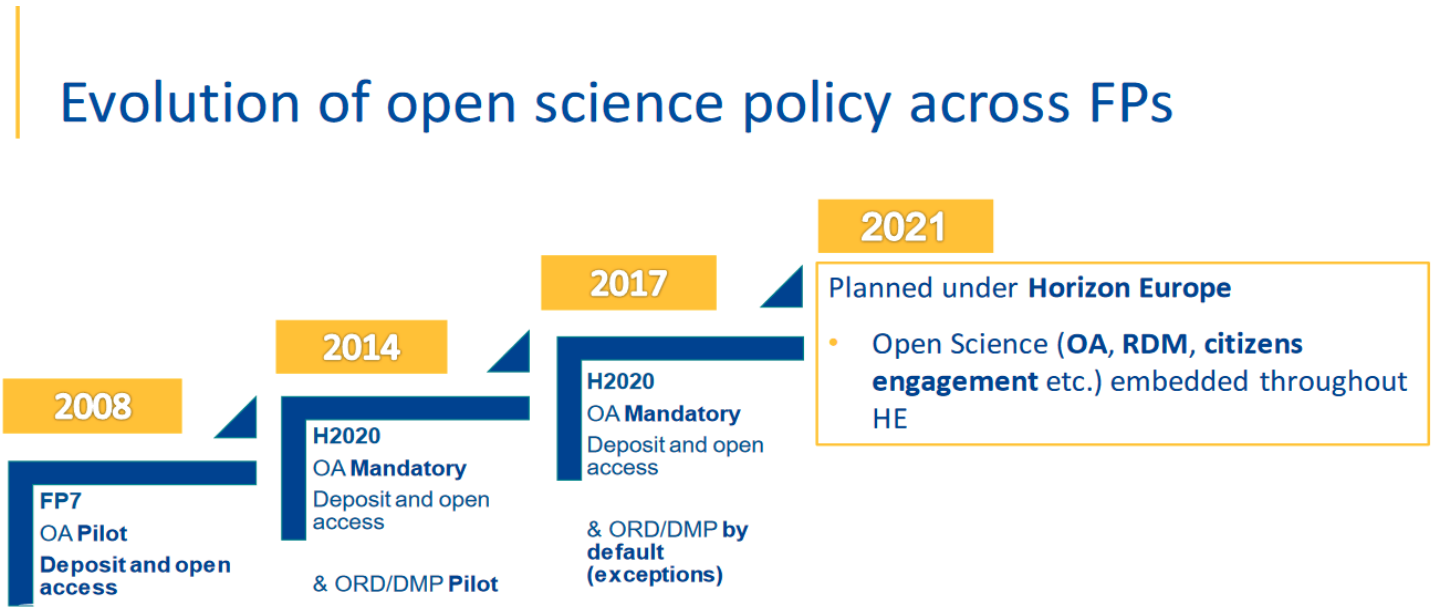
Alonso-Arévalo, J. (2019) El conocimiento es de todos y para todos ¿Qué es y qué implica la Ciencia Abierta? Universo Abierto. https://universoabierto.org/2019/09/30/el-conocimiento-es-de-todos-y-para-todos-que-es-y-que-implica-la-ciencia-abierta/

“La ciencia abierta es esto: un cambio de paradigma en la manera de hacer ciencia. El cambio no está en lo que se hace, sino en cómo se hace”

ANGLADA L, ABADAL E. (2018) “¿Qué es la ciencia abierta?”. Anuario ThinkEPI, v. 12, pp. 292-298

Marco legal. Requisitos europeos y nacionales

Del acceso abierto a la ciencia abierta en Horizonte Europa



Stakeholder workshop: Novelties in Horizon Europe MGA (9 October 2020). https://ec.europa.eu/research/participants/docs/h2020-funding-guide/other/event201009.htm

Ciencia abierta a lo largo del programa

Ciencia abierta → **Mejor difusión y explotación de los resultados de investigación e innovación, así como apoyo a la participación activa de la sociedad**

Acceso abierto obligatorio para las publicaciones: los beneficiarios se asegurarán de que ellos o los autores conservan los derechos de la propiedad intelectual necesarios para cumplir los requisitos de acceso abierto.

Garantizar el acceso abierto a los datos de investigación: de conformidad con el principio «tan abierto como sea posible y tan cerrado como sea necesario»; plan obligatorio de gestión de datos para datos FAIR (fáciles de encontrar, accesibles, interoperables y reutilizables) y datos de investigación abiertos.

- Apoyo a las habilidades de los investigadores en materia de ciencia abierta, así como sistemas de recompensa.
- Uso de la Nube Europea de la Ciencia Abierta.

Comisión Europea

https://ec.europa.eu/info/sites/default/files/research_and_innovation/strategy_on_research_and_innovation/presentations/horizon_europe_es_invertir_para_dar_forma_a_nuestro_futuro.pdf

Normativa Horizonte Europa

COMMUNICATION, DISSEMINATION, OPEN SCIENCE AND VISIBILITY
(ARTICLE 17 del General Model Grant Agreement).

Horizon Europe (HORIZON)
Euratom Research and Training Programme (EURATOM)

General Model Grant Agreement
EIC Accelerator Contract

(HE MGA — Multi & Mono)

Version 1.1
15 April 2022

Disclaimer
This document is aimed at assisting applicants. It shows the full range of provisions that may be applied to this type of agreement, and is provided for information purposes only. The legally binding agreement will be that which is signed by the parties in the system.

- El mandato de acceso abierto se extiende a libros y otras publicaciones de formato extenso.
- Acceso abierto inmediato, sin embargos, no se reembolsan las tarifas de las publicaciones en revistas híbridos.
- Se debe proporcionar **información sobre todos los demás objetos, herramientas e instrumentos académicos que se necesitan para validar las conclusiones de la publicación**.
- Mandato de intercambio de datos abierto a través de repositorios reconocidos.
- Énfasis en la gestión **de datos de investigación FAIR** y los **DMP**, al igual que en H2020, los costos de RDM son elegibles para reembolso.
- Publicación en OA a través de Open Research Europe.

Los beneficiarios (o autores) deben conservar suficientes derechos de propiedad intelectual para cumplir con los requisitos de acceso abierto.

Los metadatos de los datos o publicaciones depositados deben estar en abierto bajo una licencia CC de Dominio Público Común (CC By) o equivalente (en la medida en que se protejan los intereses o limitaciones legítimos), de acuerdo con los principios FAIR y proporcionar información al menos sobre lo siguiente:

- Para publicaciones: autor (es), título, fecha de publicación, lugar de publicación; para conjuntos de datos: descripción, fecha de depósito, autor (es), lugar y embargo);
- Datos sobre la financiación; nombre, acrónimo y número del proyecto;
- Términos de licencia;
- Identificadores persistentes para publicaciones, conjuntos de datos, autores y, si es posible, para sus organizaciones y para subvención.
- Cuando corresponda, los metadatos deben incluir identificadores persistentes para publicaciones relacionadas y otros resultados de la investigación.

NORMATIVA A NIVEL NACIONAL

Ley 14/2011, de 1 de junio, de la Ciencia, la Tecnología y la Innovación, que en su artículo 37, “Difusión en acceso abierto”,

Artículo 37. Difusión en acceso abierto.

1. Los agentes públicos del Sistema Español de Ciencia, Tecnología e Innovación impulsarán el desarrollo de repositorios, propios o compartidos, de acceso abierto a las publicaciones de su personal de investigación, y establecerán sistemas que permitan conectarlos con iniciativas similares de ámbito nacional e internacional.

2. El personal de investigación cuya actividad investigadora esté financiada mayoritariamente con fondos de los Presupuestos Generales del Estado hará pública una versión digital de la versión final de los contenidos que le hayan sido aceptados para publicación en publicaciones de investigación seriadas o periódicas, tan pronto como resulte posible, pero no más tarde de doce meses después de la fecha oficial de publicación.

3. La versión electrónica se hará pública en repositorios de acceso abierto reconocidos en el campo de conocimiento en el que se ha desarrollado la investigación, o en repositorios institucionales de acceso abierto.

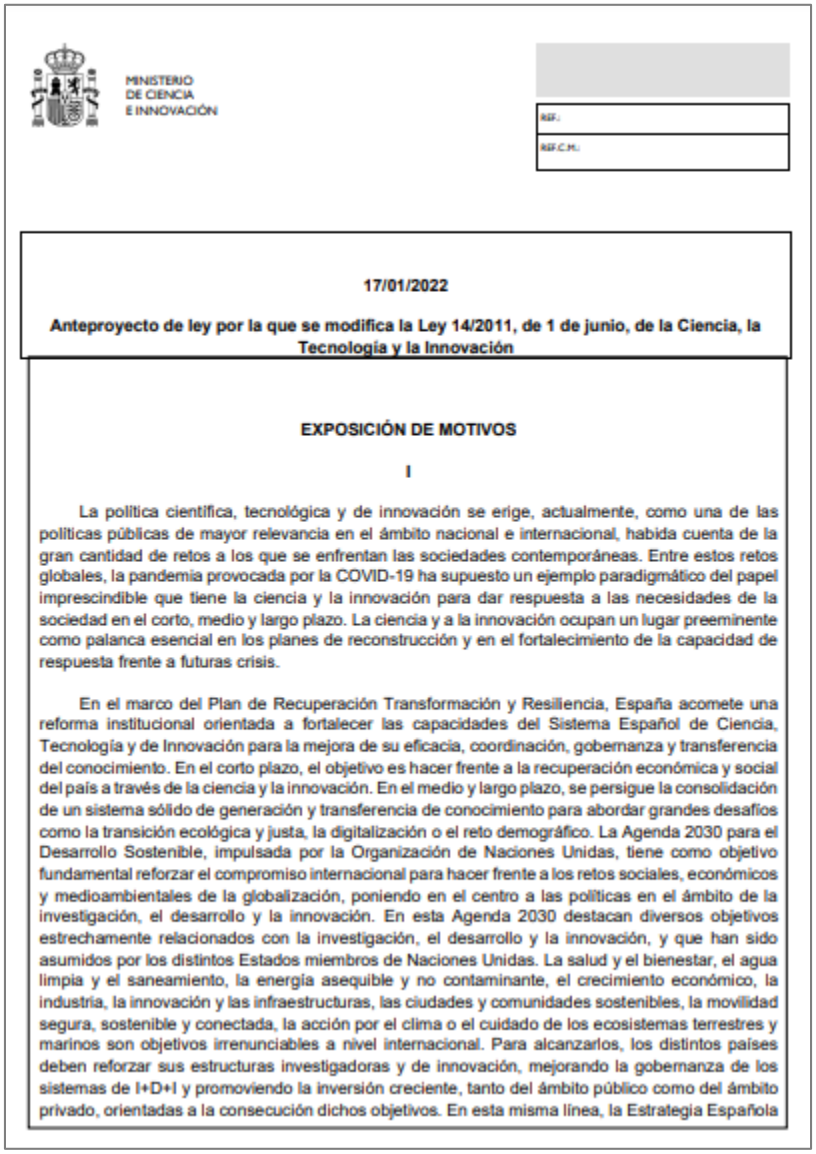
4. La versión electrónica pública podrá ser empleada por las Administraciones Públicas en sus procesos de evaluación.

5. El Ministerio de Ciencia e Innovación facilitará el acceso centralizado a los repositorios, y su conexión con iniciativas similares nacionales e internacionales.

6. Lo anterior se entiende sin perjuicio de los acuerdos en virtud de los cuales se hayan podido atribuir o transferir a terceros los derechos sobre las publicaciones, y no será de aplicación cuando los derechos sobre los resultados de la actividad de investigación, desarrollo e innovación sean susceptibles de protección.

Enero de 2022: Anteproyecto de ley de modificación de la Ley 14/2011, de 1 de junio, de la Ciencia, la Tecnología y la Innovación

Para fomentar la ciencia abierta al servicio de la sociedad, se reconoce el valor de la ciencia como bien común, y se establece que el Ministerio de Ciencia e Innovación promoverá iniciativas orientadas a facilitar el libre acceso a los datos generados por la investigación, a desarrollar infraestructuras y plataformas abiertas, y a fomentar la participación abierta de la sociedad civil en los procesos científicos.



MINISTERIO
DE CIENCIA
E INNOVACIÓN

17/01/2022

Anteproyecto de ley por la que se modifica la Ley 14/2011, de 1 de junio, de la Ciencia, la Tecnología y la Innovación

EXPOSICIÓN DE MOTIVOS

I

La política científica, tecnológica y de innovación se erige, actualmente, como una de las políticas públicas de mayor relevancia en el ámbito nacional e internacional, habida cuenta de la gran cantidad de retos a los que se enfrentan las sociedades contemporáneas. Entre estos retos globales, la pandemia provocada por la COVID-19 ha supuesto un ejemplo paradigmático del papel imprescindible que tiene la ciencia y la innovación para dar respuesta a las necesidades de la sociedad en el corto, medio y largo plazo. La ciencia y a la innovación ocupan un lugar preeminente como palanca esencial en los planes de reconstrucción y en el fortalecimiento de la capacidad de respuesta frente a futuras crisis.

En el marco del Plan de Recuperación Transformación y Resiliencia, España acomete una reforma institucional orientada a fortalecer las capacidades del Sistema Español de Ciencia, Tecnología y de Innovación para la mejora de su eficacia, coordinación, gobernanza y transferencia del conocimiento. En el corto plazo, el objetivo es hacer frente a la recuperación económica y social del país a través de la ciencia y la innovación. En el medio y largo plazo, se persigue la consolidación de un sistema sólido de generación y transferencia de conocimiento para abordar grandes desafíos como la transición ecológica y justa, la digitalización o el reto demográfico. La Agenda 2030 para el Desarrollo Sostenible, impulsada por la Organización de Naciones Unidas, tiene como objetivo fundamental reforzar el compromiso internacional para hacer frente a los retos sociales, económicos y medioambientales de la globalización, poniendo en el centro a las políticas en el ámbito de la investigación, el desarrollo y la innovación. En esta Agenda 2030 destacan diversos objetivos estrechamente relacionados con la investigación, el desarrollo y la innovación, y que han sido asumidos por los distintos Estados miembros de Naciones Unidas. La salud y el bienestar, el agua limpia y el saneamiento, la energía asequible y no contaminante, el crecimiento económico, la industria, la innovación y las infraestructuras, las ciudades y comunidades sostenibles, la movilidad segura, sostenible y conectada, la acción por el clima o el cuidado de los ecosistemas terrestres y marinos son objetivos irrenunciables a nivel internacional. Para alcanzarlos, los distintos países deben reforzar sus estructuras investigadoras y de innovación, mejorando la gobernanza de los sistemas de I+D+I y promoviendo la inversión creciente, tanto del ámbito público como del ámbito privado, orientadas a la consecución dichos objetivos. En esta misma línea, la Estrategia Española

https://bit.ly/3nx57cH

Treinta y ocho. Se modifican el título y los apartados 5 y 6 del artículo 37. "Artículo 37. Ciencia abierta."

5. El Ministerio de Ciencia e Innovación facilitará el **acceso centralizado a los repositorios** y su conexión con iniciativas similares nacionales e internacionales, e **impulsará la ciencia abierta en la Estrategia Española de Ciencia, Tecnología e Innovación**, reconociendo el valor de la ciencia como bien común.

6. Además del acceso abierto, y siempre con el objetivo de hacer la ciencia más abierta, accesible, eficiente, transparente y beneficiosa para la sociedad, el Ministerio de Ciencia e Innovación **promoverá también otras iniciativas orientadas a facilitar el libre acceso a los datos generados por la investigación** (datos abiertos), a desarrollar infraestructuras y plataformas abiertas, y a fomentar la participación abierta de la sociedad civil en los procesos científicos, tal como se desarrolla en el artículo 38.

!

IV Plan de Gobierno Abierto 2020-2024

GOBIERNO ABIERTO

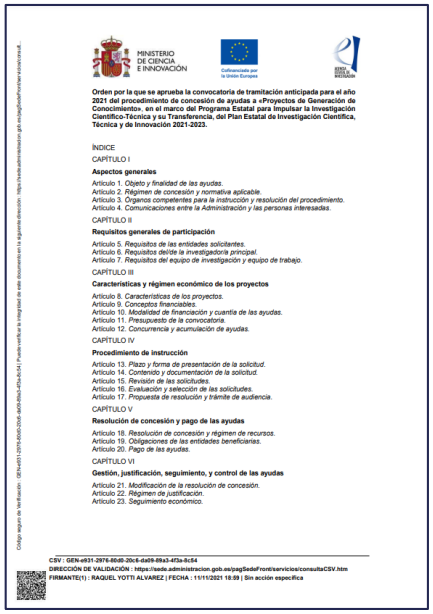
ii. Fomentar la I+D+I y su transferencia

La ciencia excelente y abierta constituye uno de los pilares del ***Objetivo 4*** (Generación de conocimiento y liderazgo científico). El impulso a un modelo de **Ciencia Abierta** favorecerá la generación de conocimiento de alta calidad e impacto, así como su transmisión a la sociedad, elemento que se relaciona directamente con el ***Eje de actuación 5*** (fomentar y apoyar la generación de capacidades científicas e innovadoras en los agentes del SECTI). Para asegurar el nivel de excelencia y su presencia internacional, tenemos que dotar a la ciencia española de **recursos humanos y materiales**. De esta manera potenciaremos, también, las capacidades estratégicas de nuestro país, que permitirán aplicar el conocimiento científico e innovador al desarrollo de nuevas tecnologías. Se favorecerá el **relevo generacional** fomentando las vocaciones científicas y tecnológicas, ofreciendo oportunidades a los jóvenes talentos, y asegurando, la asignación presupuestaria necesaria para el desarrollo de los proyectos de I+D+I. Igualmente, se potenciará el **liderazgo tecnológico**, la actividad **inter y multidisciplinar** y el abordaje de áreas y tecnologías **disruptivas**, manteniendo, a la vez, líneas de trabajo en ámbitos que den solidez al SECTI y favorezcan a la industria como usuaria final de las mismas. Se fomentará la interacción entre organismos públicos de investigación, universidades, centros tecnológicos y empresas innovadoras para favorecer la transferencia de conocimiento y la colaboración efectiva entre agentes.

El apoyo a la **ciencia de excelencia**, elemento básico del modelo de EECTI, fomentará, en consonancia con las directrices de la UE, el acceso abierto a los resultados de investigación, permitiendo que los datos sean accesibles, interoperables y reutilizables (su acrónimo en inglés *FAIR*). La difusión en el ámbito científico, junto al esfuerzo llevado a cabo por los repositorios abiertos, facilitará la accesibilidad a los avances científicos y fomentará la divulgación y comunicación científica hacia la sociedad, objetivo que se persigue en el ***Eje de actuación 14***.

Estrategia Española de Ciencia, Tecnología e Innovación 2021-2027

Orden por la que se aprueba la convocatoria de tramitación anticipada para el año 2021 del procedimiento de concesión de ayudas a «Proyectos de Generación de Conocimiento», en el marco del Programa Estatal para Impulsar la Investigación Científico-Técnica y su Transferencia, del Plan Estatal de Investigación Científica, Técnica y de Innovación 2021-2023.



MINISTERIO DE CIENCIA E INNOVACIÓN

Cofinanciado por la Unión Europea

Orden por la que se aprueba la convocatoria de tramitación anticipada para el año 2021 del procedimiento de concesión de ayudas a «Proyectos de Generación de Conocimiento», en el marco del Programa Estatal para Impulsar la Investigación Científico-Técnica y su Transferencia, del Plan Estatal de Investigación Científica, Técnica y de Innovación 2021-2023.

ÍNDICE

CAPÍTULO I

Aspectos generales

Artículo 1. *Objeto y finalidad de las ayudas.*
Artículo 2. *Régimen de concesión y normativa aplicable.*
Artículo 3. *Órganos competentes para la instrucción y resolución del procedimiento.*
Artículo 4. *Comunicaciones entre la Administración y las personas interesadas.*

CAPÍTULO II

Requisitos generales de participación

Artículo 5. *Requisitos de las entidades solicitantes.*
Artículo 6. *Requisitos delde la investigadoria principal.*
Artículo 7. *Requisitos del equipo de investigación y equipo de trabajo.*

CAPÍTULO III

Características y régimen económico de los proyectos

Artículo 8. *Características de los proyectos.*
Artículo 9. *Conceptos financiables.*
Artículo 10. *Modalidad de financiación y cuantía de las ayudas.*
Artículo 11. *Presupuesto de la convocatoria.*
Artículo 12. *Concurrencia y acumulación de ayudas.*

CAPÍTULO IV

Procedimiento de instrucción

Artículo 13. *Plazo y forma de presentación de la solicitud.*
Artículo 14. *Contenido y documentación de la solicitud.*
Artículo 15. *Revisión de las solicitudes.*
Artículo 16. *Evaluación y selección de las solicitudes.*
Artículo 17. *Propuesta de resolución y trámite de audiencia.*

CAPÍTULO V

Resolución de concesión y pago de las ayudas

Artículo 18. *Resolución de concesión y régimen de recursos.*
Artículo 19. *Obligaciones de las entidades beneficiarias.*
Artículo 20. *Pago de las ayudas.*

CAPÍTULO VI

Gestión, justificación, seguimiento, y control de las ayudas

Artículo 21. *Modificación de la resolución de concesión.*
Artículo 22. *Régimen de justificación.*
Artículo 23. *Seguimiento económico.*

CSV : GEN-e831-2976-80d0-20c6-da09-89a3-4f3a-8c54
DIRECCIÓN DE VALIDACIÓN : https://sede.administracion.gob.es/pagSedeFront/servicios/consultaCSV.htm
FIRMANTE(1) : RAQUEL YOTTI ALVAREZ | FECHA : 11/11/2021 18:59 | Sin acción específica

3. Impacto
3.1 Impacto científico-técnico de los resultados esperados

Se valorará la contribución de los resultados a las necesidades de la prioridad temática seleccionada, así como aquellos resultados que permitan fortalecer la base del conocimiento científico-técnico de carácter inter y multidisciplinar.
Se valorará el plan de publicaciones científico-técnicas, presentaciones y comunicaciones a congresos y otros foros principalmente internacionales; los planes de transferencia y valorización de los resultados, patentes y otros resultados incluidos en la propuesta, **y en caso de que resulte pertinente, el plan de gestión de datos de investigación asociados a los resultados**.

https://www.aei.gob.es/sites/default/files/convocatory_info/2021-12/Convocatoria_PID2021_.pdf

ANEXO VII. Descripción del contenido de la memoria científico-técnica del proyecto

4. Impacto científico-técnico.

Se incluirán, entre otros, los siguientes aspectos:

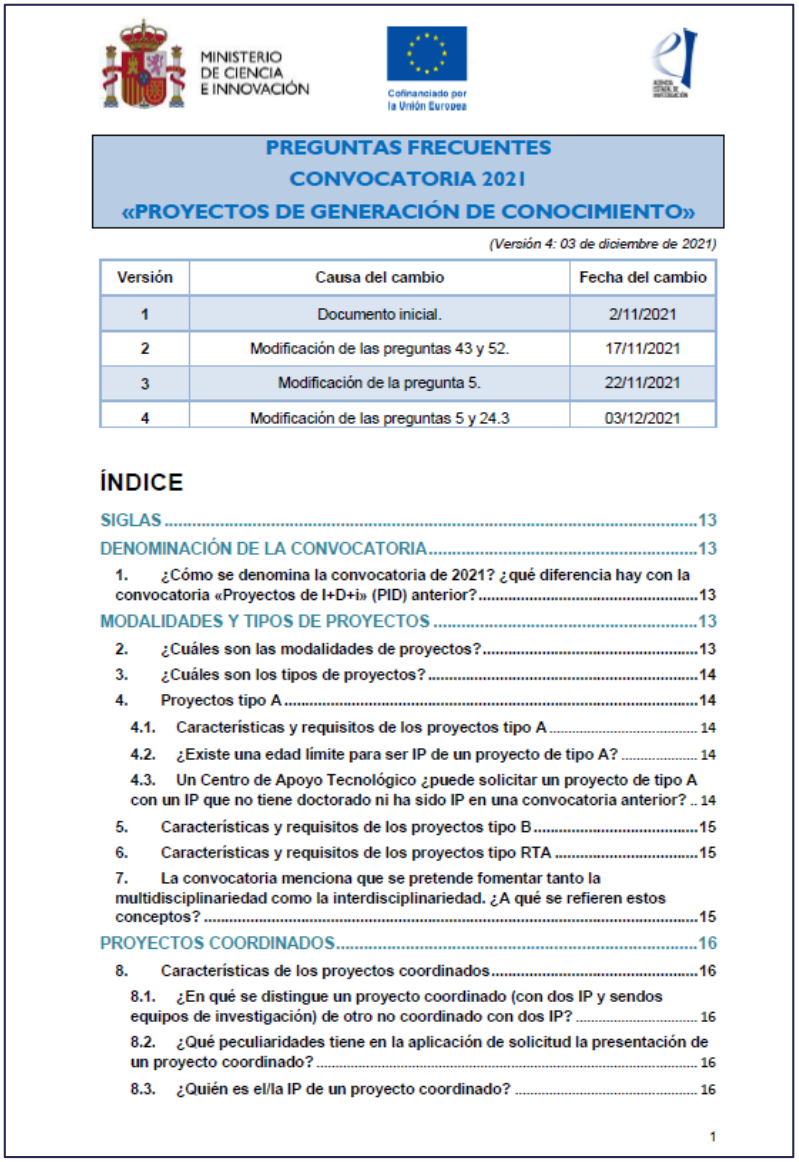
- Descripción del impacto esperado de los resultados derivados del proyecto para lograr avances en el conocimiento y/o desarrollo tecnológico, y solucionar problemas o necesidades en la prioridad temática a la que se dirige.
- Plan de difusión e internacionalización, en su caso, de los resultados, incluyendo la previsión sobre publicaciones y datos de acceso en abierto.
- En su caso, plan de transferencia y/o valorización de los resultados, así como relación con entidades o usuarios finales, que pudieran estar interesados en los mismos.
- Resumen del plan de gestión de datos previsto durante y al finalizar el proyecto.

89.6. Datos de Investigación - ¿Qué es un Plan de Gestión de Datos?

Un Plan de Gestión de Datos (PGD o DMP, *Data Management Plan*, por sus siglas en inglés) es una declaración formal que describe los datos que se utilizan y producen durante el curso de las actividades de investigación. El PGD establece cómo se van a gestionar los datos en un proyecto de investigación, tanto durante el proyecto como una vez que éste ha finalizado.

El PGD define la metodología de extracción de la información, la unidad de medida, los procesos de creación, limpieza y análisis, el formato, el software de lectura y procesamiento, las licencias y permisos de acceso y (re)utilización, su sistema de control de versiones, su autoría, su difusión, y su lugar de almacenamiento y preservación a largo plazo. Este plan se creará al comienzo del proyecto de investigación financiado y, según la naturaleza y evolución del proyecto, puede que no sea un documento definitivo, sino que variará, se completará y/o modificará a lo largo del ciclo de vida de la investigación.

PREGUNTAS FRECUENTES CONVOCATORIA 2021 «PROYECTOS DE GENERACIÓN DE CONOCIMIENTO»



MINISTERIO DE CIENCIA E INNOVACIÓN

Cofinanciado por la Unión Europea

PREGUNTAS FRECUENTES
CONVOCATORIA 2021
«PROYECTOS DE GENERACIÓN DE CONOCIMIENTO»

(Versión 4: 03 de diciembre de 2021)

Versión	Causa del cambio	Fecha del cambio
1	Documento inicial.	2/11/2021
2	Modificación de las preguntas 43 y 52.	17/11/2021
3	Modificación de la pregunta 5.	22/11/2021
4	Modificación de las preguntas 5 y 24.3	03/12/2021

ÍNDICE

SIGLAS 13
DENOMINACIÓN DE LA CONVOCATORIA 13
1. ¿Cómo se denomina la convocatoria de 2021? ¿qué diferencia hay con la convocatoria «Proyectos de I+D+i» (PID) anterior? 13
MODALIDADES Y TIPOS DE PROYECTOS 13
2. ¿Cuáles son las modalidades de proyectos? 13
3. ¿Cuáles son los tipos de proyectos? 14
4. Proyectos tipo A 14
4.1. Características y requisitos de los proyectos tipo A 14
4.2. ¿Existe una edad límite para ser IP de un proyecto de tipo A? 14
4.3. Un Centro de Apoyo Tecnológico ¿puede solicitar un proyecto de tipo A con un IP que no tiene doctorado ni ha sido IP en una convocatoria anterior? 14
5. Características y requisitos de los proyectos tipo B 15
6. Características y requisitos de los proyectos tipo RTA 15
7. La convocatoria menciona que se pretende fomentar tanto la multidisciplinariedad como la interdisciplinariedad. ¿A qué se refieren estos conceptos? 15
PROYECTOS COORDINADOS 16
8. Características de los proyectos coordinados 16
8.1. ¿En qué se distingue un proyecto coordinado (con dos IP y sendos equipos de investigación) de otro no coordinado con dos IP? 16
8.2. ¿Qué peculiaridades tiene en la aplicación de solicitud la presentación de un proyecto coordinado? 16
8.3. ¿Quién es el/la IP de un proyecto coordinado? 16

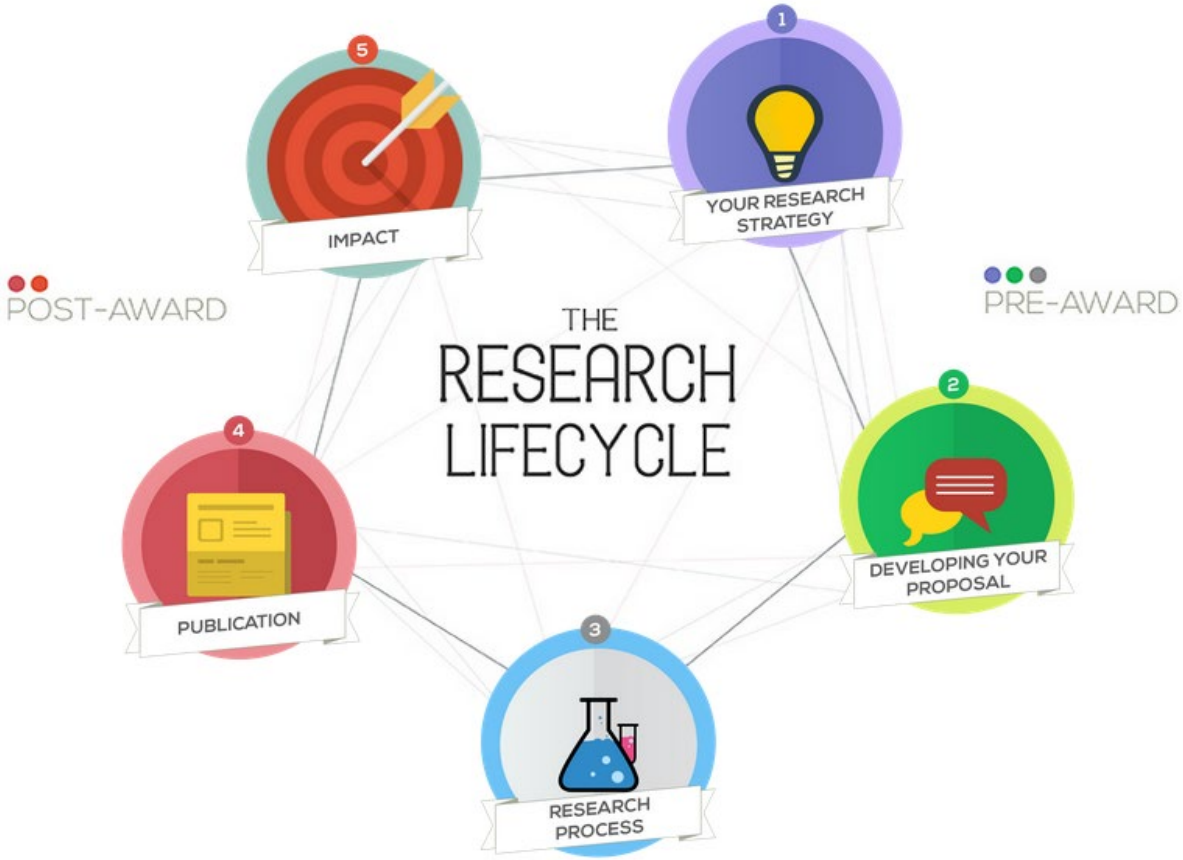
1

85. Durante la ejecución del proyecto se harán experimentos con animales en 3 sitios ¿Se debe contar con certificados de aprobación de los procedimientos de los 3 centros? 40
86. ¿Debemos iniciar el trámite correspondiente "Plan de depósito inicial de encuestas de I+D" en el CIS? 40
87. ¿Se permite la inclusión de EPO (Ente Promotor Observador) en la memoria científico-técnica? 40
88. Los documentos adjuntos ¿pueden aportarse en cualquier formato digital? 40
89. ¿Qué características debe tener la memoria científico-técnica? 41
89.1. ¿Cuántas memorias científico-técnicas hay en los proyectos coordinados? 41
89.2. Si no he tenido tiempo de rellenar algunos apartados de la memoria científico-técnica, ¿podré adjuntarla corregida en otro momento? 41
89.3. Si no adjunto la memoria científico-técnica en el modelo normalizado ¿Es posible subsanar este defecto? 41
89.4. ¿Qué sucede si no se presenta la memoria en lengua inglesa cuando es obligatorio hacerlo? 41
89.5. Datos de Investigación - ¿Que son los datos de investigación? 42
89.6. Datos de Investigación - ¿Qué es un Plan de Gestión de Datos? 42
89.7. Datos de Investigación - ¿Debo incluir en la solicitud de mi proyecto el Plan de Gestión de Datos (PGD)? 42
89.8. Datos de Investigación - ¿Cuándo debo enviar el Plan de Gestión de Datos? 42
89.9. Datos de Investigación - ¿Qué significa que los datos de investigación sean FAIR? 43
89.10. Datos de Investigación - ¿Es lo mismo datos abiertos que datos "FAIR"? 43
89.11. Datos de Investigación - ¿Qué datos de investigación se deben depositar? 43
89.12. Datos de Investigación - ¿Qué repositorios se podrán utilizar para depositar los datos? 43
89.13. ¿Es posible sustituir en la memoria las referencias bibliográficas por un enlace? Es decir, que cada vez que se cite un trabajo, esta cita lleve al *link* de ese trabajo en la web. 44
89.14. En las instrucciones para rellenar la memoria, en el apartado de capacidad formativa, se recomienda incluir una relación de tesis realizadas o en

https://www.aei.gob.es/sites/default/files/convocatory_info/2021-12/FAQs_PID2021_v7_sc.pdf

Datos de investigación y datos abiertos

El ciclo de investigación



El ciclo de vida de la investigación. Fuente: BU Research Blog. Research Lifecycle
http://blogs.bournemouth.ac.uk/research/research-lifecycle/

¿Qué son los datos de investigación?

"Datos que se utilizan como **fuentes primarias para respaldar** preguntas técnicas o científicas, investigaciones, o actividades académicas y artísticas, y datos que se utilizan como **evidencia en el proceso de investigación** y son comúnmente aceptados en la comunidad de investigadores como necesarios para **validar los hallazgos y resultados de la investigación**.

Todos los demás contenidos digitales y no digitales tienen el potencial de convertirse en datos de investigación.

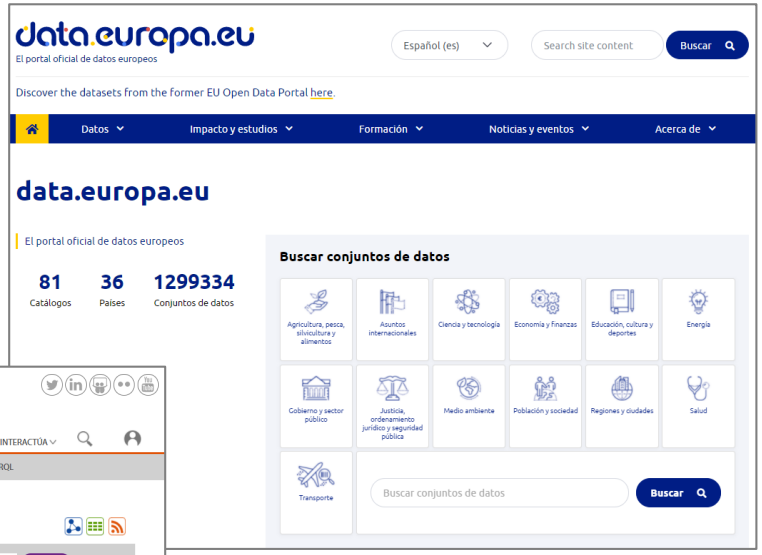
Consortia Advancing Standards in Research Administration Information (CASRAI)
https://casrai.org/term/research-data/

Qué son los datos abiertos

Los datos abiertos no se refieren de forma exclusiva a los datos de investigación, también tienen especial importancia los datos abiertos en el contexto de los **datos públicos y gubernamentales**.

http://data.un.org/

https://datos.gob.es/es/catalogo



https://data.europa.eu/es

Los datos de investigación abiertos

Son datos de investigación que:

- Están disponibles de manera libre en internet;
- Permiten ser descargados, copiados, analizados, usados , reprocesados, etc.;
- Sin barreras de tipo económico, legal o técnicas (aparte de tener acceso a internet).

Sparc Europe (2017) *What is Open Data?* [Online] http://sparceurope.org/what-we-do/open-data/what-is-open-data/

WORLD VIEW *A personal take on events*

Open your minds and share your results

An open approach is the best way to maximize the benefits of research for both scientists and the public, says **Geoffrey Boulton**.

There is a compelling case for having open access to scientific papers, to enhance the efficacy and reach of scientific communication. But important though this is, the open-access debate has drawn attention away from a deeper issue that is at the heart of the scientific process: that of 'open data'. In an attempt to focus much-needed attention on this subject, I chaired a group that produced *Science as an Open Enterprise*, a policy report from the Royal Society in London, published last week.

Open enquiry has been at the heart of science since the first scientific journals were printed in the seventeenth century. Publication of scientific theories — and the supporting experimental and observational data — permits others to identify errors, to reject or refine theories and to reuse data. Science's capacity for self-correction comes from this openness to scrutiny and challenge.

Modern techniques to gather, store and manipulate data make this more difficult. In the 1980s, I published a paper that presented seven hard-won data points showing the relationship between stress and velocity beneath a glacier. Two years ago, I was involved in an analogous experiment on the Antarctic ice sheet that created more than a billion times more data points. No journal could publish these data, so for them to be accessible, the only option was to deposit the information in a recognized repository, complete with metadata (data about data), and to signpost it in published papers, preferably through live links in the papers' electronic versions.

In the Royal Society report, we argue that this procedure must become the norm, required by journals and accepted by the scientific community as mandatory. As scientists, we have some way to go to achieve this. A recent study of the 50 highest-impact journals in biomedicine showed that only 22 required specific raw data to be made available as a condition of publication. Only 40% of papers fully adhered to the policy and only 9% had deposited the full raw data online (A. A. Alsheikh-Ali *et al. PLoS ONE* 6, e24357; 2011).

We also need to be open towards fellow citizens. The massive impact of science on our collective and individual lives has decreased the willingness of many to accept the pronouncements of scientists unless they can verify the strength of the underlying evidence for themselves. The furore surrounding 'Climategate' — rooted in the resistance of climate scientists to accede to requests from members of the public for data underlying some of the claims of climate science — was in part a motivation for the Royal Society's current report. It is vital that science is not seen to hide behind closed laboratory doors, but engages seriously with the public.

There is, of course, a problem in making data sets open to non-specialists. They are rarely in the form of an Excel spreadsheet, an illusion under which many politicians labour in their laudable but problematic calls for open data. True openness requires data to be not only accessible, but also intelligible, assessable (who produced the data, what are their qualifications, do they have conflicts of interest?) and reusable.

Everyone will benefit from a more open approach. The digital and communications revolutions bring opportunities for research that demand openness and a willingness to share data. These include the assembly of massive data sets from diverse sources, and linking them to allow data integration, dynamic updating and the manipulation of data within electronic publications. Such data-led science offers ways to explore massive data sets for patterns and relationships.

Yet this, too, presents a problem. Too often, we scientists seek patterns in data that reflect our preconceived ideas. And when we do publish the data, we too frequently publish only those that support these ideas. This cherry-picking is bad practice and should stop.

For example, there is strong evidence that the partial reporting of the results of clinical trials, skewed towards those with positive outcomes, obscures relationships between cause and effect. We should publish all the data, and we should explore them not just for preconceived relationships, but also for unexpected ones. Without rigorous use and manipulation of data, science merely creates myths. At the same time, communications technologies are displacing the printed page from its dominant role as the medium of scientific communication. This is already exploiting the collective intelligence of the scientific community and shifting the social dynamic of research towards collaboration.

This shift has not been mandated by research councils, governments or national academies, but is the consequence of scientists finding more productive and creative ways to do science. Pathfinder disciplines include bioinformatics, astronomy, mathematics, nanotechnology and social and health statistics. Likewise, to extend the reach and depth of these approaches does not need top-down orchestration. It merely requires some constraints to be removed and some enabling changes to be made.

What about costs? Data curation should be viewed as a necessary cost of research. Creative data generation should be a source of scholarly esteem and a criterion for promotion. We need a revolution in the role of the science library, with data scientists supporting the management of data strategies for both institutions and researchers. We need strategic funding to develop software tools to automate and simplify the creation and exploitation of data sets. And above all, we need scientists to accept that publicly funded research is a public resource. ■

SCIENCE'S CAPACITY FOR SELF-CORRECTION COMES FROM ITS OPENNESS TO SCRUTINY AND CHALLENGE.

NATURE.COM
Discuss this article online at:
go.nature.com/u5lazh

Geoffrey Boulton *is regius professor of geology emeritus at the University of Edinburgh, UK.*
e-mail: g.boulton@ed.ac.uk

© 2012 Macmillan Publishers Limited. All rights reserved

Boulton G (2012) *Open your minds and share your results*. Nature. 486(7404):441. doi: 10.1038/486441a1

BIBLIO MAD SALUD

Algunas iniciativas

Bits of Power: Issues in Global Access to Scientific Data 1997 (Libro)

OECD Principles and Guidelines on Access to Research Data, 2004, 2007
https://www.oecd.org/sti/inno/38500813.pdf

UK Funder Data Policies, empezaron en 2001, pero se incrementaron a partir de 2009
https://bit.ly/2ZnD0TR

National Science Foundation (NSF) Data Management Plan Requirements desde 2010
https://www.nsf.gov/open/

Royal Society Report 'Science as an Open Enterprise', 2012 https://royalsociety.org/topics-policy/projects/science-public-enterprise/report/

G8 Science Ministers Statement y el G8 Open Data Charter and Technical Appendix, junio de 2013
https://bit.ly/3vHf3TT

EC H2020 Open Data Policy Pilot, 2014. Adopción de los FAIR Data Principles, 2017 https://bit.ly/3jBIKAF

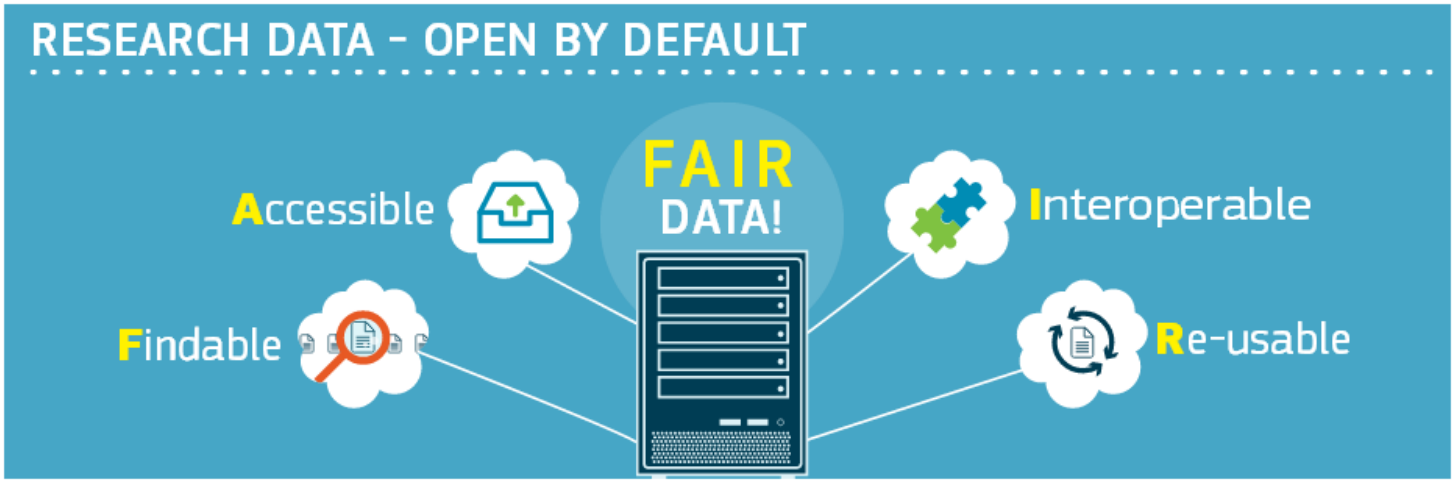
China State Council Law on Management of Research Data, Marzo 2018 https://bit.ly/3BpQFrb

EU Directive on Open Data and Public Sector Information 2019 https://digital-strategy.ec.europa.eu/en/policies/legislation-open-data

UN Conference Towards Global Open Science: Core Enabler of the UN 2030 Agenda 2019
https://www.un.org/en/library/OS21

Datos abiertos versus datos FAIR

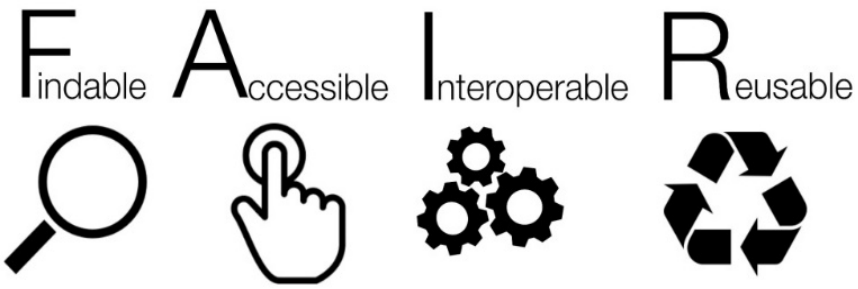
Datos FAIR

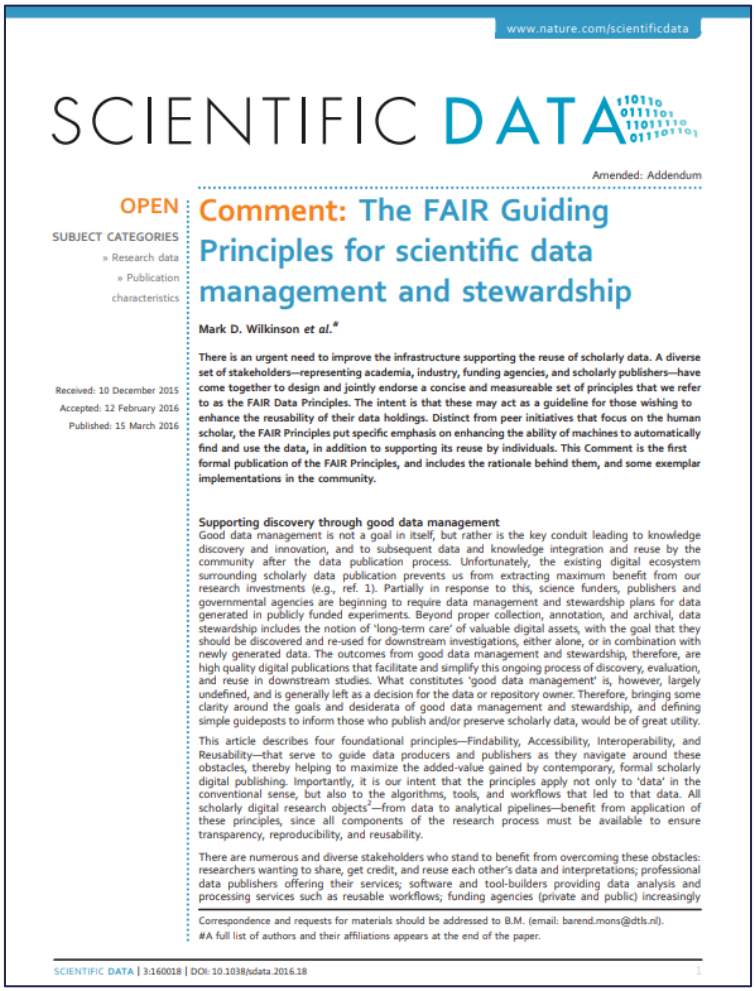


- se pueden encontrar en Internet,
- son accesibles (derechos y licencias claros),
- están en un formato utilizable,
- se identifican de una manera única y persistente para que se pueda hacer referencia a ellos.

http://ec.europa.eu/research/press/2016/pdf/opendata-infographic_072016.pdf

'Localizables'	se puede encontrar a través de metadatos, y son identificables y localizables mediante un mecanismo de identificación estándar.
'Acesibles'	siempre disponibles; incluso si los datos están restringidos, los metadatos están abiertos.
'Interoperables'	sintácticamente analizables y semánticamente comprensibles, lo que permite el intercambio y la reutilización de datos entre investigadores, instituciones, organizaciones o países.
'Reutilizables'	suficientemente descritos y compartidos con licencias mínimamente restrictivas, para permitir así una reutilización lo más amplia posible.





www.nature.com/scientificdata

SCIENTIFIC DATA

Amended: Addendum

OPEN

SUBJECT CATEGORIES
» Research data
» Publication characteristics

Comment: The FAIR Guiding Principles for scientific data management and stewardship

Mark D. Wilkinson *et al.*#

Received: 10 December 2015
Accepted: 12 February 2016
Published: 15 March 2016

There is an urgent need to improve the infrastructure supporting the reuse of scholarly data. A diverse set of stakeholders—representing academia, industry, funding agencies, and scholarly publishers—have come together to design and jointly endorse a concise and measureable set of principles that we refer to as the FAIR Data Principles. The intent is that these may act as a guideline for those wishing to enhance the reusability of their data holdings. Distinct from peer initiatives that focus on the human scholar, the FAIR Principles put specific emphasis on enhancing the ability of machines to automatically find and use the data, in addition to supporting its reuse by individuals. This Comment is the first formal publication of the FAIR Principles, and includes the rationale behind them, and some exemplar implementations in the community.

Supporting discovery through good data management

Good data management is not a goal in itself, but rather is the key conduit leading to knowledge discovery and innovation, and to subsequent data and knowledge integration and reuse by the community after the data publication process. Unfortunately, the existing digital ecosystem surrounding scholarly data publication prevents us from extracting maximum benefit from our research investments (e.g., ref. 1). Partially in response to this, science funders, publishers and governmental agencies are beginning to require data management and stewardship plans for data generated in publicly funded experiments. Beyond proper collection, annotation, and archival, data stewardship includes the notion of 'long-term care' of valuable digital assets, with the goal that they should be discovered and re-used for downstream investigations, either alone, or in combination with newly generated data. The outcomes from good data management and stewardship, therefore, are high quality digital publications that facilitate and simplify this ongoing process of discovery, evaluation, and reuse in downstream studies. What constitutes 'good data management' is, however, largely undefined, and is generally left as a decision for the data or repository owner. Therefore, bringing some clarity around the goals and desiderata of good data management and stewardship, and defining simple guideposts to inform those who publish and/or preserve scholarly data, would be of great utility.

This article describes four foundational principles—Findability, Accessibility, Interoperability, and Reusability—that serve to guide data producers and publishers as they navigate around these obstacles, thereby helping to maximize the added-value gained by contemporary, formal scholarly digital publishing. Importantly, it is our intent that the principles apply not only to 'data' in the conventional sense, but also to the algorithms, tools, and workflows that led to that data. All scholarly digital research objects[2]—from data to analytical pipelines—benefit from application of these principles, since all components of the research process must be available to ensure transparency, reproducibility, and reusability.

There are numerous and diverse stakeholders who stand to benefit from overcoming these obstacles: researchers wanting to share, get credit, and reuse each other's data and interpretations; professional data publishers offering their services; software and tool-builders providing data analysis and processing services such as reusable workflows; funding agencies (private and public) increasingly

Correspondence and requests for materials should be addressed to B.M. (email: barend.mons@dtls.nl).
#A full list of authors and their affiliations appears at the end of the paper.

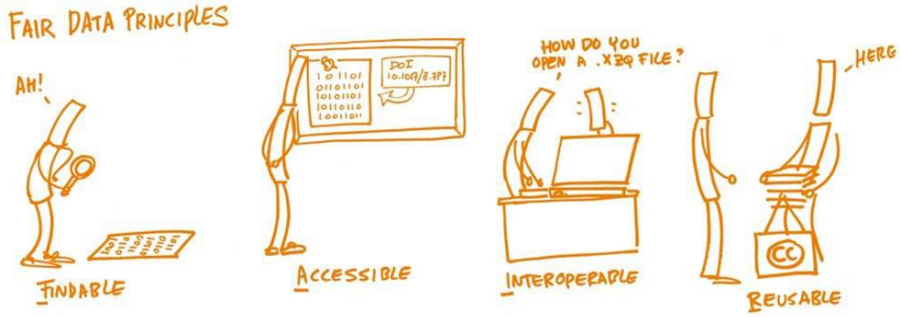
Proporcionar pautas para mejorar la capacidad de búsqueda, accesibilidad, interoperabilidad y reutilización de activos digitales.

Los principios enfatizan la capacidad de los sistemas computacionales para encontrar, acceder, interoperar y reutilizar datos con ninguna o mínima intervención humana).

Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* **3,** 160018 (2016). https://doi.org/10.1038/sdata.2016.18

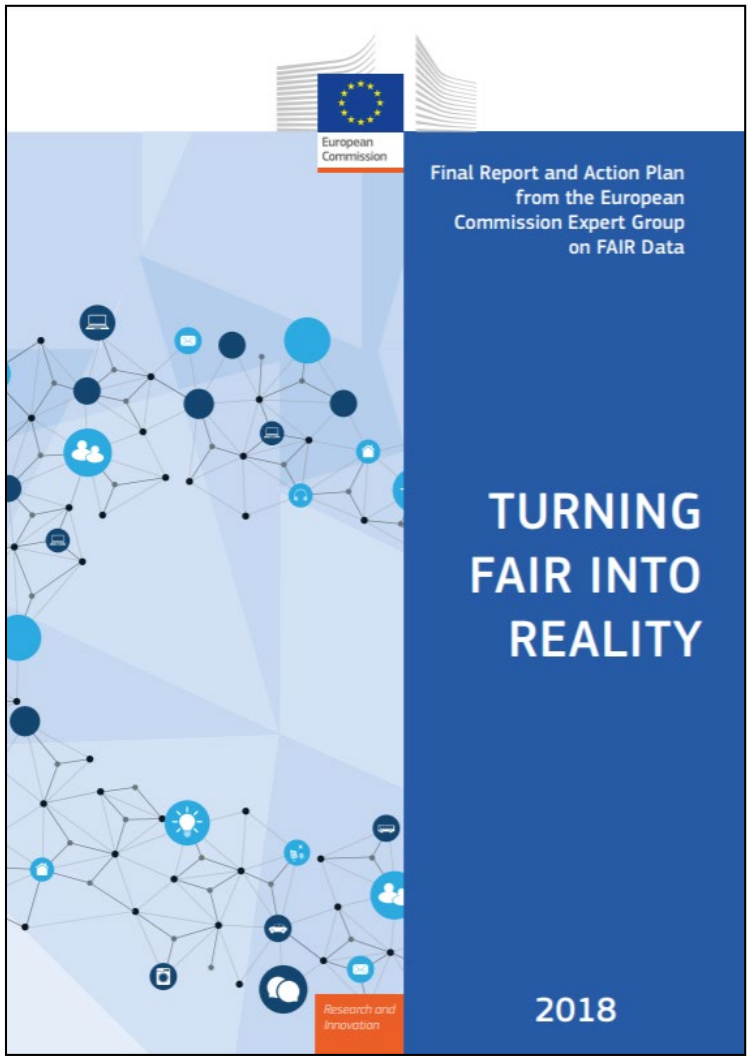
Tan abierto como sea posible, tan cerrado como sea necesario

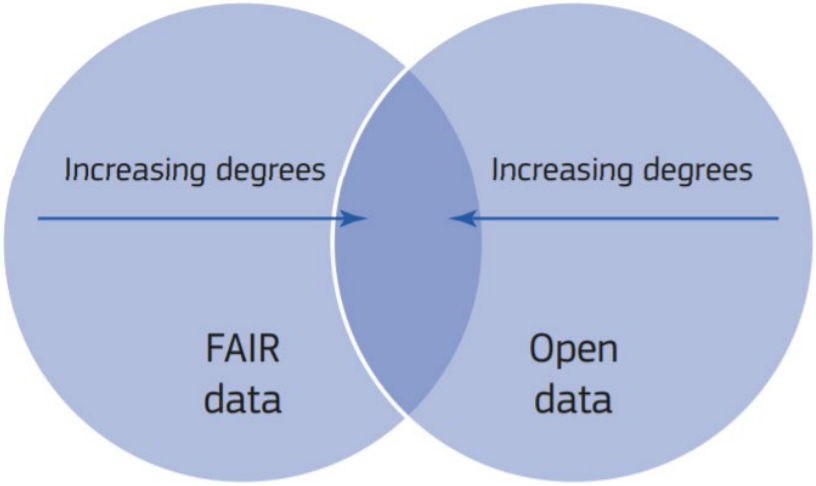
El principal objetivo de los principios FAIR es **preparar los datos de investigación para que tanto humanos y máquinas puedan recuperarlos y reutilizarlos de la mejor manera posible.**



La accesibilidad limitada, por ejemplo, debido a la protección de la privacidad de los datos, no contradice los principios FAIR.

Se han de permitir excepciones de mantener los datos con restricciones de acceso por ciertas razones: derechos de propiedad industrial, de protección datos personales, confidencialidad en materia de seguridad, por comprometer el objetivo del proyecto, o por otras razones a justificar.





https://ec.europa.eu/info/sites/info/files/turning_fair_into_reality_1.pdf

¿Cómo cumplir con los principios FAIR?

¿Qué herramientas podemos usar?

Checklist de verificación para comprobar el grado FAIRness de los datos de investigación

Jones, Sarah & Grootveld, Marjan (2017)
¿Cómo comprobar si sus datos son FAIR?
Traducido por Alicia Fátima Gómez (2021)
https://doi.org/10.5281/zenodo.5508099

¿Cómo comprobar si sus datos son FAIR?

Localizable

Cualquier ha de poder encontrar y localizar sus datos. Los metadatos enriquecidos deben estar disponibles en línea en un recurso de búsqueda y los datos han tener asignado un identificador persistente.

- ☐ Sus datos tienen un identificador persistente.
- ☐ Los datos están descritos con metadatos enriquecidos.
- ☐ Los metadatos están en línea en un recurso de búsqueda, por ejemplo, en un catálogo o en repositorio de datos.
- ☐ El registro de metadatos especifica el identificador persistente.

Accesible

Ha de ser posible que tanto personas como máquinas tengan acceso a sus datos, bajo condiciones específicas o restricciones cuando sea oportuno. ¡FAIR no significa que los datos sean siempre abiertos! Sin embargo, los metadatos sí han de estar disponibles, independientemente de que los datos sean accesibles públicamente o no.

- ☐ El ID persistente enlaza con los datos o metadatos asociados para que estos puedan ser recuperados.
- ☐ El protocolo mediante el cual se pueden recuperar los datos sigue estándares reconocidos, por ejemplo, http://.
- ☐ El procedimiento de acceso incluye pasos de autenticación y autorización en caso de que sea necesario.
- ☐ Los metadatos serán accesibles, siempre que sea posible, aunque los datos no sean.

Interoperable

Los datos y metadatos han de ajustarse a formatos y estándares reconocidos para permitir su combinación e intercambio.

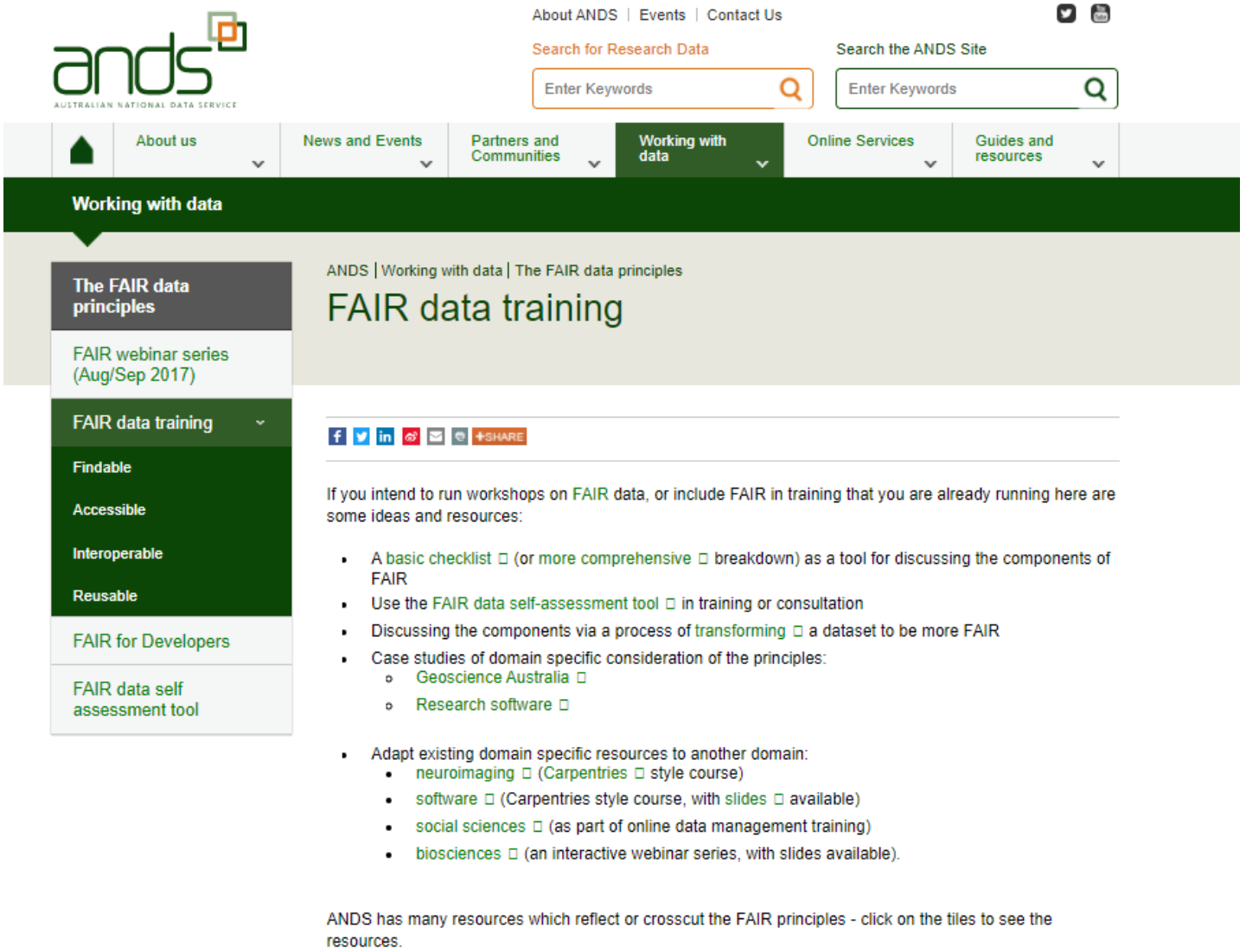
- ☐ Los datos se proporcionan en formatos conocidos y usados ampliamente, y preferiblemente abiertos.
- ☐ Los metadatos proporcionados siguen estándares reconocidos.
- ☐ Siempre que sea posible se utilizarán vocabularios controlados, palabras clave, tesauros u ontologías.
- ☐ Se incluyen referencias y enlaces a otros datos relacionados.

Reutilizable

Es importante acompañar los datos de documentación suficiente para respaldar la interpretación y reutilización de los datos. Los datos deben ajustarse a las normas de la comunidad y tener una licencia clara para que otros sepan qué tipos de reutilización están permitidos.

- ☐ Los datos son precisos y están bien descritos con todos los atributos relevantes.
- ☐ Los datos tienen una licencia de uso de datos clara y accesible.
- ☐ Queda claro cómo, por qué y quién ha creado y procesado los datos.
- ☐ Los datos y sus metadatos cumplen con los estándares que usa la comunidad de ese dominio concreto.

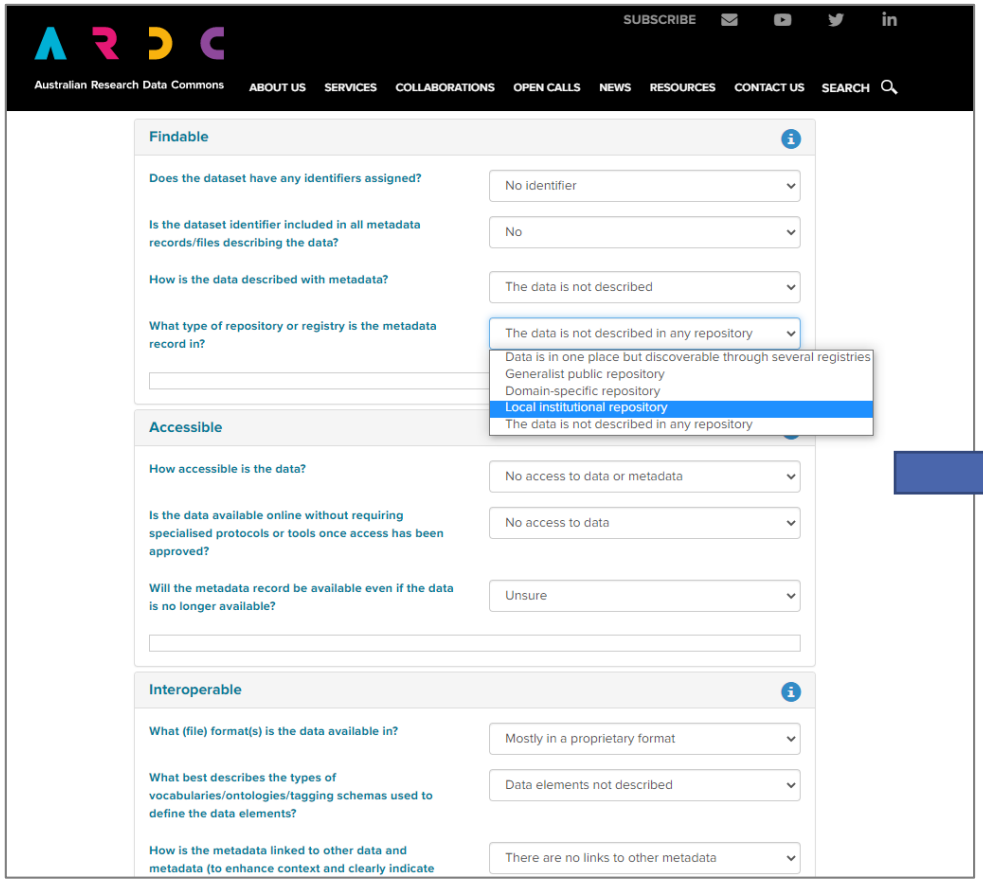
Traducido al español por Alicia Fátima Gómez del original 'How FAIR are your data?' checklist, CC-BY by Sarah Jones & Marjan Grootveld, EUDAT. Image CC-BY-SA by SangyaPundir

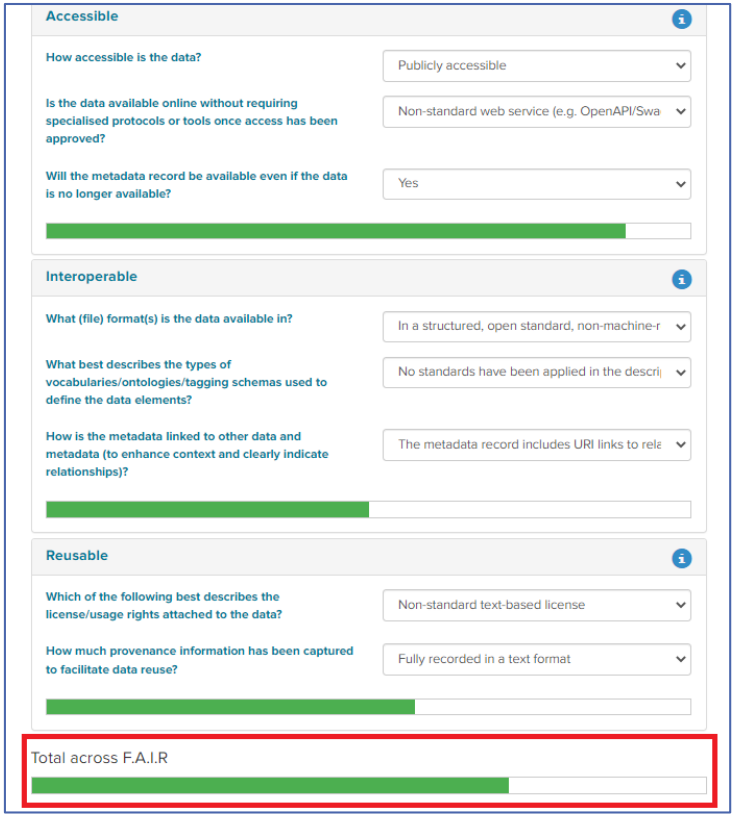


Australian National Data Service (ANDS). *FAIR data training*
https://www.ands.org.au/working-with-data/fairdata/training

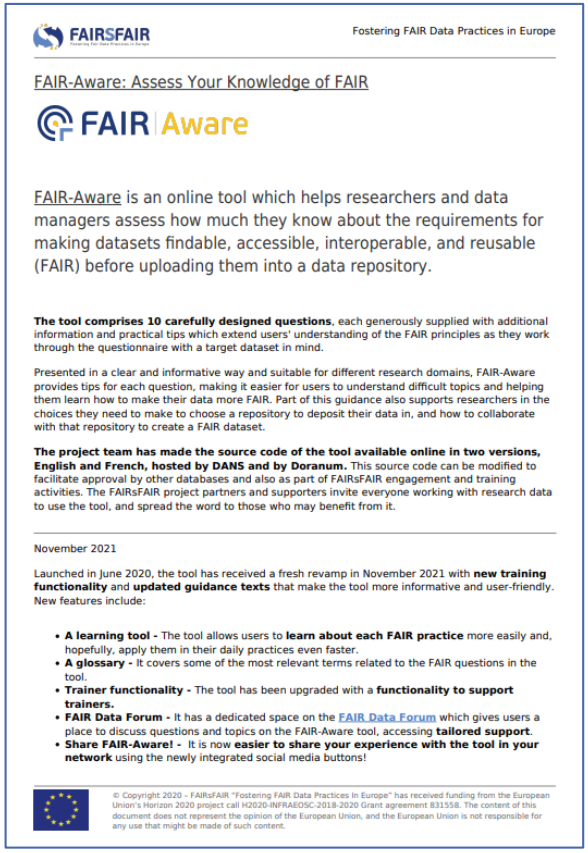
Australian Research Data Commons' FAIR data self assessment tool

https://ardc.edu.au/resources/working-with-data/fair-data/fair-self-assessment-tool/





FAIR-Aware



FAIRSFAIR — Fostering FAIR Data Practices in Europe

FAIR-Aware: Assess Your Knowledge of FAIR

FAIR Aware

FAIR-Aware is an online tool which helps researchers and data managers assess how much they know about the requirements for making datasets findable, accessible, interoperable, and reusable (FAIR) before uploading them into a data repository.

The tool comprises 10 carefully designed questions, each generously supplied with additional information and practical tips which extend users' understanding of the FAIR principles as they work through the questionnaire with a target dataset in mind.

Presented in a clear and informative way and suitable for different research domains, FAIR-Aware provides tips for each question, making it easier for users to understand difficult topics and helping them learn how to make their data more FAIR. Part of this guidance also supports researchers in the choices they need to make to choose a repository to deposit their data in, and how to collaborate with that repository to create a FAIR dataset.

The project team has made the source code of the tool available online in two versions, English and French, hosted by DANS and by Doranum. This source code can be modified to facilitate approval by other databases and also as part of FAIRsFAIR engagement and training activities. The FAIRsFAIR project partners and supporters invite everyone working with research data to use the tool, and spread the word to those who may benefit from it.

November 2021

Launched in June 2020, the tool has received a fresh revamp in November 2021 with **new training functionality** and **updated guidance texts** that make the tool more informative and user-friendly. New features include:

- **A learning tool -** The tool allows users to **learn about each FAIR practice** more easily and, hopefully, apply them in their daily practices even faster.
- **A glossary -** It covers some of the most relevant terms related to the FAIR questions in the tool.
- **Trainer functionality -** The tool has been upgraded with a **functionality to support trainers.**
- **FAIR Data Forum -** It has a dedicated space on the FAIR Data Forum which gives users a place to discuss questions and topics on the FAIR-Aware tool, accessing **tailored support**.
- **Share FAIR-Aware! -** It is now **easier to share your experience with the tool in your network** using the newly integrated social media buttons!

© Copyright 2020 - FAIRsFAIR "Fostering FAIR Data Practices In Europe" has received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 Grant agreement 831558. The content of this document does not represent the opinion of the European Union, and the European Union is not responsible for any use that might be made of such content.

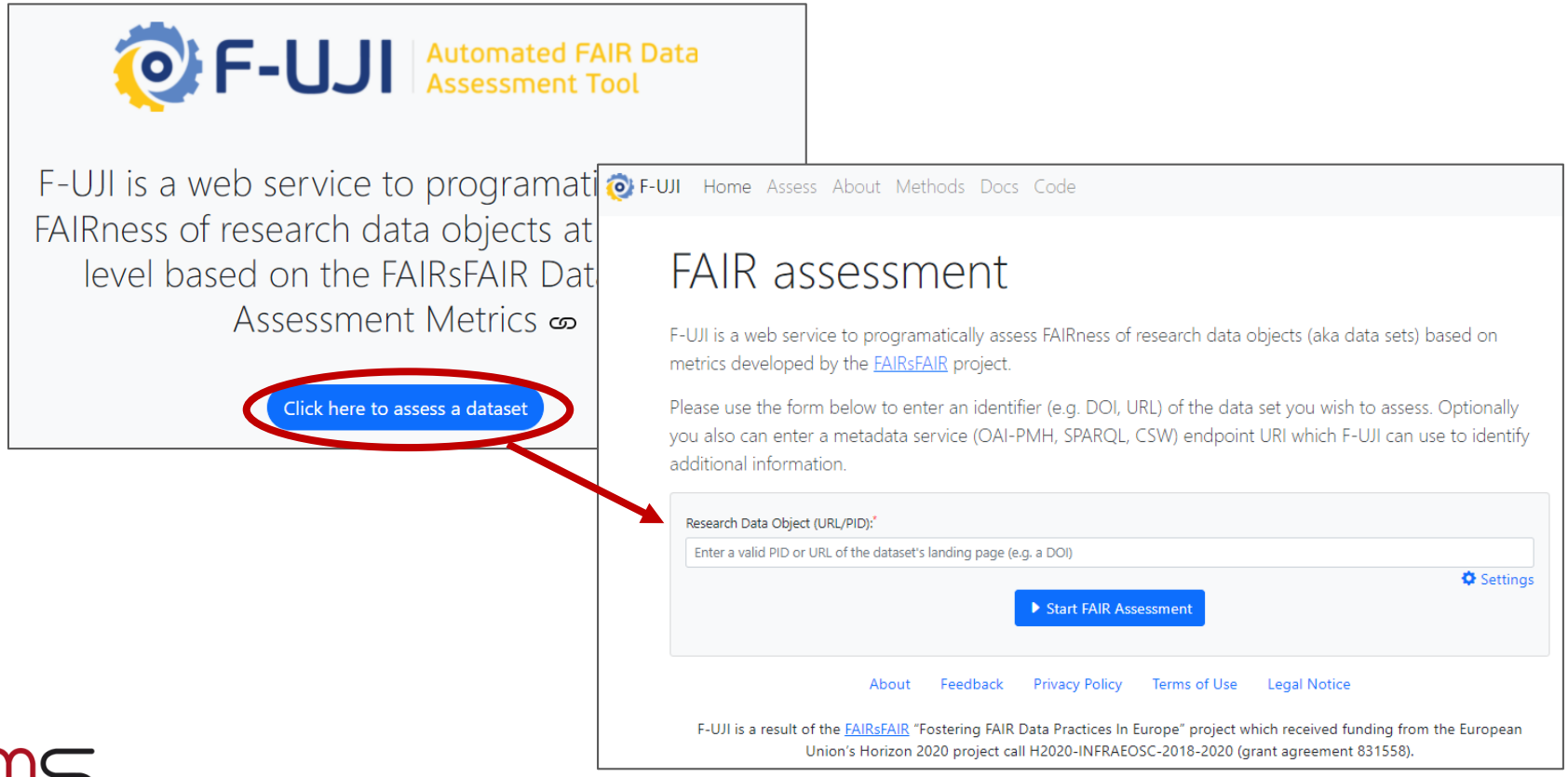
Herramienta en línea para ayudar a investigadores y gestores de datos a evaluar sus conocimientos sobre los requisitos FAIR de sus conjuntos de datos antes de cargarlos en un repositorio.

La herramienta consta de 10 preguntas, cada una de ellas generosamente provista de información adicional y consejos prácticos para entender la comprensión de los principios FAIR

https://www.fairsfair.eu/fair-aware

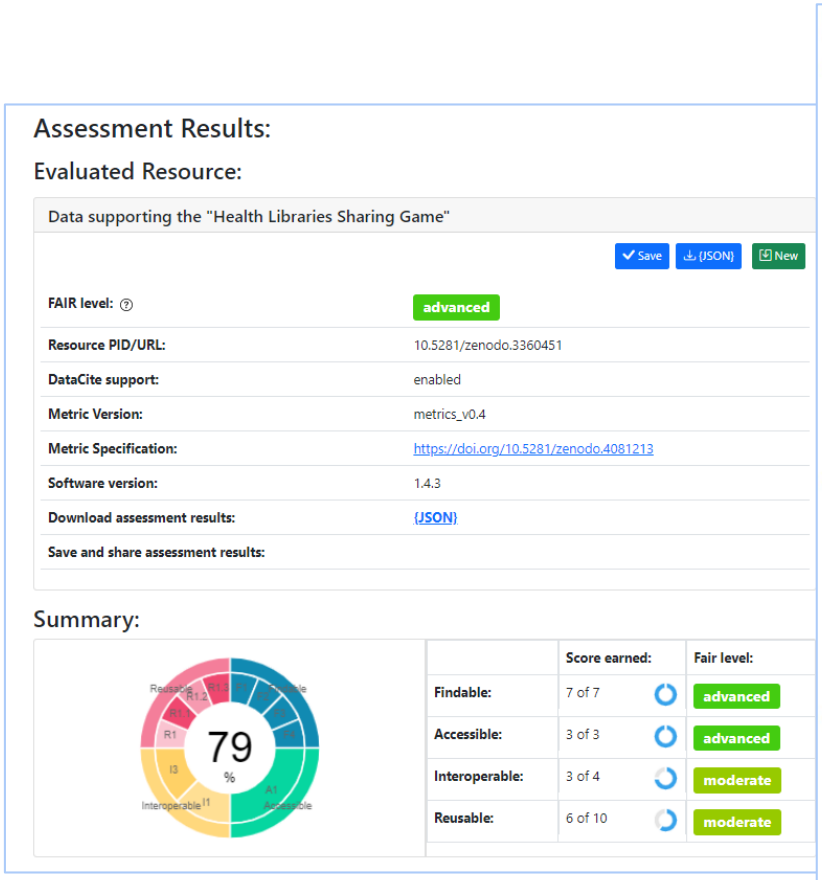
F-UJI tool to assess FAIRness of datasets

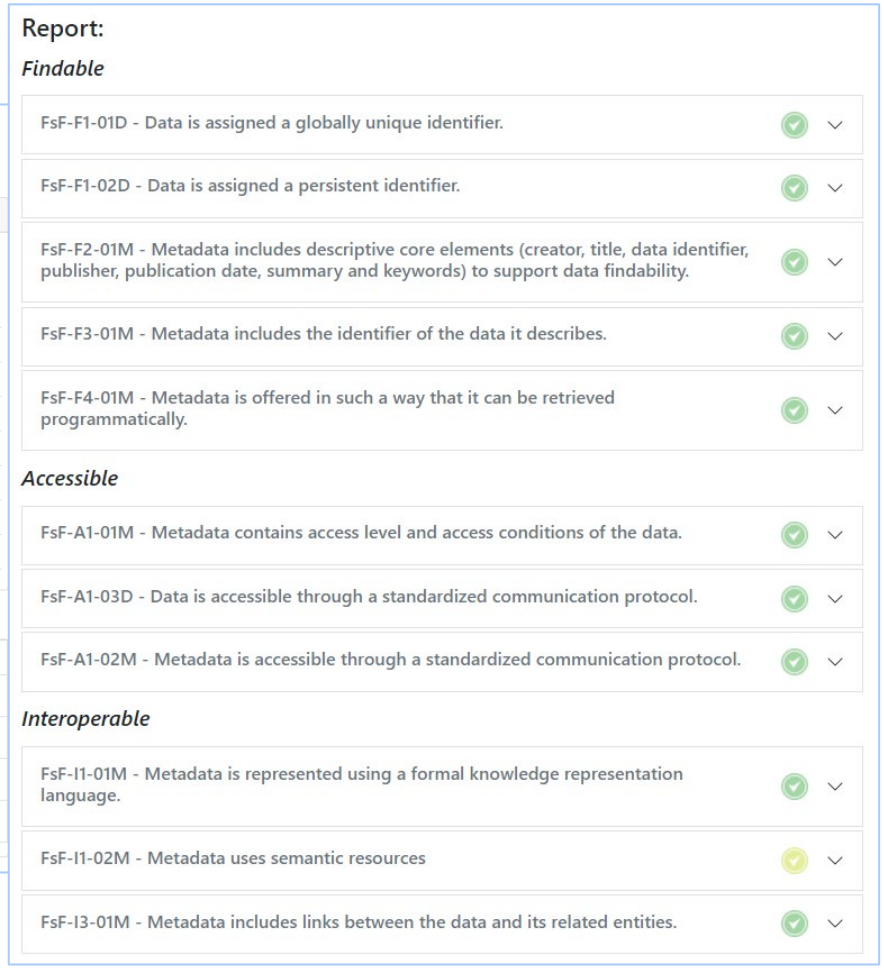
Un servicio web para evaluar el grado FAIR de los objetos de datos de investigación a nivel de conjunto de datos.



BIBLIO MAD SALUD

Data supporting the "Health Libraries Sharing Game". 10.5281/zenodo.3360451





EL PLAN DE GESTIÓN DE DATOS

BIBLIO MAD SALUD

QUÉ ES UN PLAN DE GESTIÓN DE DATOS

- Es una herramienta pensada para describir todos los aspectos relativos a la gestión de los datos a lo largo del ciclo de investigación, es decir, cómo se van a recoger, organizar, describir, almacenar, publicar, difundir y preservar los datos.
- El PGD se ha de desarrollar al inicio de la investigación, en el momento de la planificación; se trata de un instrumento vivo y flexible, que debe evolucionar e irse adaptando según avanza el proyecto y ganar en precisión.
- El PGD ha de recoger información sobre cómo los datos van a ser cumplir con los FAIR (Findable, Accesible, Interoperable and Reusable), es decir, localizables, accesibles, interoperables y reutilizables.
- El PGD es un elemento requerido cada vez por más organismos financiadores.

Al principio del proyecto:

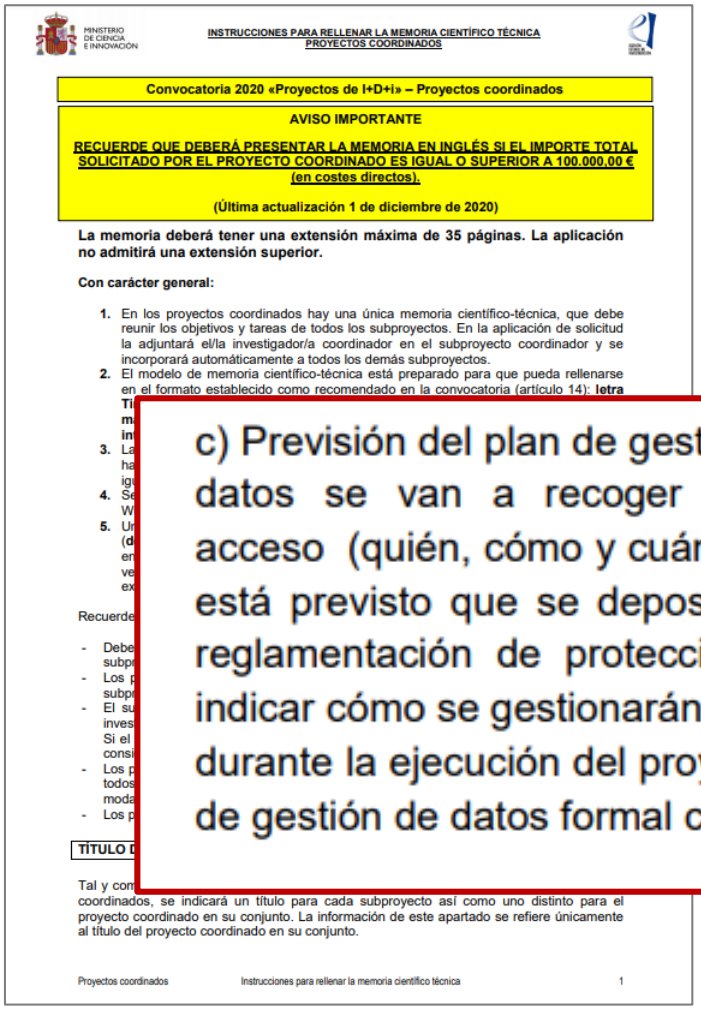
- Planificación, en el que hay que analizar las necesidades en materia de datos durante el proyecto, es cuando se redacta la primera versión del plan de gestión de datos (DMP) y se establece qué tipos de datos vamos a crear o recopilar, cómo vamos a organizarlos, denominarlos, etc.

Durante el proyecto:

- Ejecución de todas las tareas de gestión de datos propiamente dichas. En este punto se enmarcan las tareas relacionadas con la recopilación, descripción, almacenamiento y –en parte– publicación de los datos.
  Se redacta al menos una versión intermedia del DMP.

Al finalizar el proyecto:

- Es la fase de la preservación y la difusión de los datos, así como el análisis del alcance o el uso que se hace de los mismos. Se redacta la versión final del DMP.



INSTRUCCIONES PARA RELLENAR LA MEMORIA CIENTÍFICO TÉCNICA
PROYECTOS COORDINADOS

Convocatoria 2020 «Proyectos de I+D+i» – Proyectos coordinados

AVISO IMPORTANTE

RECUERDE QUE DEBERÁ PRESENTAR LA MEMORIA EN INGLÉS SI EL IMPORTE TOTAL SOLICITADO POR EL PROYECTO COORDINADO ES IGUAL O SUPERIOR A 100.000,00 € (en costes directos).

(Última actualización 1 de diciembre de 2020)

La memoria deberá tener una extensión máxima de 35 páginas. La aplicación no admitirá una extensión superior.

Con carácter general:

1. En los proyectos coordinados hay una única memoria científico-técnica, que debe reunir los objetivos y tareas de todos los subproyectos. En la aplicación de solicitud la adjuntará el/la investigador/a coordinador en el subproyecto coordinador y se incorporará automáticamente a todos los demás subproyectos.
2. El modelo de memoria científico-técnica está preparado para que pueda rellenarse en el formato establecido como recomendado en la convocatoria (artículo 14): **letra**
3. La
4. Se
5. Un

Recuerde

- Debe
- Los
- El su
- Los
- Los

TÍTULO

Tal y con
coordinados, se indicará un título para cada subproyecto así como uno distinto para el proyecto coordinado en su conjunto. La información de este apartado se refiere únicamente al título del proyecto coordinado en su conjunto.

Proyectos coordinados Instrucciones para rellenar la memoria científico técnica 1

Se recomienda que se publique en acceso abierto junto a los conjuntos de datos

c) Previsión del plan de gestión de datos de investigación en el que se indique qué datos se van a recoger o generar (tipologías y formatos), cómo será el acceso (quién, cómo y cuándo se podrá acceder a los datos) y en qué repositorio está previsto que se depositen. En el caso de datos que estén sometidos a la reglamentación de protección de los datos personales o de aspectos éticos, indicar cómo se gestionarán. En el caso de los proyectos que resulten financiados, durante la ejecución del proyecto y junto al informe final se podrá solicitar un plan de gestión de datos formal completo

Instrucciones para rellenar la memoria científico-técnica de proyectos coordinados. 1 de diciembre de 2020.

Información mínima requerida en un PGD

1. Descripción de datos y recopilación o reutilización de datos existentes
   - ¿Cómo se recopilarán o se generarán nuevos datos y cómo se reutilizarán datos existentes?
   - *¿Qué datos (tipos, formatos y volúmenes) se recopilarán o producirán?*
2. Documentación y calidad de los datos
   - *¿Qué metadatos y documentación (por ejemplo, metodología de recopilación de datos, forma de organizar los datos, etc.) acompañarán a los datos?*
   - *¿Qué medidas de control de calidad de los datos se utilizarán?*
3. Almacenamiento y respaldo durante el proceso de investigación
   - *¿Cómo se almacenarán y realizarán copias de seguridad de los datos y metadatos durante el proceso de investigación?*
   - *¿Cómo se cuidará la seguridad de los datos y la protección de los datos sensibles durante la investigación?*

4. Requisitos legales y éticos, códigos de conducta
   - *Si se procesan datos personales, ¿cómo se garantizará el cumplimiento de la legislación sobre datos personales y sobre seguridad de los datos?*
   - *¿Cómo se gestionarán otras cuestiones legales, como los derechos de propiedad intelectual? ¿Qué otra legislación es aplicable?*
   - *¿Cómo se tendrán en cuenta las posibles cuestiones éticas y se seguirán los códigos de conducta?*
5. Intercambio de datos y conservación a largo plazo
   - *¿Cómo y cuándo se compartirán los datos? ¿Existen posibles restricciones para compartir datos o motivos de embargo?*
   - *¿Cómo se seleccionarán los datos para su conservación y dónde se conservarán a largo plazo (en un repositorio o en un archivo de datos)?*
   - *¿Qué métodos o herramientas de software se necesitarán para acceder y utilizar los datos?*
   - *¿Cómo se garantizará la aplicación de un identificador único y persistente (por ejemplo, un DOI) a cada conjunto de datos?*
   - *¿Durante cuánto tiempo se guardarán los datos?*

6. Responsabilidades y recursos de gestión de datos
   - *¿Quién (función, cargo e institución) será responsable o administrador de la gestión de datos de datos?*
   - *¿Qué recursos (financieros y de tiempo) se dedicarán a la gestión de datos y a garantizar que los datos sean FAIR (localizables, accesibles, interoperables, reutilizables)?*

El DMP debe contener solo la información más importante y preferiblemente no debería exceder los 10,000 caracteres (incluidos los espacios). En caso de necesitar incluir más información, se puede proporcionar a través de enlaces a fuentes externas.

Jones S. (2011). 'How to Develop a Data Management and Sharing Plan'. *DCC How-to Guides*. Edinburgh: Digital Curation Centre. Disponible en: https://www.dcc.ac.uk/sites/default/files/documents/publications/reports/guides/How%20to%20Develop.pdf

CORE REQUIREMENTS
FOR DATA MANAGEMENT PLANS

When developing solid data management plans, researchers are required to deal with the following topics and answer the following questions:

1. **Data description and collection or re-use of existing data**
   a. How will new data be collected or produced and/or how will existing data be re-used?
   b. What data (for example the kinds, formats, and volumes) will be collected or produced?

2. **Documentation and data quality**
   a. What metadata and documentation (for example the methodology of data collection and way of organising data) will accompany data?
   b. What data quality control measures will be used?

3. **Storage and backup during the research process**
   a. How will data and metadata be stored and backed up during the research process?
   b. How will data security and protection of sensitive data be taken care of during the research?

4. **Legal and ethical requirements, codes of conduct**
   a. If personal data are processed, how will compliance with legislation on personal data and on data security be ensured?
   b. How will other legal issues, such as intellectual property rights and ownership, be managed? What legislation is applicable?
   c. How will possible ethical issues be taken into account, and codes of conduct followed?

5. **Data sharing and long-term preservation**
   a. How and when will data be shared? Are there possible restrictions to data sharing or embargo reasons?
   b. How will data for preservation be selected, and where will data be preserved long-term (for example a data repository or archive)?
   c. What methods or software tools will be needed to access and use the data?
   d. How will the application of a unique and persistent identifier (such as a Digital Object Identifier (DOI)) to each data set be ensured?

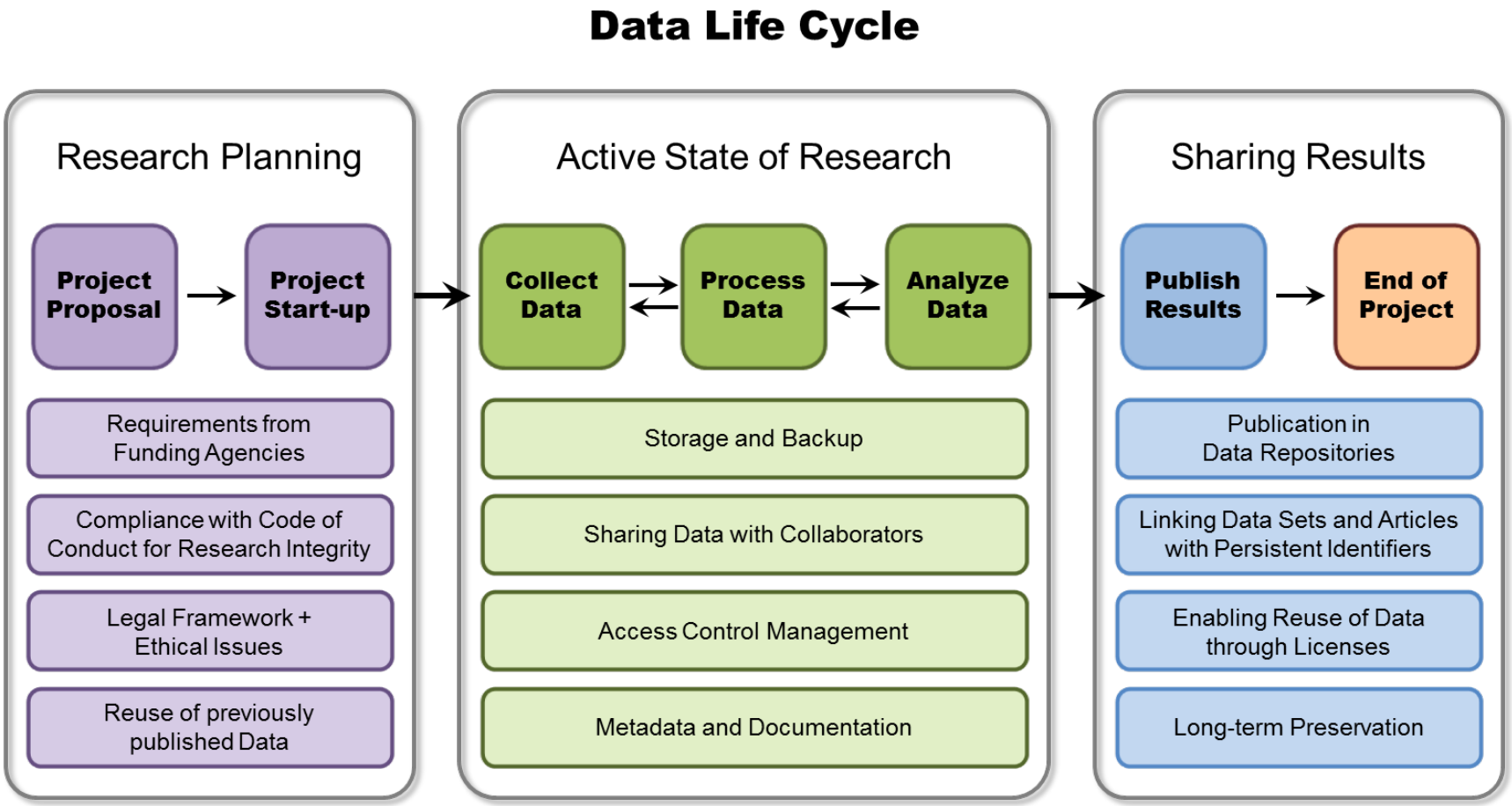
6. **Data management responsibilities and resources**
   a. Who (for example role, position, and institution) will be responsible for data management (i.e. the data steward)?
   b. What resources (for example financial and time) will be dedicated to data management and ensuring that data will be FAIR (Findable, Accessible, Interoperable, Re-usable)?

https://www.scienceeurope.org/media/jezkhnoo/se_rdm_practical_guide_final.pdf

10 pasos para elaborar un Plan de Gestión de Datos
Un Plan de Gestión de Datos (PGD) o Data Management Plan (DMP) es un documento formal, que debe presentarse al inicio de la investigación, en el que se describe qué vas a hacer con tus datos durante y después de finalizar tu investigación y que puede modificarse si se producen cambios en el proceso de la misma.
¿Por qué?
Es una buena práctica, es un elemento clave de Open Science y es obligatorio en los nuevos proyectos H2020.
Herramientas gratuitas para elaborar un PGD
PGDonline (Consorcio Madroño) http://dmp.consorciomadrono.es/
DMPonline (Digital Curation Centre, UK) https://dmponline.dcc.ac.uk/
01
Revisa los requerimientos de la entidad financiadora (H2020).
02
Identifica los datos: tipología, procedencia, volumen, formatos y ficheros.
03
Define cómo se organizarán y gestionarán los datos: nombre de los ficheros, control de versiones, software necesario...
04
Explica cómo se documentarán los datos: identifica la información a procesar, consulta si hay estándares o esquemas de metadatos, identifica herramientas que permitan gestionarlos.
05
Describe los procesos que aseguran una buena calidad de los datos.
06
Prepara una estrategia de almacenamiento (durante el proceso) y de preservación de datos (repositorio).
07
Define las políticas de datos del proyecto: cuestiones sobre propiedad intelectual y cómo se tratarán los datos sensibles y personales.
08
Describe cómo se difundirán los datos: dónde, cuáles, cuándo se van a difundir. Si publicarás los datos en un repositorio, como información suplementaria del artículo o como un "data paper".
09
Asigna roles y responsabilidades para las personas y organizaciones participantes en el proyecto.
10
Prepara un presupuesto realista: la gestión de datos cuesta tiempo y dinero en términos de software, hardware, servicios y personal.
crue Universidades Españolas
Red de Bibliotecas REBIUN
Adaptado de Michener, WK. Ten simple rules for creating a good data managent plan. PLOS Comp Biol. 2015; 11(10)
El personal de tu Biblioteca te puede asesorar
CC BY

ASPECTOS PRÁCTICOS Y RECURSOS PARA LA GESTIÓN DE DATOS DE INVESTIGACIÓN

Actividades en el ciclo de vida de los datos



Hüser, Falco Jonas; Elbæk, Mikael K.; Martinez lavanchy, Paula (2016): DTU Research Data Life Cycle. Technical University of Denmark. Figure. https://doi.org/10.6084/m9.figshare.4258019.v1

Descripción y organización de los datos, formatos y versiones

Preguntas a responder

- ¿En qué tipo de datos se basa la investigación?
- ¿Los datos van a ser derivados, se van a crear de cero, recopilar o reutilizar?
- ¿Qué formatos se manejarán?
- ¿Cuál es el tamaño total esperado de los datos recopilados?
- ¿Cómo se va a establecer la estructura de las carpetas?
- ¿Cómo se va a definir la nomenclatura de archivos?
- ¿Se implementarán estándares específicos, como convenciones de nomenclatura o estructuras de codificación estandarizadas?

Tomar decisiones acordadas por consenso

Organización y estructura de carpetas

- Comprobar si hay algún procedimiento interno en el grupo de trabajo, departamento o institución.
- Utilizar las mismas estructuras de carpetas de marea sistemática en diferentes proyectos.
- Denominar las carpetas de forma que su contenido sea claramente reconocible.
- Preferible denominar las carpetas según las áreas de trabajo relacionadas, sin asignar nombres de investigadores.
- Una vez decidido un método de denominar carpetas, todos los miembros participantes en la investigación lo han de seguir.
- Al estructurar carpetas jerárquicamente es conveniente empezar con un número limitado de carpetas para los temas principales y posteriormente crear carpetas más específicas dentro de estas.
- Es recomendable establecer un máximo de tres niveles de subcarpetas.
- Es conveniente separar los trabajos en curso de los ya finalizados.
- No es necesario guardar todos los archivos intermedios o que no aportan valor; se deben evaluar los materiales durante o al final del proyecto y eliminar los sobrantes.

Denominación de archivos (i)

La nomenclatura ha de ser significativa, coherente y descriptiva, de forma que facilite la gestión y localización de archivos.

- Numerar cada conjunto de datos de forma continua.
- Utilizar nombres únicos, no asignar nombres iguales en carpetas diferentes
- Elegir nombres cortos y representativos, utilizando únicamente letras, números, guiones bajos y guiones. Decidir qué símbolos de puntuación, mayúsculas, guiones, etc. y cuándo se van a usar.
- Evitar el uso de espacios, así como barras diagonales y caracteres especiales como: ~! @ # $% ^ & * () `; : <>? . , [] {} ,“ |
- En caso de utilizar abreviaturas, explicarlas en un índice, siguiendo unas reglas de nomenclatura de datos.
- Evitar nombres de archivos demasiado largos.
- Es conveniente que el nombre del archivo comience con el elemento más común y luego volverse más específico. Se recomienda separar los elementos individuales con un guion bajo "_" para los elementos no relacionados y guion "-" para separar los elementos conectados, si fuera necesario.

Denominación de archivos (ii)

- En el caso de varios autores o editores de los archivos, especificar de quién es la contribución utilizando iniciales o abreviaturas de apellidos y nombres.
- Designar las fechas como AAAMMDD o en su defecto como AAAA-MM-DD (ISO 8601 aplicada a la gestión de datos de investigación).
- Utilizar un sistema de numeración secuencial, esto es, con ceros a la izquierda, para asegurar el orden de los archivos. "001, 002, .. 010" en lugar de "1, 2, .. 10"
- Utilizar el control de versiones para indicar la versión más actual, p. Ej. filename_v02.xxx

Ejemplo: 01_Surveydaten2019_V3_20201121_GOM.

- ✓ *01 – Paquete de trabajo*
- ✓ *Surveydaten2019 – Datos del experimento, actividad, etc.*
- ✓ *V3 – Versión*
- ✓ *20201121 – Fecha*
- ✓ *GOM – Autor(a)*

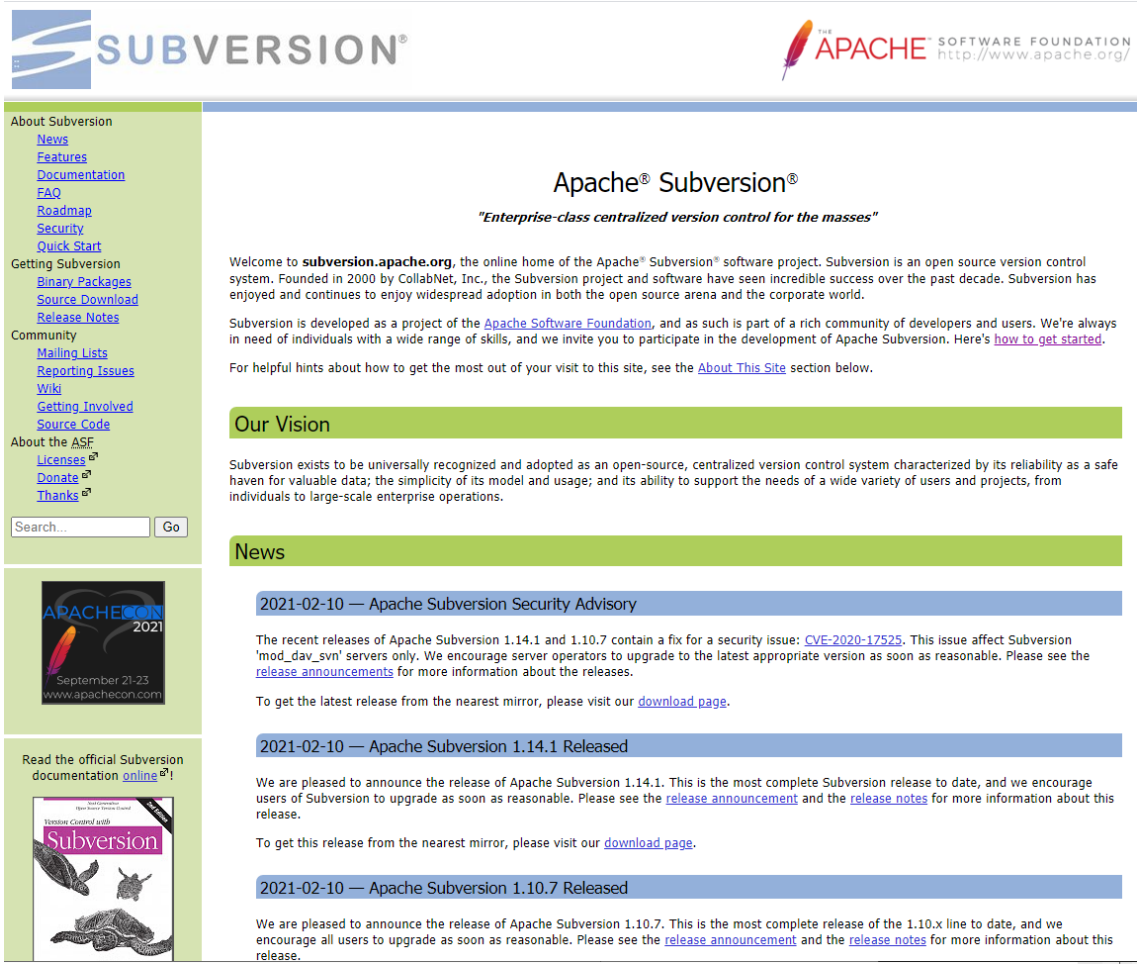
Organización de versiones

- La identificación de versiones es fundamental, en especial cuando se trata de conjuntos de datos dinámico, es decir, que requieren cambios y revisiones. Es especialmente importante para su posterior reutilización y citación.
- Se puede optar por un control automático de versiones o, en su defecto, se deberá asignar manualmente un número de versión o a través del uso de fechas.
- Es aconsejable integrar el control de versiones de archivos en las reglas que se adopten en la nomenclatura de archivos (es decir, v1, v1_2, v2; o en caso de optar por fechas, en formatos AAAAMMDD o AAAA-MM-DD).
- Se recomienda evitar el uso de términos ambiguos como "final" o "revisión". En caso de que se opte por marcar la versión definitiva como "final", es imprescindible que se eviten situaciones como “final_final_final" o “final1”, “final2”, etc.

Herramientas de software para gestionar versiones

Apache® Subversion®

Un sistema de control de versiones de código abierto fundado en 2000 por CollabNet, Inc.



https://subversion.apache.org/

Formatos de archivos

Es importante contar con formatos estándar y a poder ser abiertos, que aseguren el acceso a largo plazo a los datos.

Según el Open Data Handbook:

- Un formato abierto es aquel donde las especificaciones del software están disponibles para cualquier persona, de forma gratuita, así cualquiera puede usar dichas especificaciones en su propio software sin ninguna limitación en su reutilización que fuere impuesta por derechos de propiedad intelectual.
- Si el formato del archivo es 'cerrado', esto puede ser debido a que el formato es propietario y sus especificaciones no están disponibles públicamente, o porque el formato es propietario y aunque las especificaciones se han hecho públicas, su reutilización es limitada.

Open Knowledge Foundation. The open data handbook.
http://opendatahandbook.org/guide/en/

Consideraciones al momento de elegir formatos de archivo

- **Elegir formatos comunes al campo disciplinar al que se está trabajando**: Para asegurar la interoperabilidad y la reutilización de los datos.
- **Tener en cuenta el tiempo en que se espera conservar los datos:** Cuanto mayor sea el periodo de tiempo que se desea conservar los datos, mayor será la necesidad de seleccionar formatos abiertos, estandarizados y bien documentados.
- **La conversión de archivos puede provocar la pérdida de datos**: Se han de considerar formatos de multiplataforma común que respondan a estándares específicos para evitar dicha pérdida.
- **Verificar los requisitos del repositorio de datos:** Muchas revistas, archivos y repositorios requieren de formatos específicos al momento de cargar la información.

Biblioteca CEPAL. Biblioguía sobre la Gestión de datos de investigación
https://biblioguias.cepal.org/gestion-de-datos-de-investigacion/formatos

Type of data	Recommended formats	Acceptable formats
Tabular data with extensive metadata variable labels, code labels, and defined missing values	SPSS portable format (.por) delimited text and command ('setup') file (SPSS, Stata, SAS, etc.) structured text or mark-up file of metadata information, e.g. DDI XML file	proprietary formats of statistical packages: SPSS (.sav), Stata (.dta), MS Access (.mdb/.accdb)
Tabular data with minimal metadata column headings, variable names	comma-separated values (.csv) tab-delimited file (.tab) delimited text with SQL data definition statements	delimited text (.txt) with characters not present in data used as delimiters widely-used formats: MS Excel (.xls/.xlsx), MS Access (.mdb/.accdb), dBase (.dbf), OpenDocument Spreadsheet (.ods)
Geospatial data vector and raster data	ESRI Shapefile (.shp, .shx, .dbf, .prj, .sbx, .sbn optional) geo-referenced TIFF (.tif, .tfw) CAD data (.dwg) tabular GIS attribute data Geography Markup Language (.gml)	ESRI Geodatabase format (.mdb) MapInfo Interchange Format (.mif) for vector data Keyhole Mark-up Language (.kml) Adobe Illustrator (.ai), CAD data (.dxf or .svg) binary formats of GIS and CAD packages

UK Data Service. *Format your data. "Create well organised and sustainable data"* https://www.ukdataservice.ac.uk/manage-data/format/file-formats.aspx

Type of data	Recommended formats	Acceptable formats
Textual data	Rich Text Format (.rtf) plain text, ASCII (.txt) eXtensible Mark-up Language (.xml) text according to an appropriate Document Type Definition (DTD) or schema	Hypertext Mark-up Language (.html) widely-used formats: MS Word (.doc/.docx) some software-specific formats: NUD*IST, NVivo and ATLAS.ti
Image data	TIFF 6.0 uncompressed (.tif)	JPEG (.jpeg, .jpg, .jp2) if original created in this format GIF (.gif) TIFF other versions (.tif, .tiff) RAW image format (.raw) Photoshop files (.psd) BMP (.bmp) PNG (.png) Adobe Portable Document Format (PDF/A, PDF) (.pdf)
Audio data	Free Lossless Audio Codec (FLAC) (.flac)	MPEG-1 Audio Layer 3 (.mp3) if original created in this format Audio Interchange File Format (.aif) Waveform Audio Format (.wav)
Video data	MPEG-4 (.mp4) OGG video (.ogv, .ogg) motion JPEG 2000 (.mj2)	AVCHD video (.avchd)
Documentation and scripts	Rich Text Format (.rtf) PDF/UA, PDF/A or PDF (.pdf) XHTML or HTML (.xhtml, .htm) OpenDocument Text (.odt)	plain text (.txt) widely-used formats: MS Word (.doc/.docx), MS Excel (.xls/.xlsx) XML marked-up text (.xml) according to an appropriate DTD or schema, e.g. XHMTL 1.0

U.S. National Archives and Records Administration

https://www.archives.gov/records-mgmt/policy/transfer-guidance-tables.html

Formatos de archivos de datos recomendados.

Fuente: *U.S. National Archives and Records Administration*

https://www.archives.gov/records-mgmt/policy/transfer-guidance-tables.html

Digital Audio

The Digital audio category encompasses formats used to encode recorded sound as machine readable files by converting acoustic sound waves into digital signals. Digital audio formats are generally composed of both a wrapper format, usually the common name associated with the file extension, and an encoding method or codec.

General requirements for digital audio records:

- Digitize to standards appropriate for the accurate preservation of the original audio, when converting analog material (e.g., audio cassettes, record albums, and reel-to-reel audio tapes). Examples of appropriate methods and formats are available on NARA's Digitization Services Products and Services page;
- Transfer digital audio at a minimum of 16 bits per sample, but 24 bits per sample is encouraged; and
- Transfer digital audio at a minimum sample rate of at least 44.1 KHz, but sampling at 96 KHz is encouraged.

Preferred Formats

Preferred Formats	Format Version	Codecs	Format Specifications
Broadcast Wave (BWF)	0, 1 & 2	Linear Pulse Code Modulated Audio (LPCM)	European Broadcast Union (EBU). Tech Specification of the Broadcast Wave Format (BWF) – Version 1: (http://web.archive.org/web/20091229093941/http://tech.ebu.ch/docs/tech/tech3285.pdf) Specification of the Broadcast Wave Format (BWF) - Version 2: (https://tech.ebu.ch/docs/tech/tech3285.pdf)
Free Lossless Audio Codec (FLAC)	1.21	FLAC	FLAC Format Specification version 1.21: (http://flac.sourceforge.net/format.html)

Acceptable Formats

Acceptable Formats	Format Version	Codecs	Format Specifications
Audio Interchange Format (AIFF)	1.3	Linear Pulse Code Modulated Audio (LPCM)	Audio Interchange File Format: "AIFF" A Standard for Sampled Sound Files Version 1.3 Apple Computer, Inc.: (http://www-mmsp.ece.mcgill.ca/Documents/AudioFormats/AIFF/Docs/AIFF-1.3.pdf)
MPEG Audio Layer III (MP3)		MP3enc, Lame	ISO/IEC-11172-3 Information technology – Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s – Part 3: Audio: (http://www.iso.org/iso/iso_catalogue/catalogue_tc/catalogue_detail.htm?csnumber=22412) ISO/IEC 13818-3:1995 Information technology – Generic coding of moving pictures and associated audio information – Part 3: Audio: (http://www.iso.org/iso/home/store/catalogue_ics/catalogue_detail_ics.htm?csnumber=26797)
Wave Waveform Audio File Format (Wave)		Linear Pulse Code Modulated Audio (LPCM)	Multimedia Programming Interface and Data Specifications 1.0: (http://www-mmsp.ece.mcgill.ca/Documents/AudioFormats/WAVE/Docs/riffmci.pdf)

Esquemas de metadatos

Metadatos bibliográficos y administrativos

Proporcionan información sobre el origen de un conjunto de datos completo.

- Título: nombre del conjunto de datos o proyecto de investigación
- Autor / investigador principal: nombre, institución, identificador de la persona
- Colaboradores (contributor): personas / instituciones que no participan principalmente en la creación de datos (por ejemplo, curadores de datos, administradores de datos
- Identificador: número individual que identifica los datos
- Tipo de datos: tipo de datos, formato de archivo y tamaño de archivo
- Derechos: derechos de uso y licencias
- Fechas: períodos de tiempo asociados con los datos (por ejemplo, inicio del proyecto, conclusión del proyecto, período de observación, fecha de publicación)
- Idioma: idioma o idiomas del contenido de los datos de la investigación
- Lugar: referencias a una ubicación física o cobertura territorial (por ejemplo, coordenadas)
- Resumen de contenido y palabras clave o frases que describen el contenido.
- Investigación: organismo de financiación, número de subvención.
- Relaciones: información sobre las relaciones con otros recursos.

Formatos y estándares para metadatos

Los metadatos a menudo se guardan como XML (lenguaje de marcado extensible) u otro formato de lenguaje de marcado (por ejemplo, JSON).

Los textos de origen XML son legibles por máquina y por humanos y se pueden transferir a otros formatos (por ejemplo, con la ayuda de un conversor XML - JSON en JSON).

Para aumentar la efectividad de los metadatos, los repositorios y la comunidad científica utilizaron estándares definidos para metadatos. La estandarización permite vincular y editar juntos metadatos de diferentes fuentes. A menudo, un estándar también se puede transferir a otro mediante mapeo.

La Research Data Alliance (RDA) tiene una lista de estándares de metadatos disciplinarios para datos científicos.

Un estándar de metadatos simple y generalizado es Dublin Core. Consta de 15 elementos (por ejemplo, <dc: creator> Max Mustermann </ dc: creator) y varios elementos secundarios elementos.

Vocabularios controlados, clasificaciones , etc.

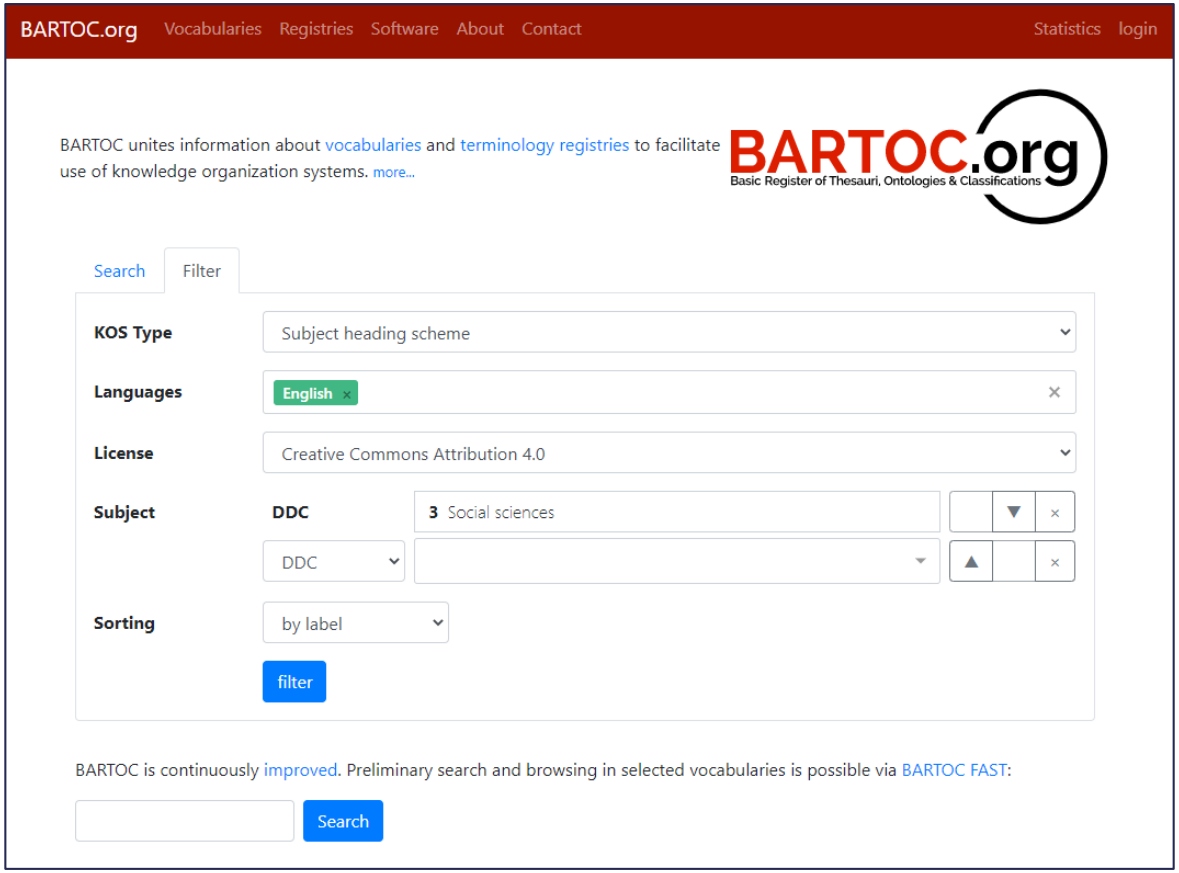
Para optimizar las búsquedas y facilitar el procesamiento automático, debe utilizar términos fijos para los valores individuales en sus metadatos. El uso de términos estandarizados e identificadores claros también ayuda a evitar ambigüedades y redundancias.

Ejemplos de vocabularios controlados (tesauros y clasificaciones) y archivos de autoridades.

- ISBN para libros
- ISO 8601 para representaciones de fecha y hora; ISO 639 para idiomas; ISO 3166-1 alpha-2 para abreviaturas de países
- GeoNames para nombres geográficos y objetos topográficos
- AGROVOC para terminología de agricultura y nutrición
- ICD para enfermedades

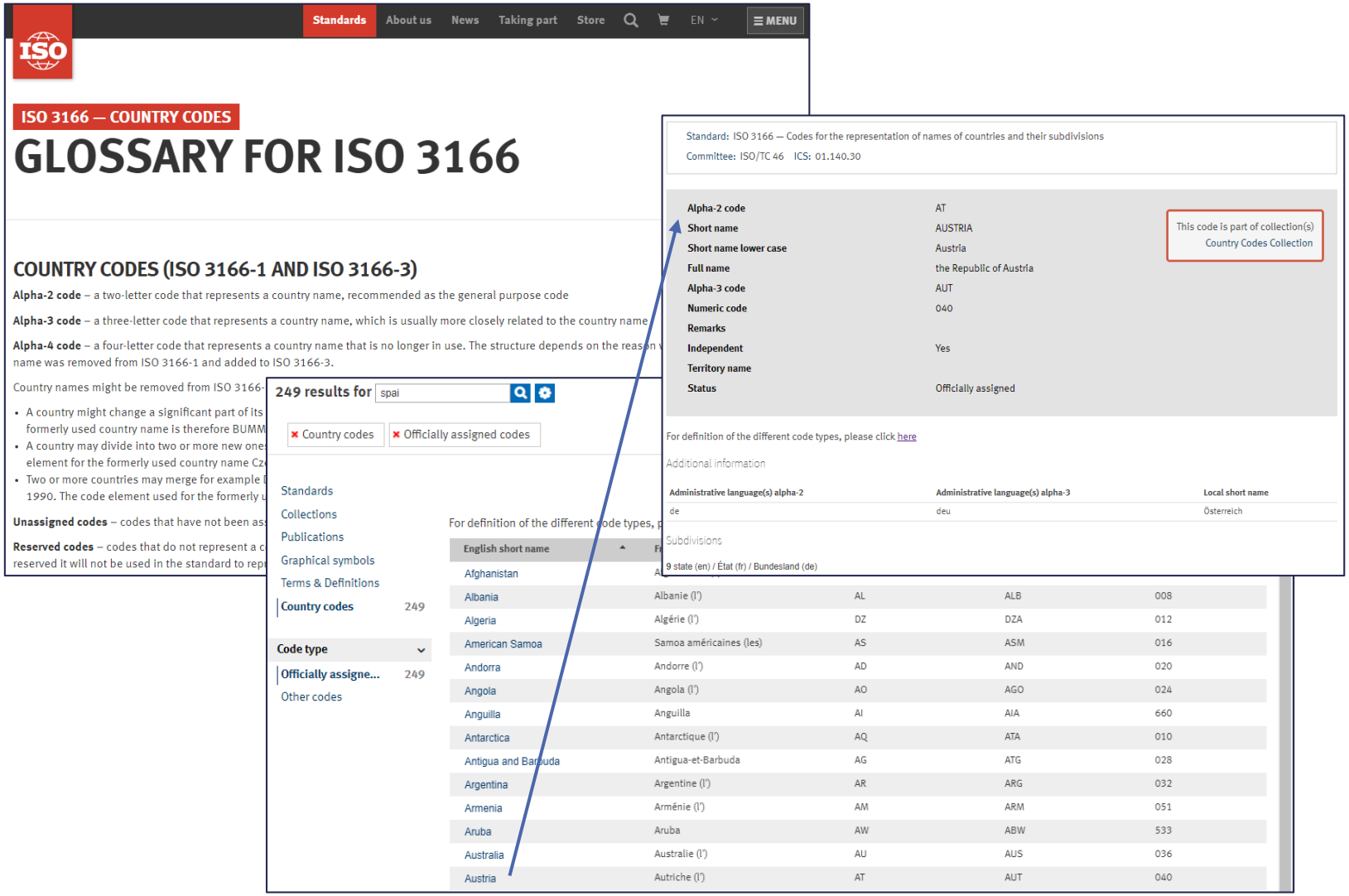
BARTOC.org – Basic Register of Thesauri, Ontologies & Classifications

BARTOC incluye cualquier tipo de KOS (Sistemas de Organización de Conocimiento, en inglés Knowledge Organization Systems) de cualquier área temática, en cualquier idioma, cualquier formato de publicación y cualquier forma de accesibilidad.
La interfaz está disponible en 20 idiomas europeos.



http://bartoc.org/

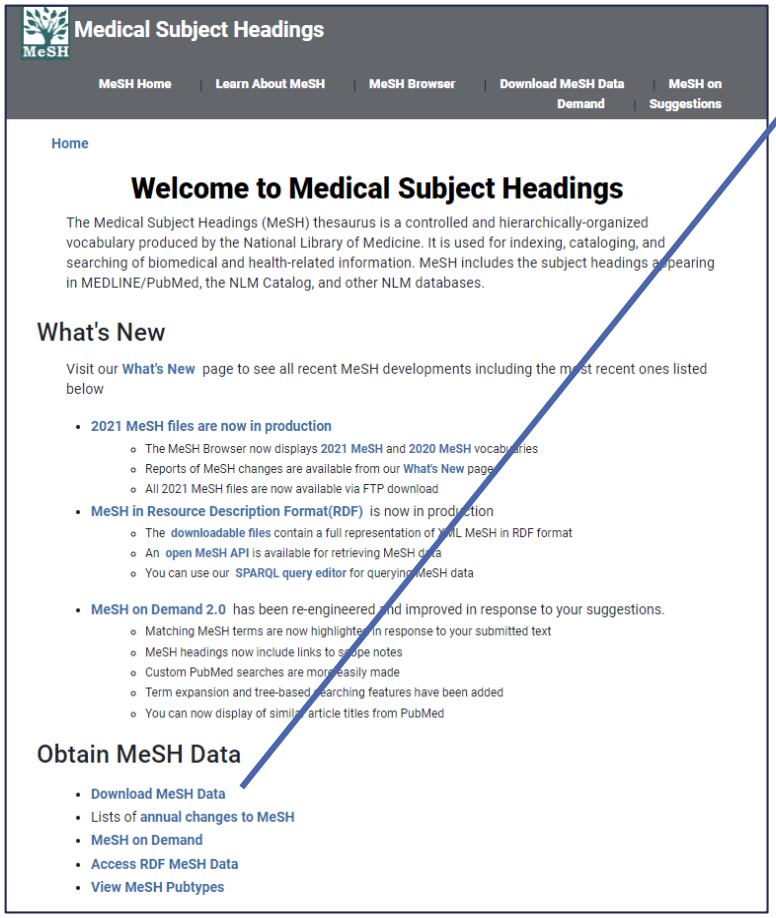
Normas ISO, por ej. para códigos de países

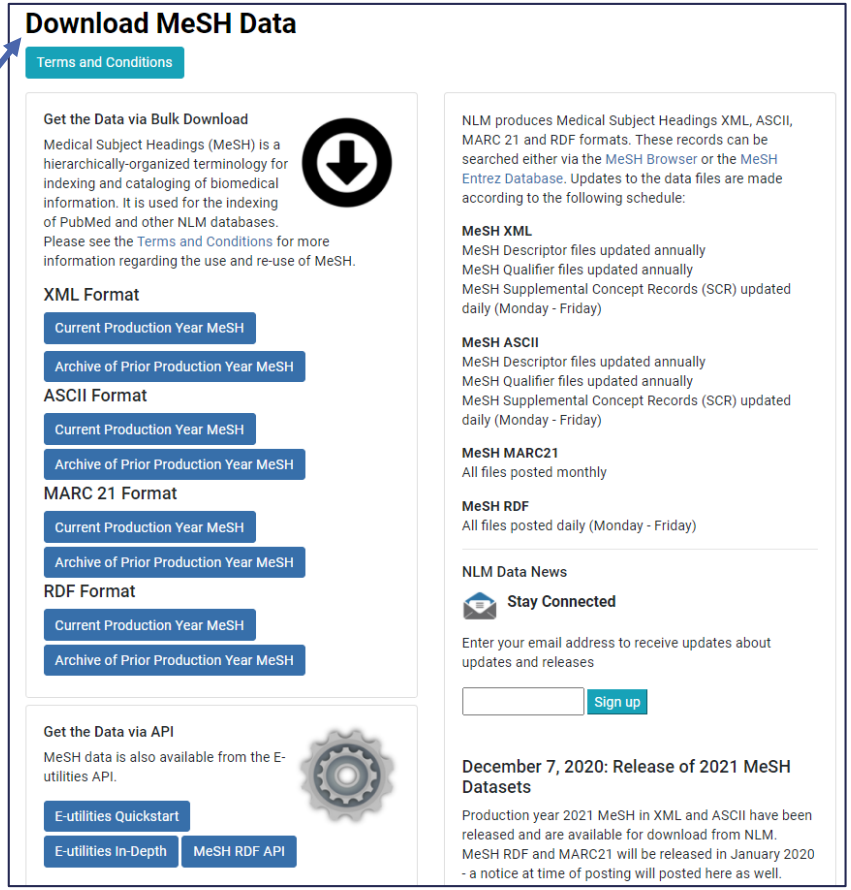


https://www.iso.org/obp/ui/#search

Vocabularios controlados

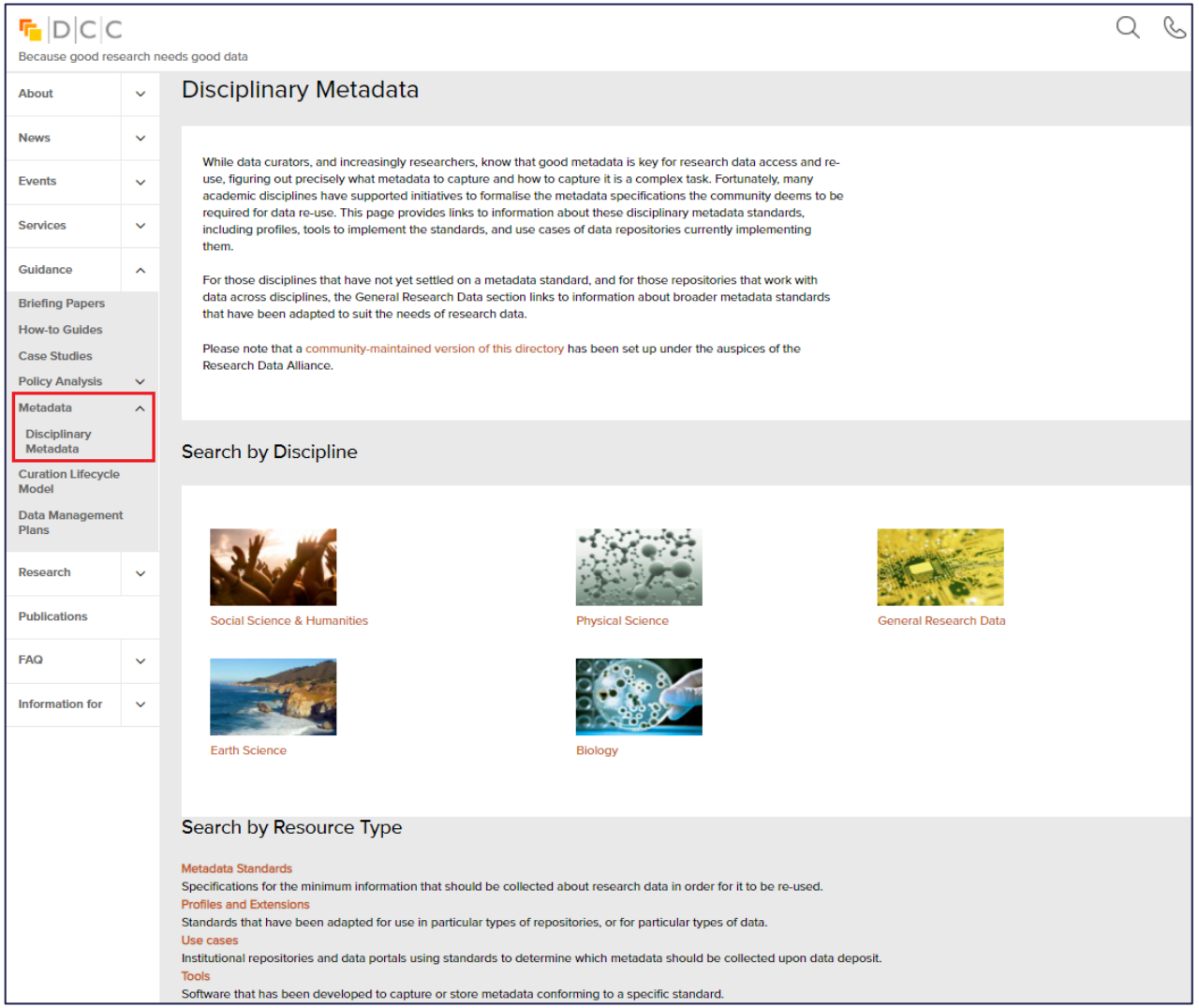
Medical Subject Headings (MeSH) de la National Library of Medicine





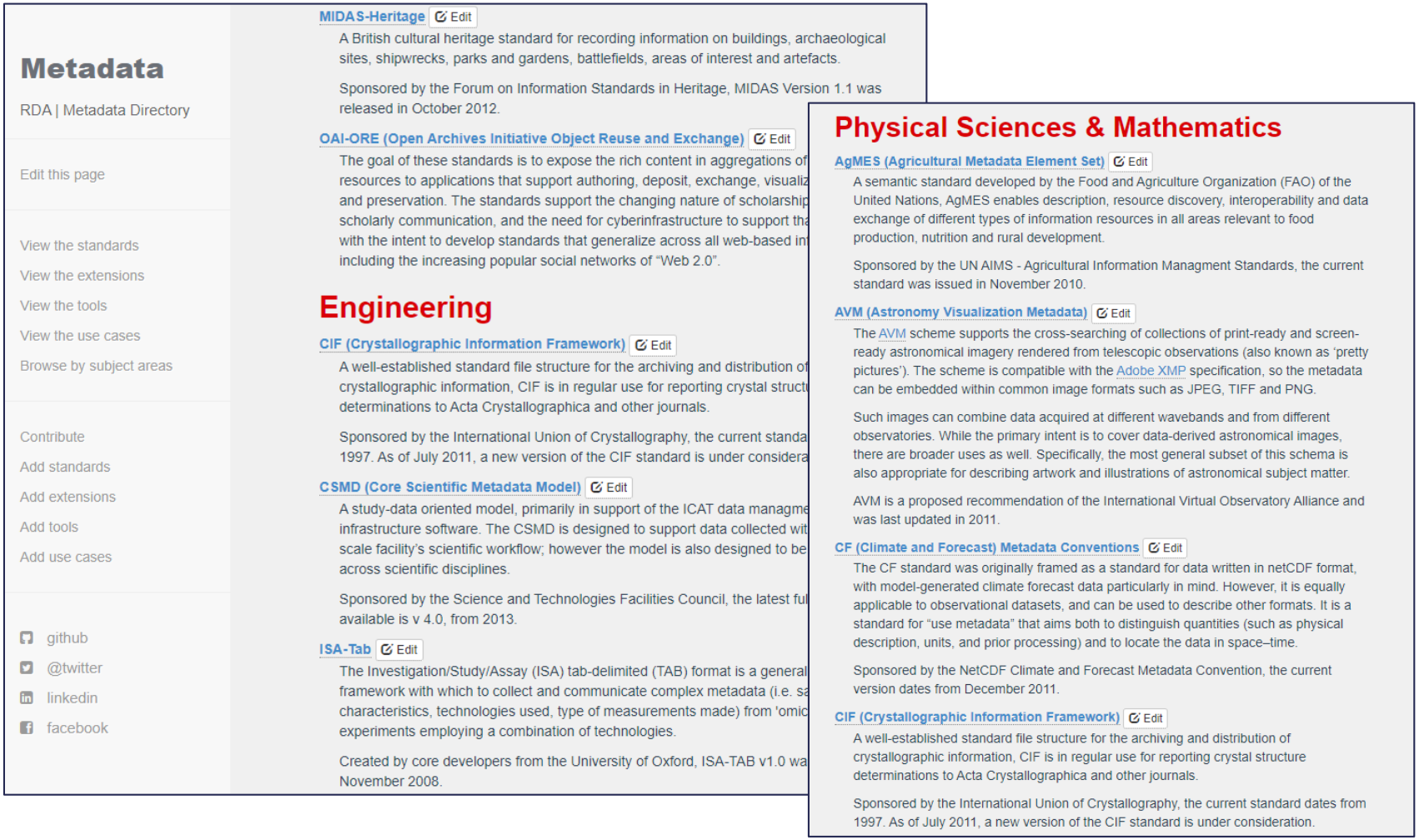
https://www.nlm.nih.gov/mesh/meshhome.html

Herramientas de metadatos del Digital Curation Center (DCC)



https://www.dcc.ac.uk/guidance/standards/metadata

Directorio de metadatos del RDA Metadata Standards Directory Working Group



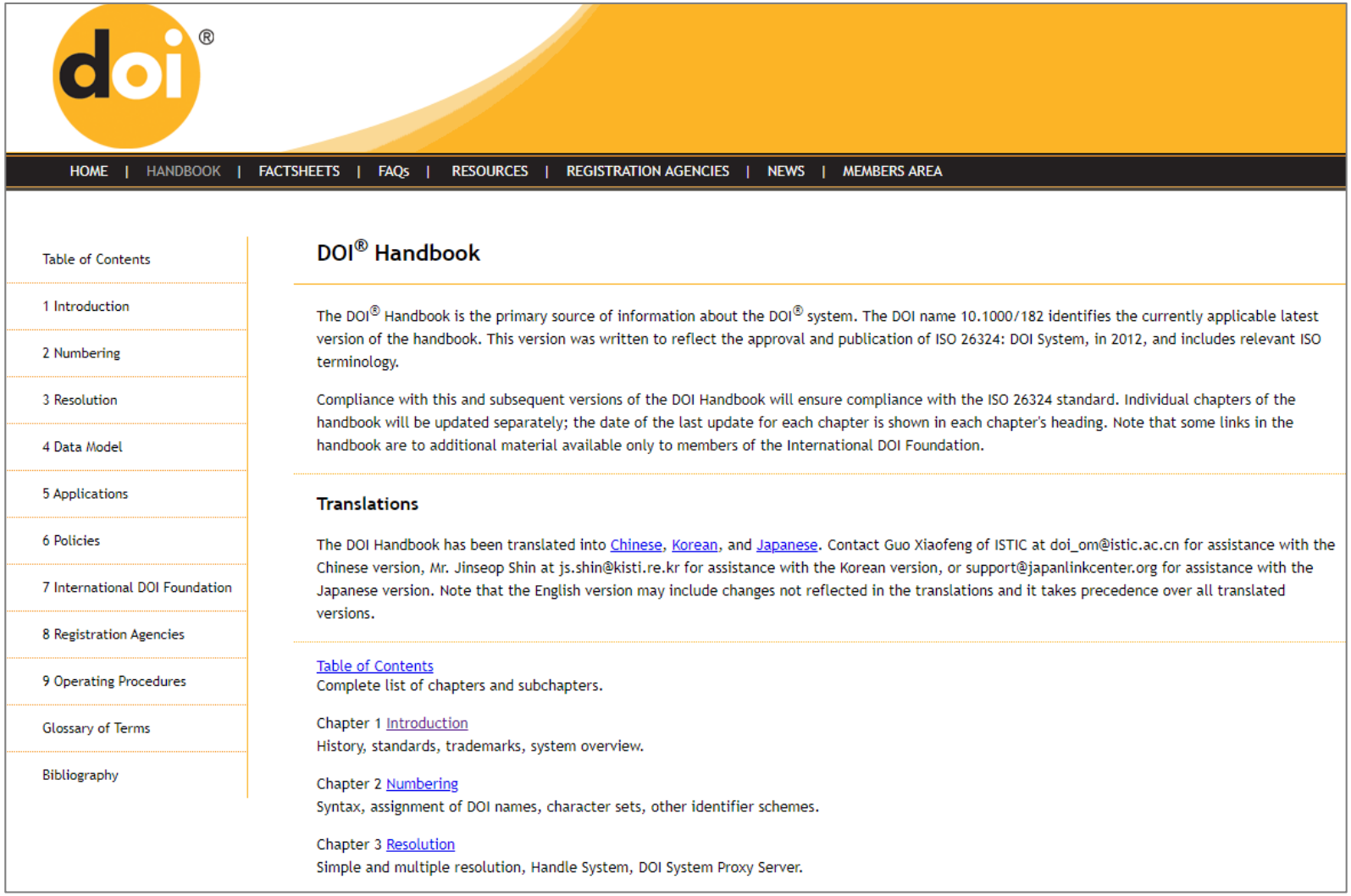
http://rd-alliance.github.io/metadata-directory/

Asignación de identificadores persistentes

- Handle.
  Un sistema de resolución de identificadores no comerciales, diseñado en 1994 por la Corporación para Iniciativas de Investigación Nacional norteamericana (Corporation for National Research Initiatives, CNRI). Se utiliza como base de otros sistemas de nivel superior como los DOIs. Se usa ampliamente en repositorios institucionales en el mundo académico y en las instituciones de investigación.

- DOI – Digital object identifier (Identificador de objeto digital)
  Combina un modelo de metadatos con el sistema Handle como infraestructura de resolución (es decir, los DOI se basan en Handles). Se introdujo en 1998 con fondos de la Fundación Internacional DOI (IDF) y en 2012 se convirtió en una norma ISO (ISO 26324).

- URN - Uniform Resource Name (Nombre uniforme de recurso)
  Un identificador persistente diseñado específicamente por la Internet Engineering Task Force (IETF) para ser interoperable con los sistemas de identificación estándar existentes, como el ISSN. Se introdujo en 1994, y se formalizó en 1997, convirtiéndose en una norma IETF. Es utilizado por las principales bibliotecas nacionales de Europa. Los ISBN de los libros o los ISSN de las series forman parte del sistema URN.



https://www.doi.org/hb.html

Agencias para registro de DOIs

https://www.doi.org/registration_agencies.html

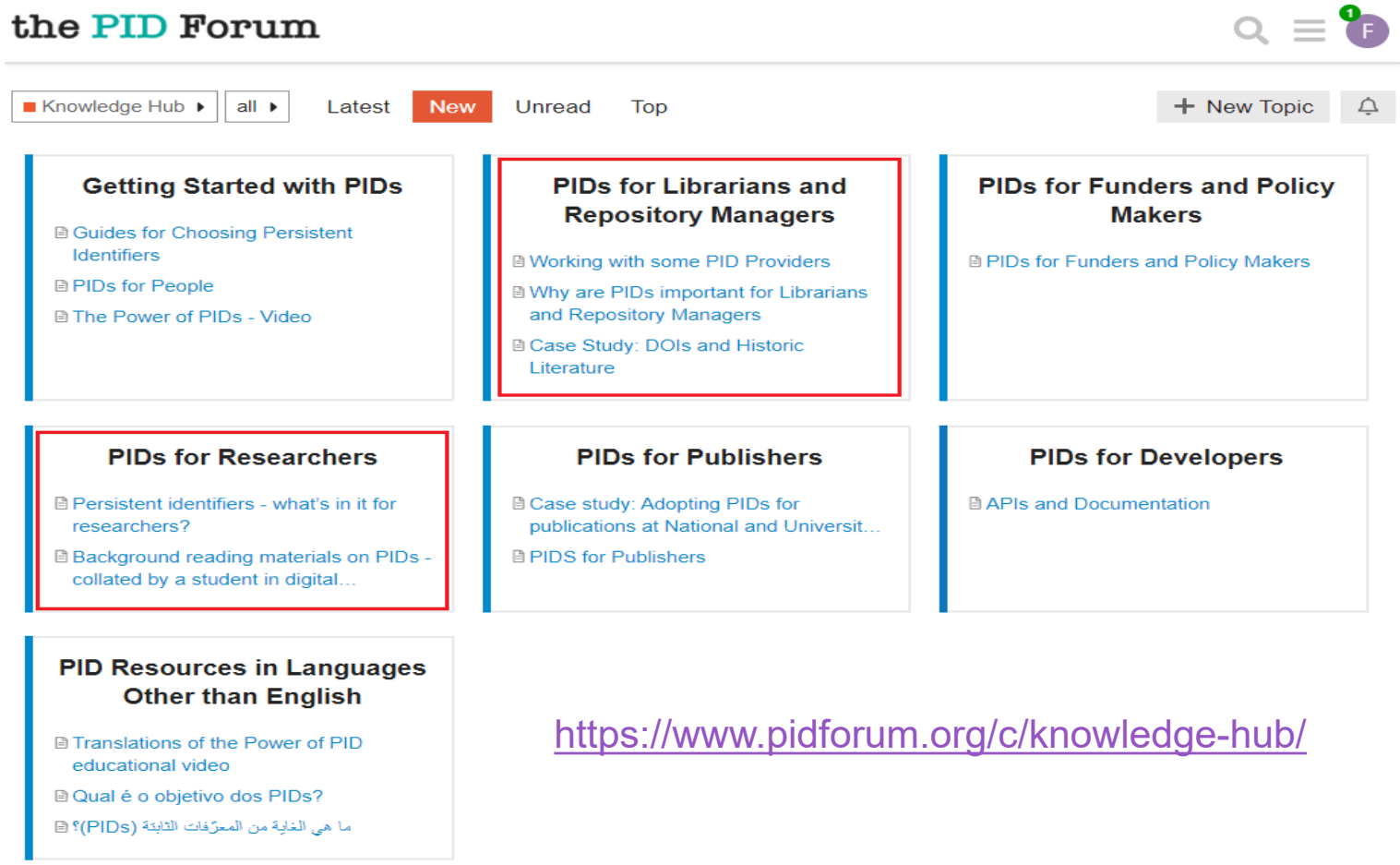
Registration Agency	Coverage
Airiti, Inc.	•DOI applications to Traditional Chinese materials.
China National Knowledge Infrastructure (CNKI)	•China-based information resources, including Chinese politics, economics, humanities, social science, science, and technology. •CNKI publishes databases containing e-journals, newspapers, dissertations, proceedings, yearbooks, references works, and more.
Crossref	•Scholarly and professional research content. •Journal articles, books, conference proceedings, etc. •Reference linking and searchable metadata database.
DataCite	•DataCite is a global non-profit organization that provides DOIs for research data and all other research outputs. •By assigning DataCite DOIs, research outputs become discoverable and associated metadata is made available to the community. •DataCite develops additional services to make it easy to connect and share research outputs with the broader research ecosystem and to assess the use of outputs within that ecosystem. •All organizations within the research community can join DataCite to start registering DOIs.
Multilingual European DOI Registration Agency (*m*EDRA)	•Persistent citation system for Internet documents. •Relation tracking between intellectual property entities. •Certification of voluntary deposit including time stamping and digital signatures.
Publications Office of the European Union (OP)	•OP is the official publisher of the institutions, bodies, offices and agencies of the European Union. As such, it is responsible for assigning DOI names on behalf of these clients. Coverage includes the identification of all EU monographs, the Official Journal of the EU and its individual acts, as well as a number of scientific articles.

Otros identificadores persistentes PIDs

- Para personas, el más destacado es ORCID https://orcid.org/.
- Para instituciones (ROR - Organization Registry Community https://ror.org/; Ringgold https://www.ringgold.com/.
- Para financiadores (FundRef, el registro de organismos financiadores de Crossref, https://www.crossref.org/services/funder-registry/).

Para más información sobre la conexión e integración entre los distintos sistemas de identificadores persistentes es interesante consultar el **PID Forum** https://www.pidforum.org/.

El PID Forum



https://www.pidforum.org/c/knowledge-hub/

Almacenamiento durante el proyecto

Tipos de almacenamiento

El que se requiere durante la ejecución del proyecto (en general interno y privado), y el lugar de almacenamiento de cara a la difusión y a la preservación de los materiales.

Las cuestiones esenciales a las que se ha de dar respuesta son:

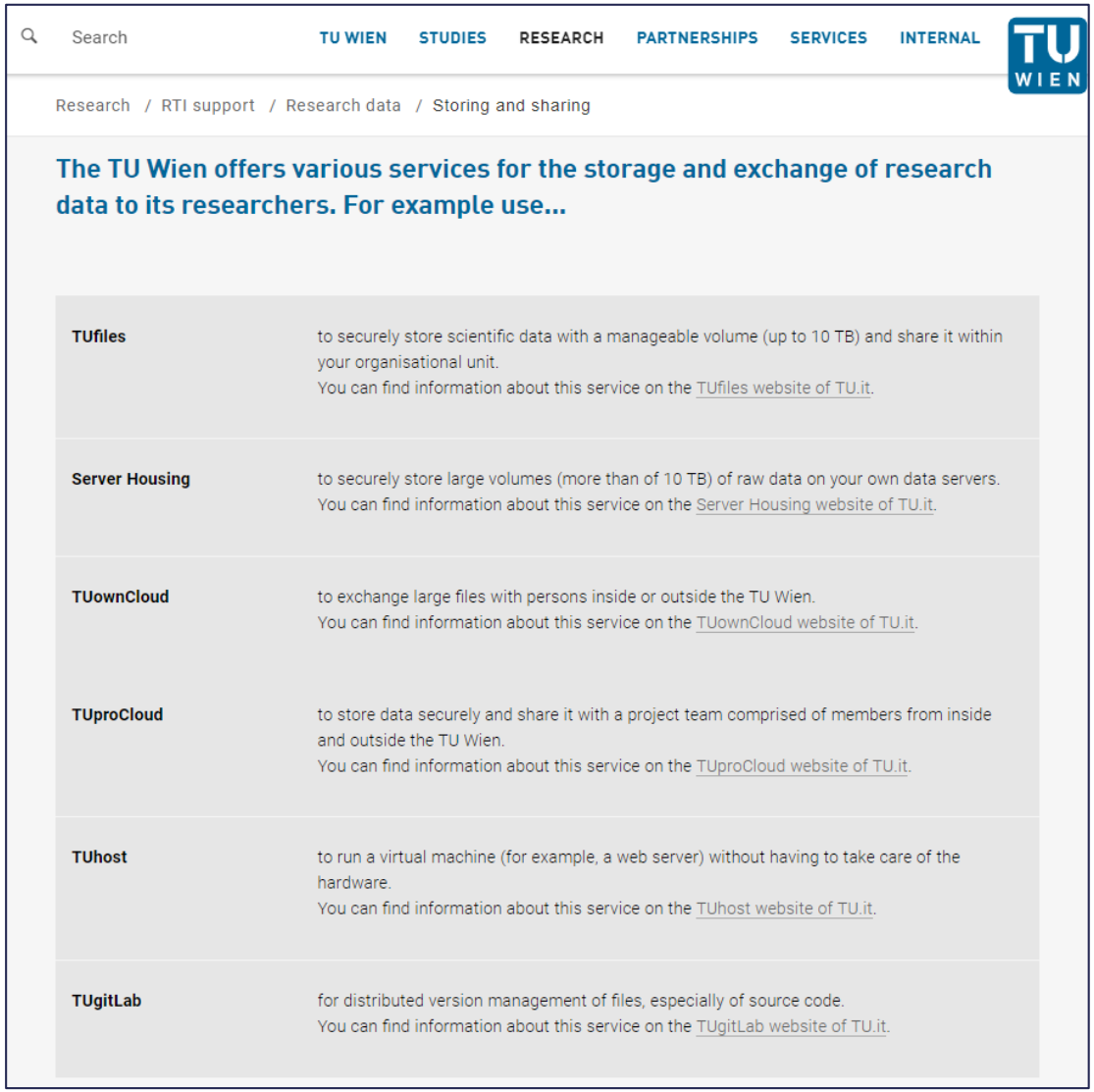
- ¿Dónde se almacenarán sus datos y cómo se harán las copias de seguridad?
- ¿Quién será responsable de controlar el acceso a sus datos y cómo se controlará que el acceso sea seguro?

Para agilizar el acceso es se podrán utilizar sistemas de almacenamiento en la nube, preferentemente que estén aprobados por las instituciones y que no presenten riesgos legales.

Por ejemplo Dropbox y Google Drive, incluyen en sus contratos renuncias de responsabilidad legal, de modo que en caso de que hubiera algún problema con la seguridad de archivos a terceros el responsable del tratamiento de datos es el usuario.

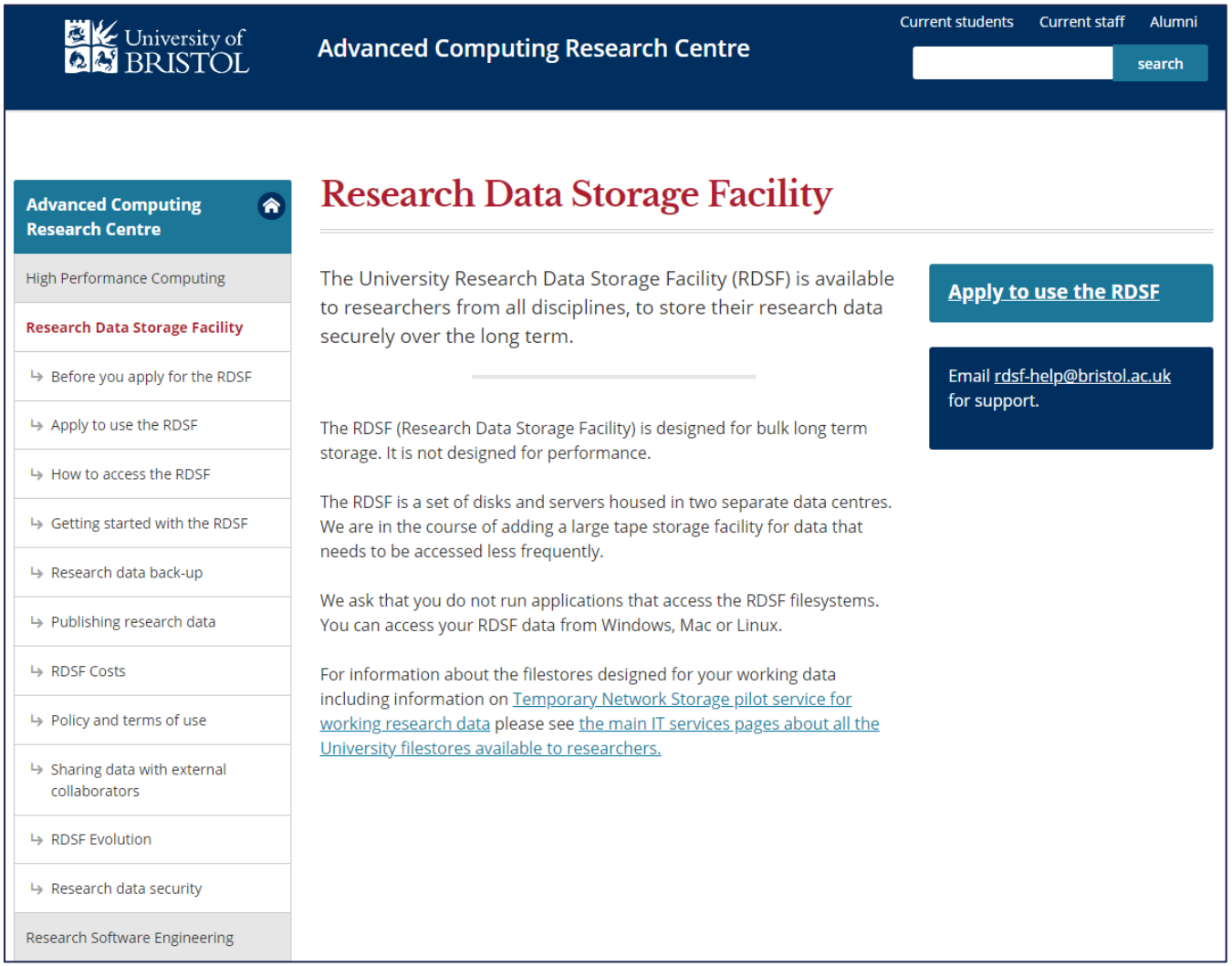
Opción de almacenamiento institucional

Comprobar si se pueden utilizar servicios institucionales, preferentemente en la nube, que permitan el acceso tanto al personal interno como a los colaboradores externos del proyecto.



https://www.tuwien.at/en/research/rti-support/research-data/storing-and-sharing

Ejemplo de almacenamiento institucional de la Universidad de Bristol



http://www.bristol.ac.uk/acrc/research-data-storage-facility/

Almacenamiento temático

En algunas disciplinas hay también servicios de almacenamiento, entre otros muchos recursos:

- *BlueBridge* https://bluebridge.d4science.org/home para las ciencias del mar
- *Elixir* https://elixir-europe.org/ para ciencias de la vida
- *Software Heritage* https://www.softwareheritage.org/?lang=es, para software en formato de código fuente)

Para un almacenamiento a largo plazo, que además permita la difusión y preservación de los datos sin duda lo más conveniente es la elección de un repositorio apropiado. Los repositorios pueden ser institucionales, temáticos o generales.

Licencias y aspectos legales y éticos

Licencias para la reutilización de datos

- Una **licencia** es un contrato legal que especifica **restricciones y permisos estándar para un trabajo**, que luego puede ser utilizado por cualquier persona a quien se le otorgue la licencia, aunque los derechos de propiedad no se transfieren.
- Las licencias dependerán del tipo de datos, de si están sujeto a legislación, así como de los usos permitidos o restricciones aplicables.
- Recursos:
  - *Margoni T, Tsiavos P. (2018). Toolkit for Researchers on Legal Issues. Zenodo. DOI 10.5281/zenodo.2574618*
  - *OpenAIRE (2020) Guides for Researchers. How do I know if my research data is protected? [en línea] https://www.openaire.eu/how-do-i-know-if-my-research-data-is-protected*

La licencia se puede:

- elegir al cargar sus datos en un repositorio;
- especificar haciendo referencia a la licencia en la página desde la que se obtendrán los datos de investigación;
- adjuntar a los metadatos que acompañan a los datos de la investigación;
- especificar en el archivo Readme.txt que acompaña a los datos.

Algunos de los aspectos que se han de tener en cuenta son:

- Si una parte de los datos de la investigación no puede estar disponible.
- Si los datos están sujetos a derechos de autor o derecho sui generis.
- Si el investigador tiene alguna obligación de confidencialidad o ética con respecto a los datos.

Licencias Creative Commons

- La idea surgió a finales de los años 90 con el objetivo de establecer un modelo normalizado para proteger la propiedad intelectual de los autores, permitiendo su reutilización bajo condiciones concretas.
- Sirven para que el creador de una obra pueda decidir bajo qué condiciones de uso se puede utilizar su música, imagen, obra de arte, etc.
- Creative Commons (CC) se refiere a una organización sin fines de lucro que surgió a fines de 2001 en los Estados Unidos y fue fundada por los profesores Lawrence Lessing, de la Universidad de Stanford, y James Boyle, de la Facultad de Derecho de Duke.

Las licencias Creative Commons contemplan 4 posibles condiciones:

	Atribución (BY)	El beneficiario de la licencia tiene el derecho de copiar, distribuir, exhibir y representar la obra y hacer obras derivadas siempre y cuando reconozca y cite la obra de la forma especificada por el autor o el licenciante.
	No Comercial (NC)	El beneficiario de la licencia tiene el derecho de copiar, distribuir, exhibir y representar la obra y hacer obras derivadas para fines no comerciales.
	Sin Obra Derivada (ND)	El beneficiario de la licencia solamente tiene el derecho de copiar, distribuir, exhibir y representar copias literales de la obra y no tiene el derecho de producir obras derivadas.
	Compartir Igual (SA)	El beneficiario de la licencia tiene el derecho de distribuir obras derivadas bajo una licencia idéntica a la licencia que regula la obra original.

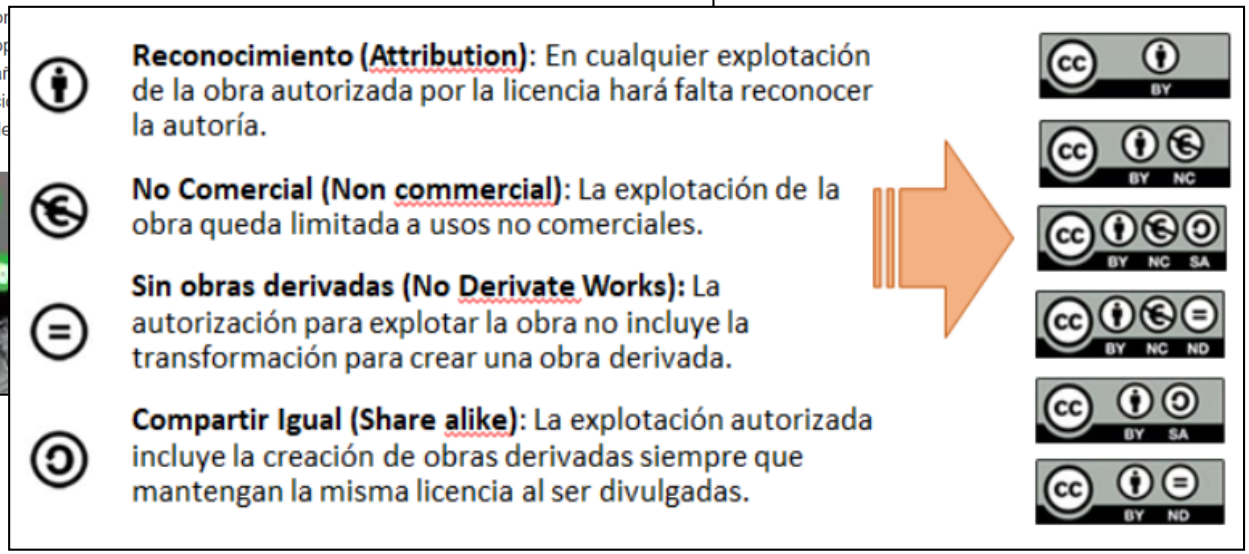
En función de cómo se combinen dan lugar a 6 tipos de licencias:

	Atribución	CC BY
	Atribución – Compartir Igual	CC BY-SA
	Atribución – Sin Obra Derivada	CC BY-ND
	Atribución – No Comercial	CC BY-NC
	Atribución – No Comercial – Compartir Igual	CC BY-NC-SA
	Atribución – No Comercial – Sin Obra Derivada	CC BY-NC-ND

Recomendaciones de OpenAIRE en el uso de licencias CC:

- Si los datos de investigación se refieren a un artículo de revista o un software, la CC BY 4.0 suele ser la mejor opción.
- El uso de Share Alike (SA) también es compatible con la definición de acceso abierto.
- Se debe **evitar el uso no comercial ya que no es compatible con Open Access**. La **opción de no derivado** es compleja y debe evitarse, ya que puede resultar no ser compatible con la definición de acceso abierto.
- Si los datos de investigación son una base de datos o un conjunto de datos (datos no estructurados que no cumplen con la definición de la base de datos), por lo general, la mejor opción es un CC0, que implica una renuncia a todos los derechos sobre la base de datos.
- Las licencias CC solo se refieren a derechos de autor y asuntos relacionados con los derechos de autor. Los datos personales no están cubiertos por las CC y se han de tratar por separado.

OpenAIRE. *Licenses for Research Data.* [en línea] Disponible en: https://www.openaire.eu/how-do-i-license-my-research-data. [Consultada: noviembre de 2020].



https://creativecommons.org/licenses/?lang=es_ES

Open Data Commons (i)

- *Open Data Commons Open Database License* o licencia de base de datos abierta de Open Data Commons (ODbL): Atribución y uso compartido para datos o bases de datos.

  Esta licencia permite al usuario compartir (copiar, distribuir y utilizar la base de datos), crear (producir derivados de la base de datos) y adaptar (modificar, transformar y crear a partir de dicha base de datos), siempre y cuando:

– Se atribuya de forma explícita la autoría del conjunto de datos originales y a sus términos de uso expresados en la licencia;

– Se comparta igual, si se realizan obras derivadas, se han de ofrecer bajo la misma licencia de uso (oDbL);

– Se mantengan abiertas, si se realizan versiones o adaptaciones con restricciones de acceso se ha de garantizar la disponibilidad de una copia en acceso abierto.

Open Data Commons Open Database License (ODbL) [en línea]
Disponible en: https://opendatacommons.org/licenses/odbl/

Open Data Commons (ii)

- *Open Data Commons Attribution License* o licencia de atribución de Open Data Commons (ODC-By): Atribución de datos o bases de datos.

  Al igual que la anterior, esta licencia permite al usuario compartir (copiar, distribuir y utilizar la base de datos), crear (producir derivados de la base de datos) y adaptar (modificar, transformar y crear a partir de dicha base de datos), siempre y cuando (aquí está la diferencia):

– Se atribuya de forma explícita la autoría del conjunto de datos originales y a sus términos de uso expresados en la licencia.

- *Open Data Commons Public Domain Dedication* and License o licencia de dominio público de Open Data Commons (PDDL), que permite difundir, reutilizar o adaptar los datos sin restricción alguna

Open Data Commons Open Database License (ODbL) [en línea] Disponible en: https://opendatacommons.org/licenses/odbl/

Open Data Commons

LEGAL TOOLS FOR OPEN DATA

Home

Open Data Commons is the home of a set of legal tools and licenses to help you publish, provide and use open data.

Learn more about how to apply the licenses to your material:

- Open Data Commons Open Database License (ODbL)
- Open Data Commons Attribution License
- Open Data Commons Public Domain Dedication and License (PDDL)

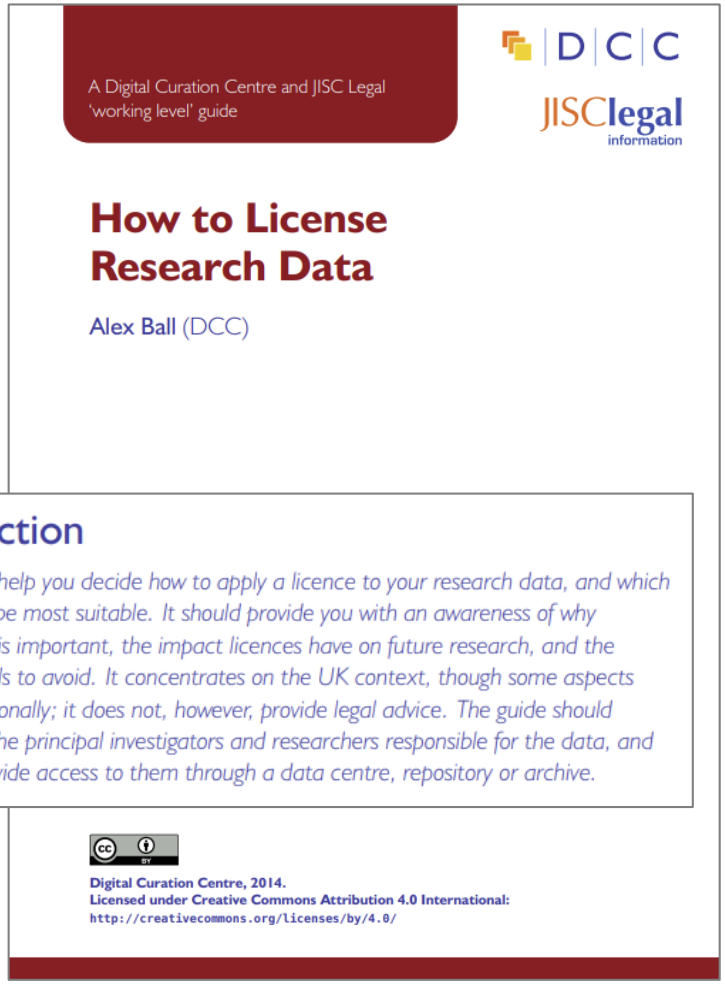
To discover more about open data and content, use these resources:

- What is open data and how does it relate to open knowledge? Consult our guide
- How is "openness" defined in relation to data and content? Read the Open Definition
- Have more detailed questions about the Open Data Commons licenses? Our frequently asked questions may provide answers
- Use the Licenses Service to get data on more than 100 open source, open data and open content licenses in JSON and API friendly form

Open Data Commons is maintained by the Open Knowledge Foundation. If you have any questions, join our discussion forum or email info@okfn.org to contact our team.

https://opendatacommons.org/

https://www.dcc.ac.uk/sites/default/files/documents/publications/reports/guides/How_To_License_Research_Data.pdf



A Digital Curation Centre and JISC Legal 'working level' guide

DCC

JISClegal information

How to License Research Data

Alex Ball (DCC)

Introduction

This guide will help you decide how to apply a licence to your research data, and which licence would be most suitable. It should provide you with an awareness of why licensing data is important, the impact licences have on future research, and the potential pitfalls to avoid. It concentrates on the UK context, though some aspects apply internationally; it does not, however, provide legal advice. The guide should interest both the principal investigators and researchers responsible for the data, and those who provide access to them through a data centre, repository or archive.

Digital Curation Centre, 2014.
Licensed under Creative Commons Attribution 4.0 International:
http://creativecommons.org/licenses/by/4.0/

Otras licencias copyleft

- **Copyleft** es un método general para hacer libre un programa u otro tipo de trabajo, que requiere que todas las versiones modificadas y extendidas sean también libres.
  Se aplican principalmente al software, pero no de manera exclusiva, usándose también en el mundo del arte, de la comercialización artística, y cada vez más en las patentes.
  Permiten que el usuario de una obra pueda a su vez modificar y redistribuir el propio trabajo y las versiones derivadas del mismo, permitiendo o no su uso comercial en función de los derechos cedidos por el autor original.
- **Licencia Pública General de GNU** o más conocida por su nombre en inglés GNU General Public License (o simplemente sus siglas en inglés GNU GPL).
  licencia de derecho de autor muy usada en el mundo del software libre y código abierto, cuyo objetivo es garantizar a los usuarios finales (personas, organizaciones, compañías) la libertad de usar, estudiar, compartir (copiar) y modificar el software.

Recursos sobre licencias

- Bezjak, Sonja, el al. (2019). "Licencias abiertas y formatos de ficheros." Manual de Capacitación sobre Ciencia Abierta. Zenodo. http://doi.org/10.5281/zenodo.2588214
- Free Software Foundation. GNU General Public License (2017) [en línea] Disponible en: https://www.gnu.org/licenses/gpl-3.0.en.html
- Choose an open source license. Un selector de licencias que guía al autor en función de su situación: https://choosealicense.com/
- Open Source Initiative, https://opensource.org/faq#which-license, que incluye un apartado con información sobre licencias.
- TLDRLegal https://tldrlegal.com/, un buscador de licencias de software más utilizadas populares resumidas.
- License Selector, https://ufal.github.io/public-license-selector, un recurso muy útil para decidir qué licencia de uso se ajusta mejor a cada tipo de set de datos o de software.

Reglamento General de Protección de Datos (RGPD)

- Es el reglamento europeo relativo a la protección de las personas físicas en lo que respecta al tratamiento de sus datos personales y a la libre circulación de estos datos. Entró en vigor el 25 de mayo de 2016 y fue de aplicación el 25 de mayo de 2018.
- En España, la legislación acorde a la RGDP es la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales (LOPD-GDD), que sustituyó a la anterior Ley Orgánica 15/1999 de Protección de Datos de Carácter Personal (LOPD).

Reglamento General de Protección de Datos (RGPD) de 27 de abril de 2017. [en línea] Disponible en: https://www.boe.es/doue/2016/119/L00001-00088.pdf.
Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales. [en línea] Disponible en: https://www.boe.es/buscar/doc.php?id=.

Los datos afectados por el RGDP pueden ser:

- Datos personales, relacionados con una persona física que se puede identificar por atributos como un nombre, un número de identificación, edad, ubicación, etc.
- Datos personales sensibles. referidos a información relacionada con el origen racial o étnico de una persona, opiniones políticas, religión, datos de salud, vida sexual, genéticos, biométricos o delitos penales.

Soluciones:

- Datos seudonimizados, que siguen siendo datos personales pero que no permiten identificar a un sujeto, ya que la información adicional que permitiría dicha identificación se mantiene por separado (ejemplo de seudonimización es la sustitución de un nombre por un código o por un seudónimo).
- Datos anonimizados, referidos a información que no puede asociarse a una persona identificable o datos personales anónimos de tal manera que el interesado no es o ya no es identificable. La anonimización pueden ser por generalización, por aleatorización y por eliminación.

Según el RGPD, **para que el tratamiento de los datos sea lícito**, deberá contar con al menos una de las siguientes condiciones:

- Que los datos sean de interés público, algo que por supuesto ha de ser demostrado.
- Contar con intereses legítimos del responsable.
- Que haya una obligación contractual o legal.
- Que el tratamiento de los datos corresponda a intereses vitales.
- Tener el consentimiento del interesado, habiendo este dado su conformidad de forma explícita, de forma libre, una vez que ha sido informado con claridad de cómo y con qué fin se van a usar esos datos, y con el derecho a retirar dicho consentimiento.

Reutilización o uso secundario de datos personales

- La reutilización o uso secundario de datos personales está permitida en el RPGD para temas de investigación, pero solo cuando el responsable disponga de medidas técnicas de seguridad y privacidad que garanticen la protección de los derechos de los interesados.
- Estas medidas pueden ser la seudonimización, la anonimización, la encriptación, el control restringido o limitado de acceso.
- No obstante, aunque el uso secundario de datos sea legal, los interesados mantienen su derecho a ser informados sobre los nuevos tratamientos ejercidos sobre los datos.

Recomendaciones para la formulación del consentimiento para la reutilización de datos personales

- Especificar desde un inicio que uno de los propósitos es compartir los datos con la comunidad científica.
- Para compartir datos de investigación es importante contar con un acuerdo de transferencia de datos o *Data Transfer Agreement* que aseguren que se cumple con todos los requisitos legales.
- Es conveniente explicar con claridad qué datos son los que podrán ser publicados.
- Es preferible no prometer que los datos van a ser destruidos, para ello hay que estar completamente seguros de que los datos no van a ser necesarios para llevar a cabo comprobaciones ni otras acciones.
- Es mejor no afirmar que los datos serán vistos únicamente por el equipo de investigación, en caso de que así sea conviene especificar qué datos podrán ser compartidos en abierto o a terceros, y que datos serán de uso privado entre el equipo de investigación.

Anonimización y seudonimización

Será necesario anonimizar o seudonimizar datos para compartir dichos datos con investigadores y terceros sin comprometer la privacidad del usuario cuando:

- se quiera compartir o ceder datos a destinatarios con los que no se ha firmado un acuerdo de confidencialidad.
- se quiera publicar datos abiertamente.
- la reducción de la calidad de la información sea aceptable y no afecte al uso de los datos.

Una vez que los datos se anonimizan, no sería necesario el consentimiento. Si las personas pudieran ser identificadas de alguna forma es contar con un formulario de consentimiento informado, firmado por los participantes.

El proceso de seudonimización y anonimización implica que los datos sobre los individuos se vean alterados a través de distintos procesos: pueden ser suprimidos, sustituidos, distorsionados, generalizados o agregados. Es importante que esto no afecte al posterior uso de la información.

Ejemplos de eliminación y generalización de datos:

id	Zipcode	Age	National.	Disease
1	13053	28	Russian	Heart Disease
2	13068	29	American	Heart Disease
3	13068	21	Japanese	Viral Infection
4	13053	23	American	Viral Infection
5	14853	50	Indian	Cancer
6	14853	55	Russian	Heart Disease
7	14850	47	American	Viral Infection
8	14850	49	American	Viral Infection
9	13053	31	American	Cancer
10	13053	37	Indian	Cancer
11	13068	36	Japanese	Cancer
12	13068	35	American	Cancer

~~id~~	Zipcode	Age	National.	Disease
~~1~~	130**	<30	*	Heart Disease
~~2~~	130**	<30	*	Heart Disease
~~3~~	130**	<30	*	Viral Infection
~~4~~	130**	<30	*	Viral Infection
~~5~~	1485*	≥40	*	Cancer
~~6~~	1485*	≥40	*	Heart Disease
~~7~~	1485*	≥40	*	Viral Infection
~~8~~	1485*	≥40	*	Viral Infection
~~9~~	130**	3*	*	Cancer
~~10~~	130**	3*	*	Cancer
~~11~~	130**	3*	*	Cancer
~~12~~	130**	3*	*	Cancer

Alicia Fatima Gomez Sanchez, & Elli Papadopoulou. (2020, November). *OpenAIRE tools and resources: supporting research data management services for libraries and researchers*. http://doi.org/10.5281/zenodo.4317181

Seudonimización

- ✓ Los identificadores directos se reemplazan por otros arbitrarios, que deben mantenerse por separado.
- ✓ “El procesamiento de datos personales hace que los datos ya no puedan atribuirse a un sujeto de datos específico sin el uso de información adicional".(WG)
- ✓ La seudonimización no es un método de anonimización

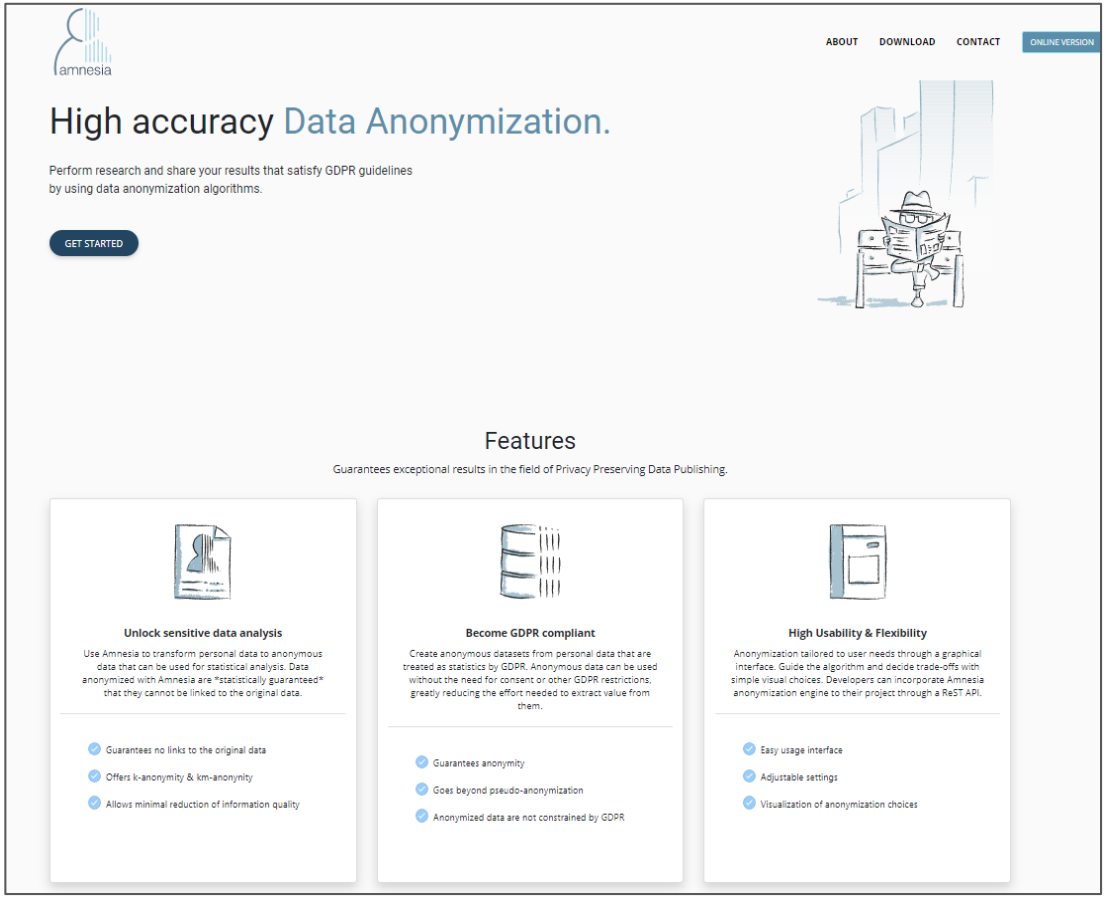
Anonimización

- ✓ Es irreversible
- ✓ Los datos dejan de ser personales

Generalización

- ✓ Se utilizan jerarquías para reemplazar los valores detallados por otros más genéricos.
- ✓ La idea clave aquí es que los valores que son lo suficientemente específicos como para ser identificadores (CP) se reemplazan por otros más genéricos (la ciudad) de modo que ya no puedan revelar la identidad

Amnesia



20200610_OpenAIRE webinar: Amnesia

Amnesia is a flexible data anonymization tool that transforms relational and transactional databases to dataset where formal privacy guaranties hold. Amnesia transforms original data to provide k-anonymity and km-anonymity

RDC Webinar: Data Deidentification and Anonymization: An Introduction

This webinar will introduce the concepts of deidentification and anonymization of research data, including direct identifiers and quasi-identifiers, risk assessment, and common strategies to reduce risk.

https://amnesia.openaire.eu/

Depósito en repositorios

Repositorios

Requisitos a tener en cuenta en la selección de un repositorio:

- Que cumpla con los criterios FAIR
- Que permita almacenar datos abiertos, con embargo, restringidos o cerrados, en especial para datos sensibles
- Que sirva para materiales diversos de investigación (distintos tipos, formatos, etc.)
- Que proporcione visibilidad, para facilitar la citación y aumentar el impacto
- Que se integre con otras infraestructuras
- Que sea transparente y fiable

Repositorios institucionales

- Muchas instituciones están trabajando en sus repositorios de datos. El CSUC por ejemplo:

  "Trabajamos en el desarrollo de un repositorio de datos que permita la publicación de conjuntos de datos de investigación en modo FAIR por parte de las universidades catalanas y algunos centros de investigación CERCA. El futuro repositorio, que se prevé que se abra al público a lo largo de 2021, también servirá para desarrollar experiencia y buenas prácticas en curación de datos."

https://www.csuc.cat/es/servicios/gestion-de-datos-de-investigacion

Repositorios institucionales

- Otras instituciones han adaptado su repositorio para alojar datos, como es el caso de Digital.CSIC.

  "DIGITAL.CSIC es un repositorio válido para almacenar y dar difusión a una amplia variedad de tipos de datos de investigación, facilita DOIs para datos de investigación y software a través de DataCite y está certificado por diversas iniciativas internacionales que evalúan repositorios de datos tales como Data Seal of Approval y Re3data."

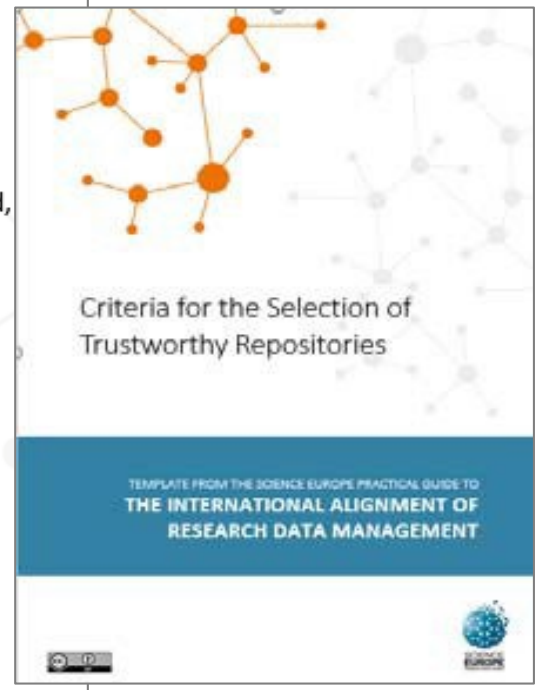
http://bibliotecas.csic.es/node/293

1. **Provision of Persistent and Unique Identifiers (PID)**
   a. Allow data discovery and identification
   b. Enable searching, citing, and retrieval of data
   c. Provide support for data versioning

2. **Metadata**
   a. Enable finding of data
   b. Enable referencing to related relevant information, such as other data and publications
   c. Provide information that is publicly available and maintained, even for non-published, protected, retracted, or deleted data
   d. Use metadata standards that are broadly accepted (by the scientific community)
   e. Ensure that metadata are machine-retrievable

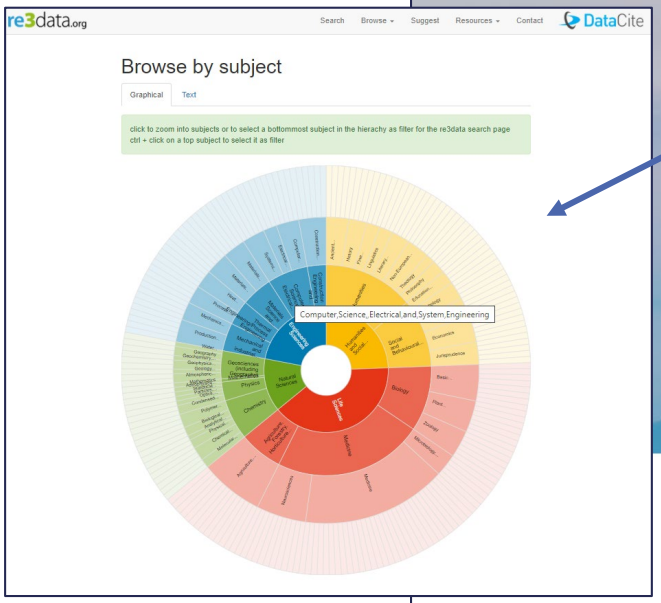
3. **Data Access and Usage Licences**
   a. Enable access to data under well-specified conditions
   b. Ensure data authenticity and integrity
   c. Enable retrieval of data
   d. Provide information about licensing and permissions (in ideally machine-readable form)
   e. Ensure confidentiality and respect rights of data subjects and creators

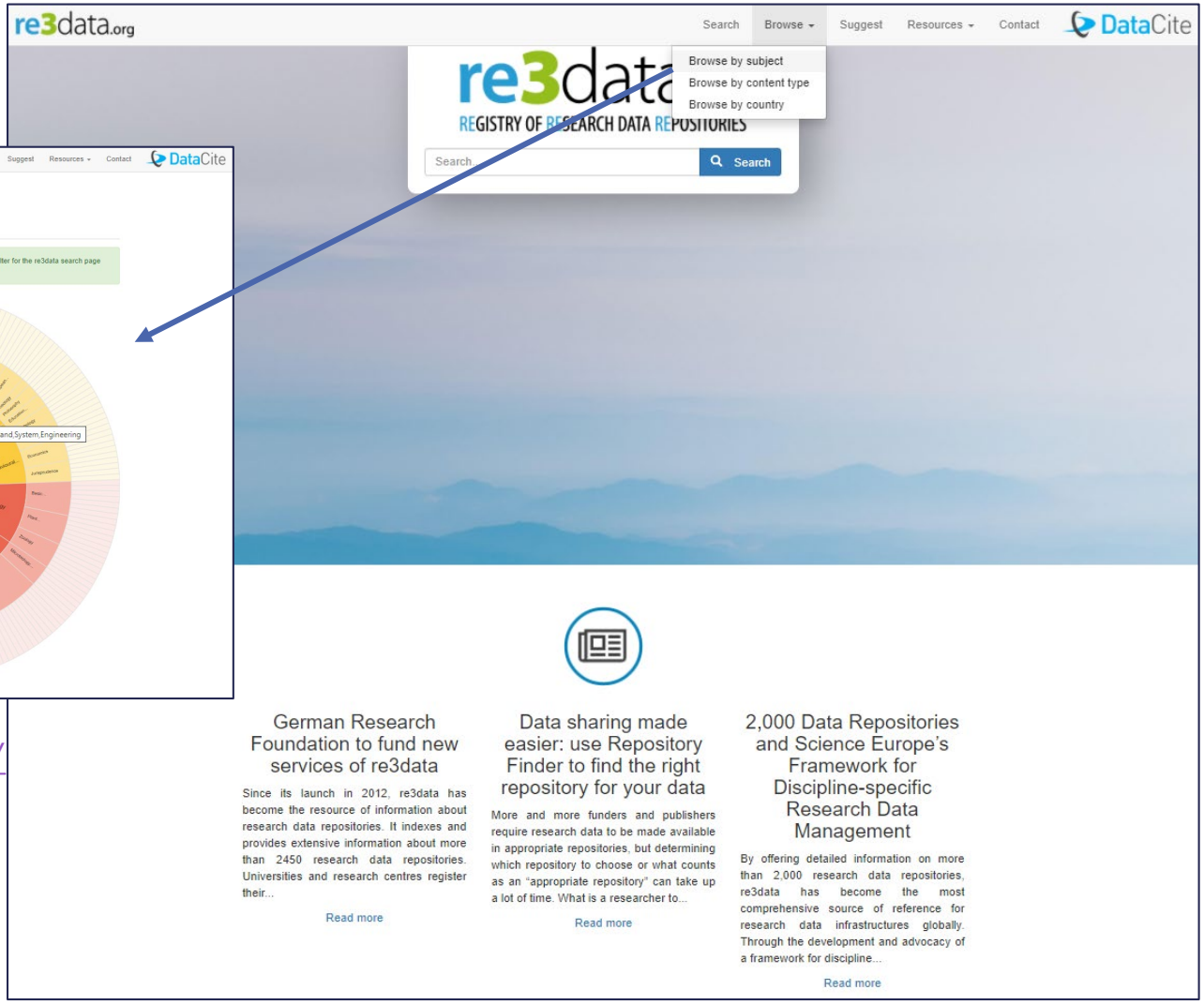
4. **Preservation**
   a. Ensure persistence of metadata and data
   b. Be transparent about mission, scope, preservation policies, and plans (including governance, financial sustainability, retention period, and continuity plan)



https://www.scienceeurope.org/media/ffkb51ei/se-rdm-template-2-criteria-for-the-selection-of-trustworthy-repositories.docx

re3data





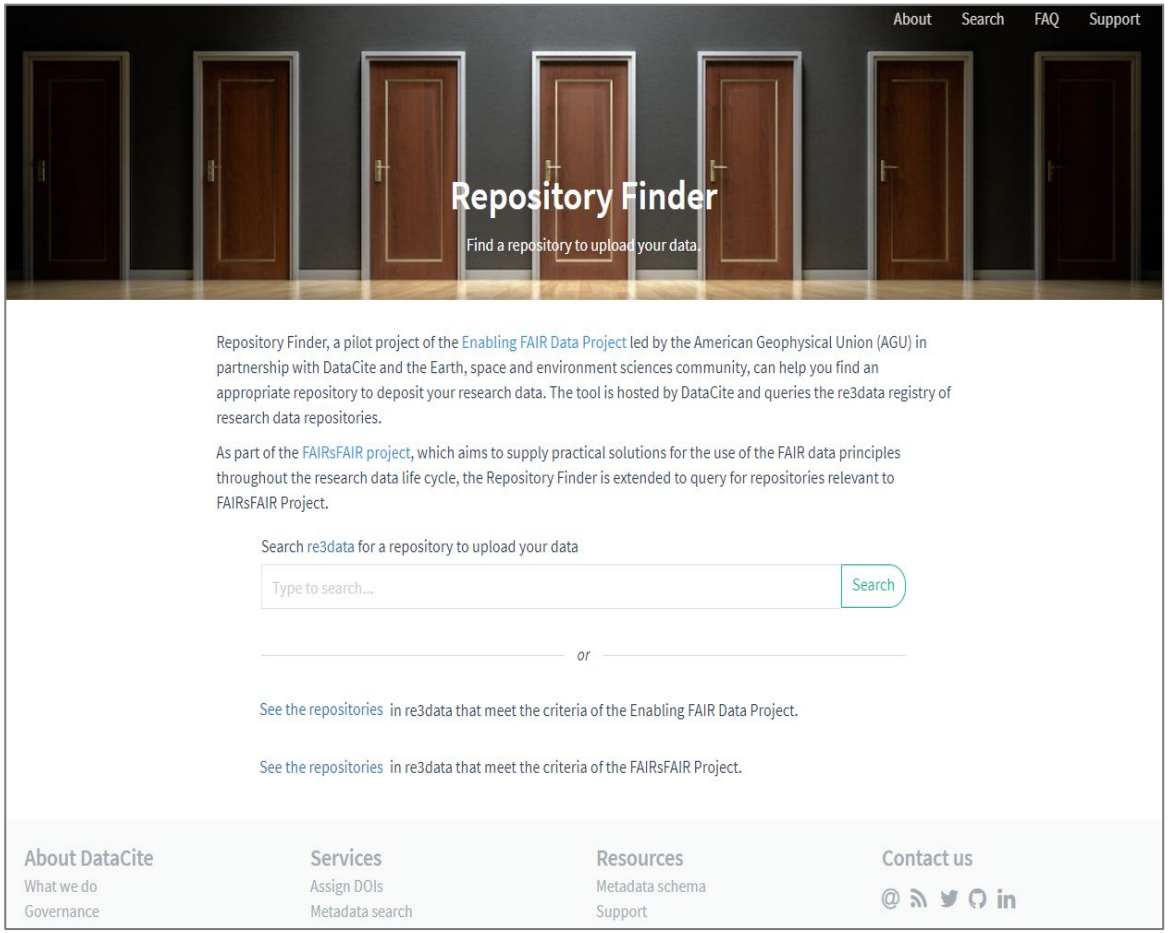
https://www.re3data.org/

La Comisión Europea recomienda el uso de re3data en sus "Directrices sobre acceso abierto a publicaciones científicas y datos de investigación en Horizonte 2020" (Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020)

Repository Finder

Proyecto piloto del Proyecto *Enabling FAIR Data* dirigido por la American Geophysical Union (Unión Geofísica Estadounidense, AGU) en colaboración con DataCite y la comunidad de ciencias de la Tierra, el espacio y el medio ambiente.

Repository Finder **incluye repositorios temáticos que siguen los Principios FAIR**

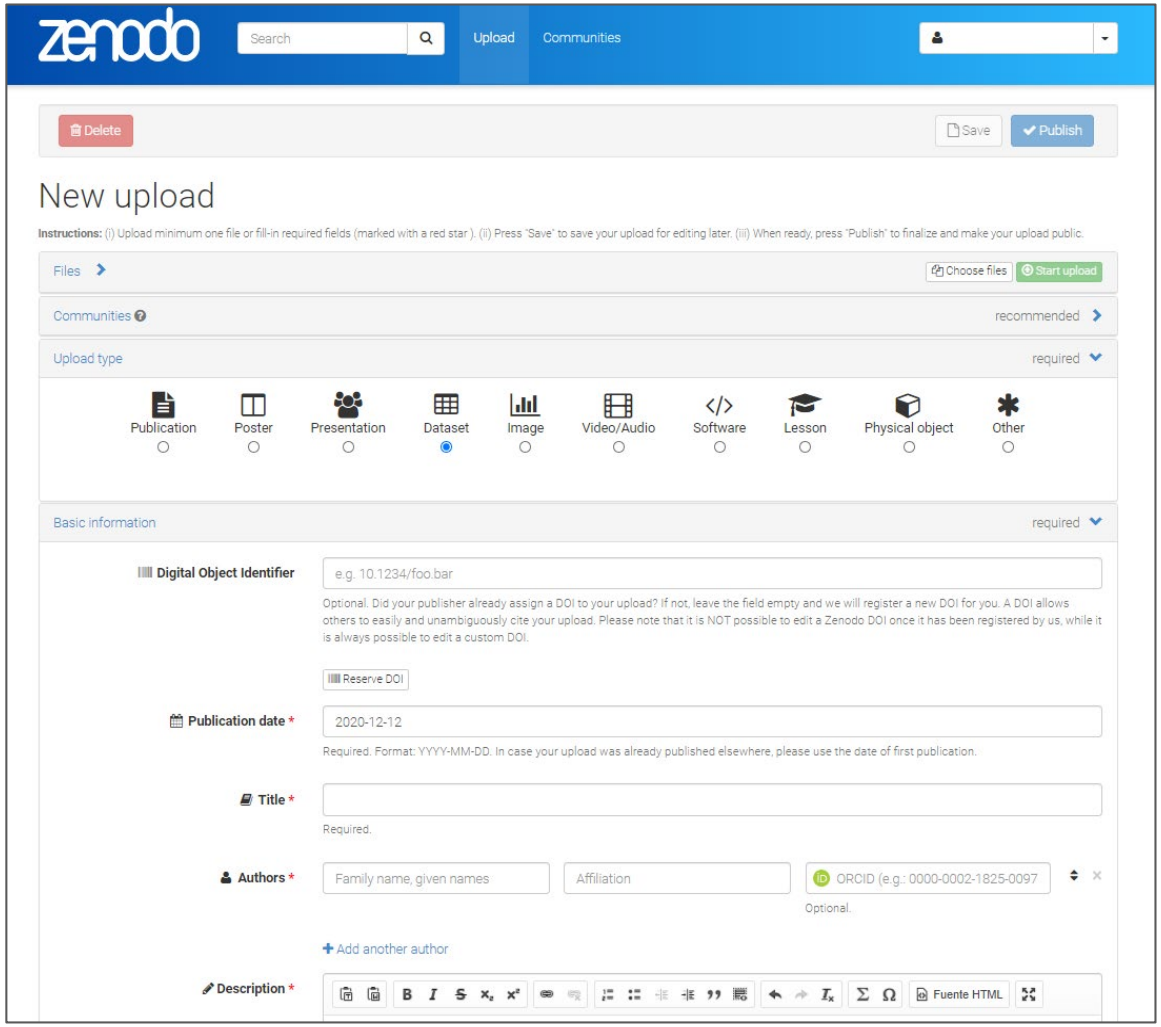


https://repositoryfinder.datacite.org/

Zenodo

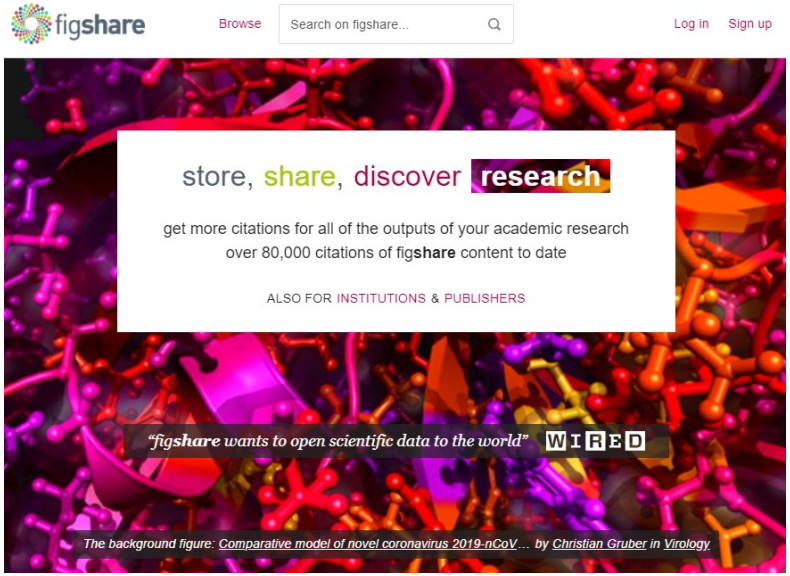
Repositorio de acceso abierto desarrollado en 2013 bajo el programa europeo OpenAIRE y gestionado por y alojado en la Organización Europea para la Investigación Nuclear (CERN).

Desde 2015 admite también el depósito de conjuntos de datos, permite carga de archivos hasta 50 GB. Proporciona DOIs al y asigna metadatos de acceso abierto, licencias, etc.

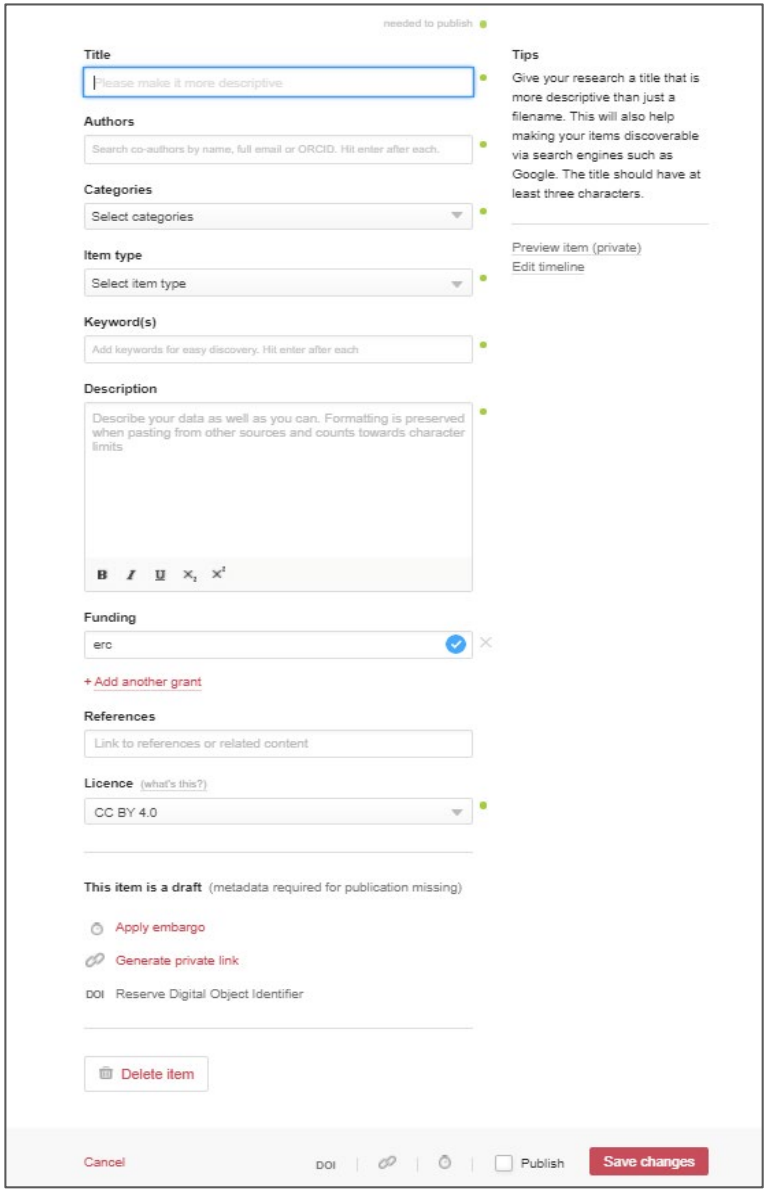


Interfaz de Zenodo para subir un objeto (un set de datos)
https://zenodo.org/deposit/new
Para más información https://help.zenodo.org/

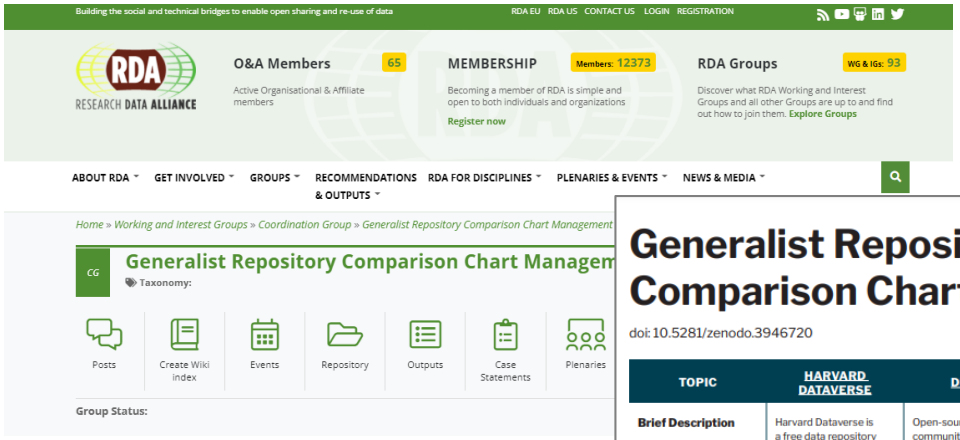
Figshare



Figshare es un repositorio de gestión privada. Se lanzó en 2011 por Mark Hahnel y desde 2012 está financiado por empresa Digital Science.
Permite cargar cualquier tipo de resultados de investigación y en cualquier formato y con un tamaño límite de 20 GB, atribuye DOIs, permite establecer embargos, añadir licencias, enlazar con proyectos financiados, etc.



Interfaz de Figshare para subir un objeto https://figshare.com/



https://www.rd-alliance.org/groups/generalist-repository-comparison-chart-management-group

Generalist Repository Comparison Chart

This chart is designed to assist researchers in finding a generalist repository should no domain repository be available to preserve their research data. Generalist repositories accept data regardless of data type, format, content, or disciplinary focus. For this chart, we included a repository available to all researchers specific to clinical trials (Vivli) to bring awareness to those in this field.

doi: 10.5281/zenodo.3946720

https://fairsharing.org/collection/GeneralRepositoryComparison

TOPIC	HARVARD DATAVERSE	DRYAD	FIGSHARE	MENDELEY DATA	OSF	VIVLI	ZENODO
Brief Description	Harvard Dataverse is a free data repository open to all researchers from any discipline, both inside and outside of the Harvard community, where you can share, archive, cite, access, and explore research data.	Open-source, community-led data curation, publishing, and preservation platform for CC0 publicly available research data Dryad is an independent non-profit that works directly with: • researchers to publish datasets utilizing best practices for discovery and reuse • publishers to support the integration of data availability statements and data citations into their workflows • institutions to enable scalable campus support for research data managment best practices at low cost	A free, open access, data repository where users can make all outputs of their research available in a discoverable, reusable, and citable manner. Users can upload files of any type and are able to share diverse research products including datasets, code, multimedia files, workflows, posters, presentations, and more. With discoverable metadata supporting FAIR principles, file visualizations, and integrations, researchers can make their work more impactful and move research further faster.	Mendeley Data is a free repository specialized for research data. Search more than 20+ million datasets indexed from 1000s of data repositories and collect and share datasets with the research community following the FAIR data principles.	OSF is a free and open source project management tool that supports researchers throughout their entire project lifecycle in open science best practices.	Vivli is an independent, non-profit organization that has developed a global data-sharing and analytics platform. Our focus is on sharing individual participant-level data from completed clinical trials to serve the international research community.	Powering Open Science, built on Open Source. Built by reserachers for researchers. Run from the CERN data centre, whose purpose is long term preservation for the High Energy Physics discipline, one of the largest scientific datasets in the world
Size limits	No byte size limit per dataset. Harvard Dataverse currently sets a file size limit of 2.5GB.	300GB/dataset	Soft limit of 20GB/file for free accounts. System limit of 5000GB/file. Unlimited storage of public data but 20GB storage for private data for free accounts. Email info@figshare.com to have upload and storage limits raised.	10GB per dataset	Projects currently have not storage limit. There is a 5GB/file upload limit for native OSF Storage. There is no limit imposed by OSF for the amount of storage used across add-ons connected to a given project.	If more than 10GB per study data, reach out to us	50GB per dataset, contact us via https://zenodo.org/support for higher limits
Storage space per researcher	1 TB per researcher	No limit	No limit	No limit	No limit	No limit	No limit
Persistent, Unique Identifier Support	DOI, Handle	DOI	DOI	DOI	DOI	DOI	DOI

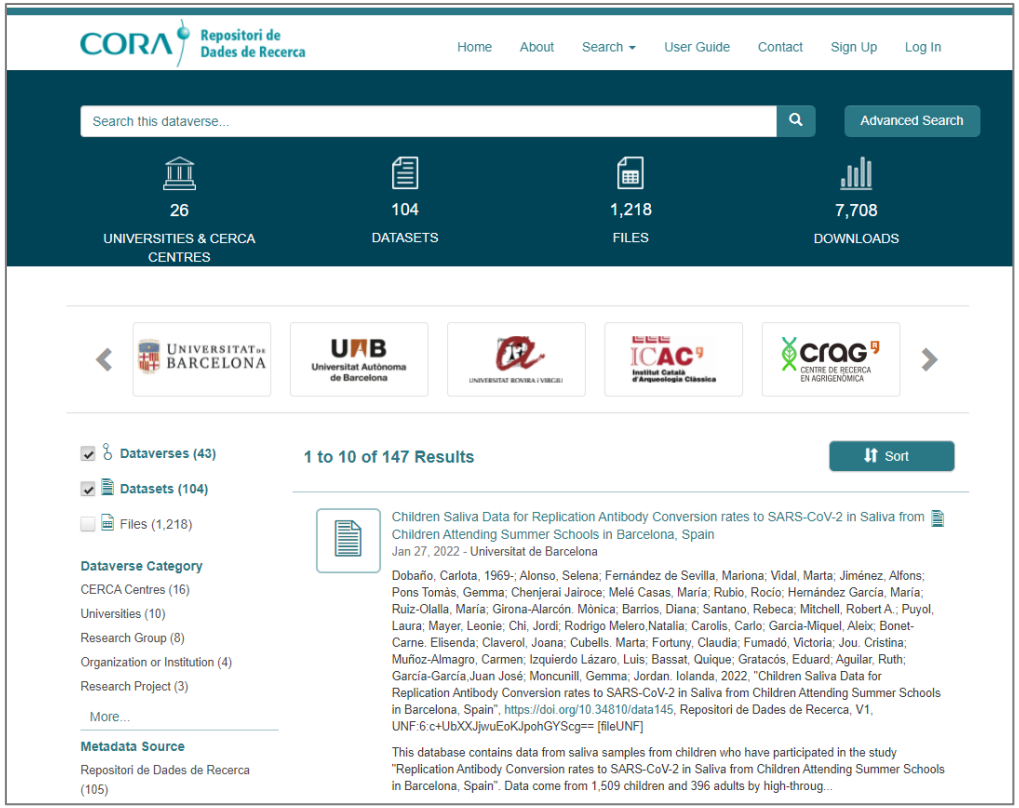
https://www.rd-alliance.org/sites/default/files/Generalist%20Repository%20Comparison%20Chart.pdf

Repositorios institucionales adaptados a datos

http://bibliotecas.csic.es/node/293

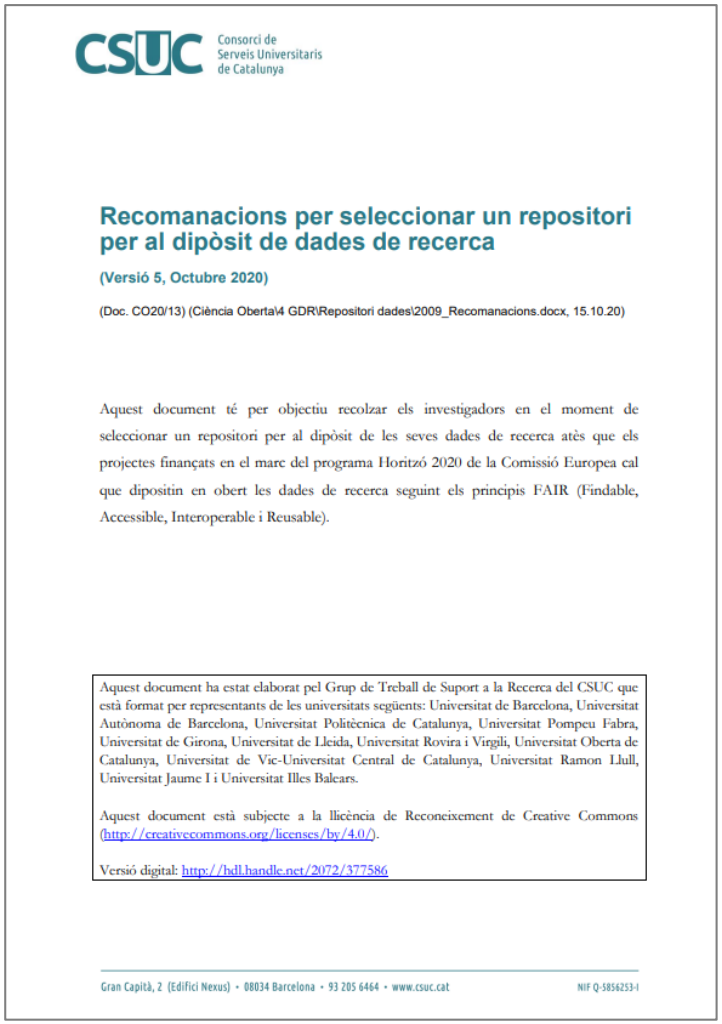
DIGITAL.CSIC es un repositorio válido para almacenar y dar difusión a una amplia variedad de tipos de datos de investigación, facilita DOIs para datos de investigación y software a través de DataCite y está certificado por diversas iniciativas internacionales que evalúan repositorios de datos tales como Data Seal of Approval y Re3data

Repositorios institucionales específicos de datos



https://dataverse.csuc.cat/

CORA es un repositorio de datos federado y multidisciplinar que permite a las universidades catalanas, a los centros de investigación del CERCA y a otras entidades que realizan investigación, publicar conjuntos de datos de investigación en modo FAIR y siguiendo las directrices de la EOSC.



CSUC Consorci de Serveis Universitaris de Catalunya

Recomanacions per seleccionar un repositori per al dipòsit de dades de recerca

(Versió 5, Octubre 2020)

(Doc. CO20/13) (Ciència Oberta\4 GDR\Repositori dades\2009_Recomanacions.docx, 15.10.20)

Aquest document té per objectiu recolzar els investigadors en el moment de seleccionar un repositori per al dipòsit de les seves dades de recerca atès que els projectes finançats en el marc del programa Horitzó 2020 de la Comissió Europea cal que dipositin en obert les dades de recerca seguint els principis FAIR (Findable, Accessible, Interoperable i Reusable).

Aquest document ha estat elaborat pel Grup de Treball de Suport a la Recerca del CSUC que està format per representants de les universitats següents: Universitat de Barcelona, Universitat Autònoma de Barcelona, Universitat Politècnica de Catalunya, Universitat Pompeu Fabra, Universitat de Girona, Universitat de Lleida, Universitat Rovira i Virgili, Universitat Oberta de Catalunya, Universitat de Vic-Universitat Central de Catalunya, Universitat Ramon Llull, Universitat Jaume I i Universitat Illes Balears.

Aquest document està subjecte a la llicència de Reconeixement de Creative Commons (http://creativecommons.org/licenses/by/4.0/).

Versió digital: http://hdl.handle.net/2072/377586

Gran Capità, 2 (Edifici Nexus) · 08034 Barcelona · 93 205 6464 · www.csuc.cat NIF Q-5856253-I

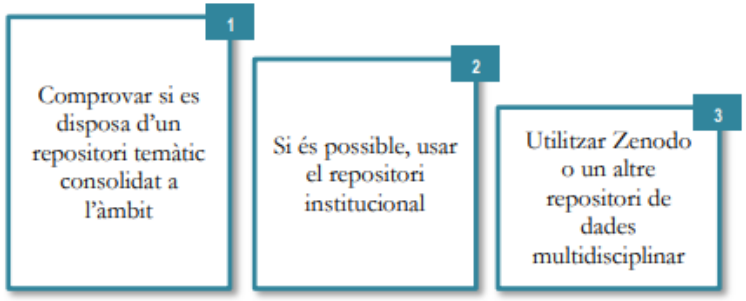


Figura 1. Proposta basada en les recomanacions d'OpenAIRE (vegeu bibliografia)

http://hdl.handle.net/2072/377586

CSUC

	zenodo CERN	B2SHARE EUDAT	figshare Digital Science	DRYAD NCSU	Dataverse Harvard University	MENDELEY DATA Elsevier
1. Àmbit temàtic i geogràfic	Multidisciplinar Internacional	Multidisciplinar Internacional	Multidisciplinar Internacional	Ciència i medicina Internacional	Multidisciplinar Internacional	Multidisciplinar Internacional
2. Tipologia de dades	Tot tipus de formats i també documents publicats	Tot tipus de formats	Tot tipus de formats	Tot tipus de formats incloent arxius comprimits de diversos fitxers amb preferència per formats de codi obert	Tot tipus de formats amb funcionalitats addicionals per alguns	Tot tipus de formats
3. Mida aproximada dels fitxers a incloure	Fins a 50 GB per fitxer. Per a mides superiors, contactar amb ells	Fins a 10 GB per fitxer (màxim de 20 GB per registre). Per a mides superiors, contactar amb ells	Fins a 5 GB per fitxer.	300 GB per un conjunt de dades. Per a mides superiors, contactar amb ells	Fins a 2 GB	Fins 10 GB per conjunt de dades
4. Dades obertes, embargades, restringides o tancades	- Obertes - Embargades - Restringides - Tancades	- Obertes - Embargades - Restringides	- Obertes - Embargades - Restringides - Tancades (per defecte fins que no es publiquen)	- Obertes -Embargades (segons editors i sempre que l'autor ho requereixi mentre no es publica l'article) -Restringides (sempre que l'autor ho requereixi mentre es revisa per parells)	- Obertes - Restringides	- Obertes - Tancades (per defecte fins que es validen els requisits) - Restringides (Per institucions) - Embargades
5. Llicències amb les que es volen difondre les dades	Permet varietat de llicències encara que recomanen llicències obertes. La descripció de les dades està sota CC0.	Permet varietat de llicències per a dades i per a software.	- CC0 per a dades i metadades - CC-BY per a no dades - MIT per a codi i software - Altres llicències: GNU, Apache 2.0, BSD 3 clause.	CC0	Per defecte CC0 però es poden especificar altres termes d'ús.	Permet varietat de llicències per a dades (CC0, CC-BY, CC-BY-NC), software (MIT, Apache, BSD, GPL, LGPL, MPL, CeCILL) i maquinari (CERN, TAPR).
6. Identificador permanent	DOI	Handle i DOI	DOI	DOI	Handle i DOI	DOI
7. Condicions per retirar contingut	- Es permet retirar les dades justificant-ne el perquè. - El DOI i la URL quedaran retinguts.	Sense informació.	Es pot sol·licitar la retirada de contingut públic justificant-ne el motiu.	- Cessió de dades irrevocable. - Es reserva el dret a eliminar el contingut justificant-ne el motiu.	Es permet retirar les dades, tot i que les metadades quedaran sempre visibles.	Per retirar el contingut contactar amb Mendeley.
8. Cost	Gratuït però hi ha la possibilitat de més espai (FAQS).	Gratuït per investigadors i científics europeus.	Gratuït (fins 20 GB). Per més espai i altres prestacions, ofereixen versió de pagament..	- 120$ per un conjunt de dades. - Tarifes i càrrecs addicionals.	Gratuït.	Gratuït.

9. Dades enllaçades a les publicacions	Permet afegir enllaços a documents relacionats.	Permet afegir enllaços a documents relacionats.	Permet afegir enllaços al text complet o a documents relacionats.	- Les dades han d'estar associades amb un document publicat o en procés de publicació o dades associades a documents no revisats per parells de fonts acadèmiques. - Integració amb les publicacions.	Permet afegir enllaços a documents relacionats.	Permet afegir enllaços a documents i programaris relacionats a les publicacions associades.
10. Preservació	Dades emmagatzemades durant mínim 20 anys al CERN Data Center	Com a mínim 2 anys i amb garanties de migració a altres repositoris (el centre de dades d'emmagatzematge està certificat per la ISO).	Les dades públiques es troben emmagatzemades a Amazon S3 i DuraCloud / Chronopolis.	- Dades emmagatzemades i permanentment disponibles a California Digital Library's Merritt Repository. - Rèplica a DataONE.	- Harvard Dataverse Preservation Policy. - Conveni amb Data Preservation Alliance per preservar les dades a llarg termini - Integració amb Archivematica, DuraCloud/Chronopolis.	Dades emmagatzemades a Amazon S3, a Alemanya. Conveni amb DANS per preservar les dades a llarg termini.
11. Versions dels documents	Permet tenir diferents versions d'un arxiu.	Sense informació.	Permet tenir diferents versions d'un arxiu ja públic.	Permet tenir diferents versions d'un arxiu.	Permet tenir diferents versions d'un arxiu.	Permet tenir diferents versions d'un arxiu.
12. Compleix els requisits dels ajuts europeus	Recomanat per projectes europeus. Desenvolupat pel CERN (projecte OpenAIRE 7PM de la UE).	Recomanat pels projectes europeus. Projecte finançat per H2020 amb participació del BSC, DANS, LIBER i la Univ. d'Edimburg entre d'altres.	Sense informació.	Acord de col·laboració amb Zenodo (signat el juliol de 2019).	- Es poden fer exportacions compatibles amb OpenAIRE (a partir de la versió 4.14) - Compleix amb els principis FAIR.	Sense informació.
13. Més informació (URL)	- FAQS - Policies - Features - About	- FAQS - DARUP	- FAQS - Features - Blog - Tools	- FAQS - Terms of service - Features and technology - Blog	- Dataverse Guides - Demo (proves Harvard) - About (general) - Blog	- FAQS - Terms of use
14. Altres	- Integració amb ORCID, GitHub - Espais col·laboratius - Permet exportar cites a gestors bibliogràfics i xarxes socials.	- B2FIND permet cercar dades d'aquest i d'altres repositoris - Es pot utilitzar de manera independent B2DROP com una infrastructura per actualitzar i intercanviar dades.	- Integració amb ORCID - Espais col·laboratius - Estadístiques de descàrregues -Permet afegir al perfil personal enllaços a xarxes socials i a les nostres publicacions - Permet exportar cites a gestors bibliogràfics i xarxes socials.	- Integració amb ORCID - Estadístiques de descàrregues - Usa Make Data Count - Permet tenir la citació en format RIS i BibTex - Permet exportar cites a Mendeley - Permet compartir enllaços a continguts web a través de Reddit.	- Integració amb ORCID, Dropbox, Rspace, OSF, DataCite, OJS - Aplicació de codi obert (qualsevol pot crear el seu propi Dataverse) - Usa Make Data Count per recopilar i mostrar mètriques - Permet tenir la citació en format XML o RIS - Permet compartir un conjunt de dades a les xarxes socials.	- Integració amb ORCID, Dropbox, Drive Certificació "Data Seal of Approval" de la indústria dels repositoris de dades. - Ofereix una API oberta - Estadístiques de visualitzacions i descàrregues per dataset.

Readme.txt

Fichero en que se describe la información necesaria para que los conjuntos de datos sean comprensibles y reutilizables: autoría, título, descripción, metodología, proyectos financiadores, cobertura temporal y geográfica, derechos de uso y privacidad, etc.

https://edatos.consorciomadrono.es/readme.xhtml

Cornell University

Home About Services Data Management Planning Best Practices

RESEARCH DATA MANAGEMENT SERVICE GROUP

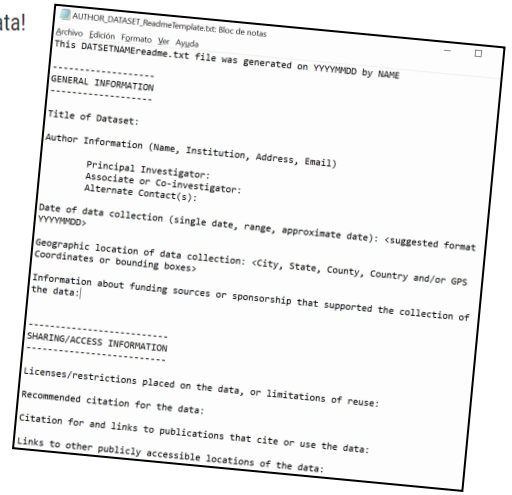
Comprehensive Data Management Planning & Services

Guide to writing "readme" style metadata

A readme file provides information about a data file and is intended to help ensure that the data can be correctly interpreted, by yourself at a later date or by others when sharing or publishing data. Standards-based metadata is generally preferable, but where no appropriate standard exists, for internal use, writing "readme" style metadata is an appropriate strategy.

Want a template? **Download one** and adapt it for your own data!

- Best practices
- Recommended content
  - General information
  - Data and file overview
  - Sharing and access information
  - Methodological information
  - Data-specific information
- References
- Related information


AUTHOR_DATASET_ReadmeTemplate.txt: Bloc de notas

Archivo Edición Formato Ver Ayuda

```
This DATSETNAMEreadme.txt file was generated on YYYYMMDD by NAME

-------------------
GENERAL INFORMATION
-------------------

Title of Dataset:

Author Information (Name, Institution, Address, Email)

        Principal Investigator:
        Associate or Co-investigator:
        Alternate Contact(s):

Date of data collection (single date, range, approximate date): <suggested format
YYYYMMDD>

Geographic location of data collection: <City, State, County, Country and/or GPS
Coordinates or bounding boxes>

Information about funding sources or sponsorship that supported the collection of
the data:

------------------------
SHARING/ACCESS INFORMATION
------------------------

Licenses/restrictions placed on the data, or limitations of reuse:

Recommended citation for the data:

Citation for and links to publications that cite or use the data:

Links to other publicly accessible locations of the data:
```

https://data.research.cornell.edu/content/readme#fileoverview

AUTHOR_DATASET_ReadmeTemplate.txt - Editor

Datei Bearbeiten Format Ansicht Hilfe

```
This DATSETNAMEreadme.txt file was generated on YYYY-MM-DD by NAME
<help text is included in angle brackets, and can be deleted before saving>


GENERAL INFORMATION

1. Title of Dataset:

2. Author Information
        A. Principal Investigator Contact Information
                Name:
                Institution:
                Address:
                Email:

        B. Associate or Co-investigator Contact Information
                Name:
                Institution:
                Address:
                Email:

        C. Alternate Contact Information
                Name:
                Institution:
                Address:
                Email:

3. Date of data collection (single date, range, approximate date) <suggested format YYYY-MM-
DD>:

4. Geographic location of data collection <latitude, longiute, or city/region, State,
Country, as appropriate>:

5. Information about funding sources that supported the collection of the data:


SHARING/ACCESS INFORMATION

1. Licenses/restrictions placed on the data:

2. Links to publications that cite or use the data:

3. Links to other publicly accessible locations of the data:

4. Links/relationships to ancillary data sets:

5. Was data derived from another source? yes/no
        A. If yes, list source(s):

6. Recommended citation for this dataset:
```

Zeile 71, Spalte 37 90% Windows (CRLF) UTF-8 mit BOM

AUTHOR_DATASET_ReadmeTemplate.txt - Editor

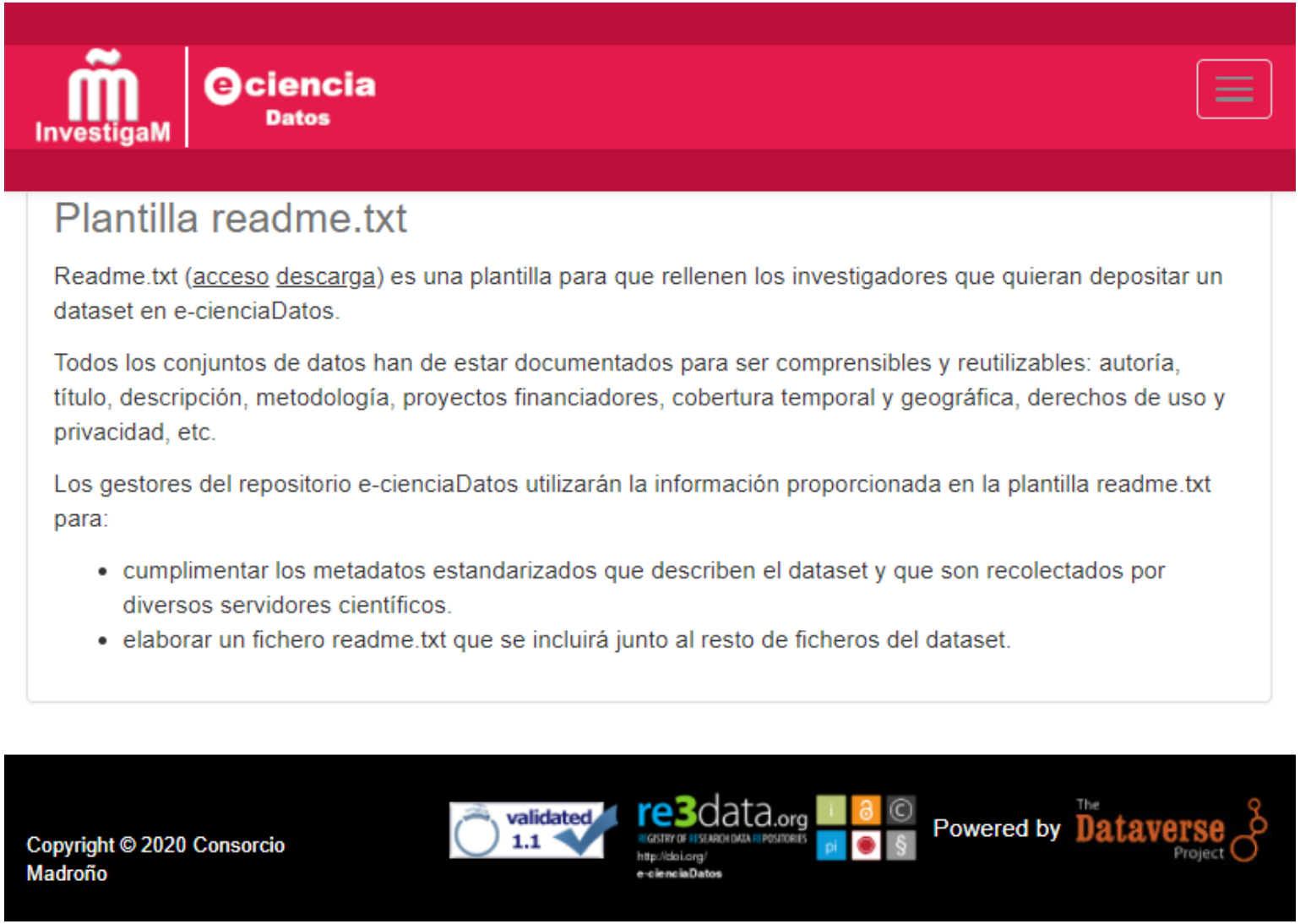
Datei Bearbeiten Format Ansicht Hilfe

```
DATA & FILE OVERVIEW

1. File List:
<list all files (or folders, as appropriate for dataset organization) contained in the dataset,
with a brief description>

2. Relationship between files, if important:

3. Additional related data collected that was not included in the current data package:

4. Are there multiple versions of the dataset? yes/no
        A. If yes, name of file(s) that was updated:
                i. Why was the file updated?
                ii. When was the file updated?


METHODOLOGICAL INFORMATION

1. Description of methods used for collection/generation of data:
<Include links or references to publications or other documentation containing experimental
design or protocols used in data collection>

2. Methods for processing the data:
<describe how the submitted data were generated from the raw or collected data>

3. Instrument- or software-specific information needed to interpret the data:
<include full name and version of software, and any necessary packages or libraries needed to
run scripts>

4. Standards and calibration information, if appropriate:

5. Environmental/experimental conditions:

6. Describe any quality-assurance procedures performed on the data:

7. People involved with sample collection, processing, analysis and/or submission:


DATA-SPECIFIC INFORMATION FOR: [FILENAME]
<repeat this section for each dataset, folder or file, as appropriate>

1. Number of variables:

2. Number of cases/rows:

3. Variable List:
<list variable name(s), description(s), unit(s)and value labels as appropriate for each>

4. Missing data codes:
<list code/symbol and definition>n

5. Specialized formats or other abbreviations used:
```

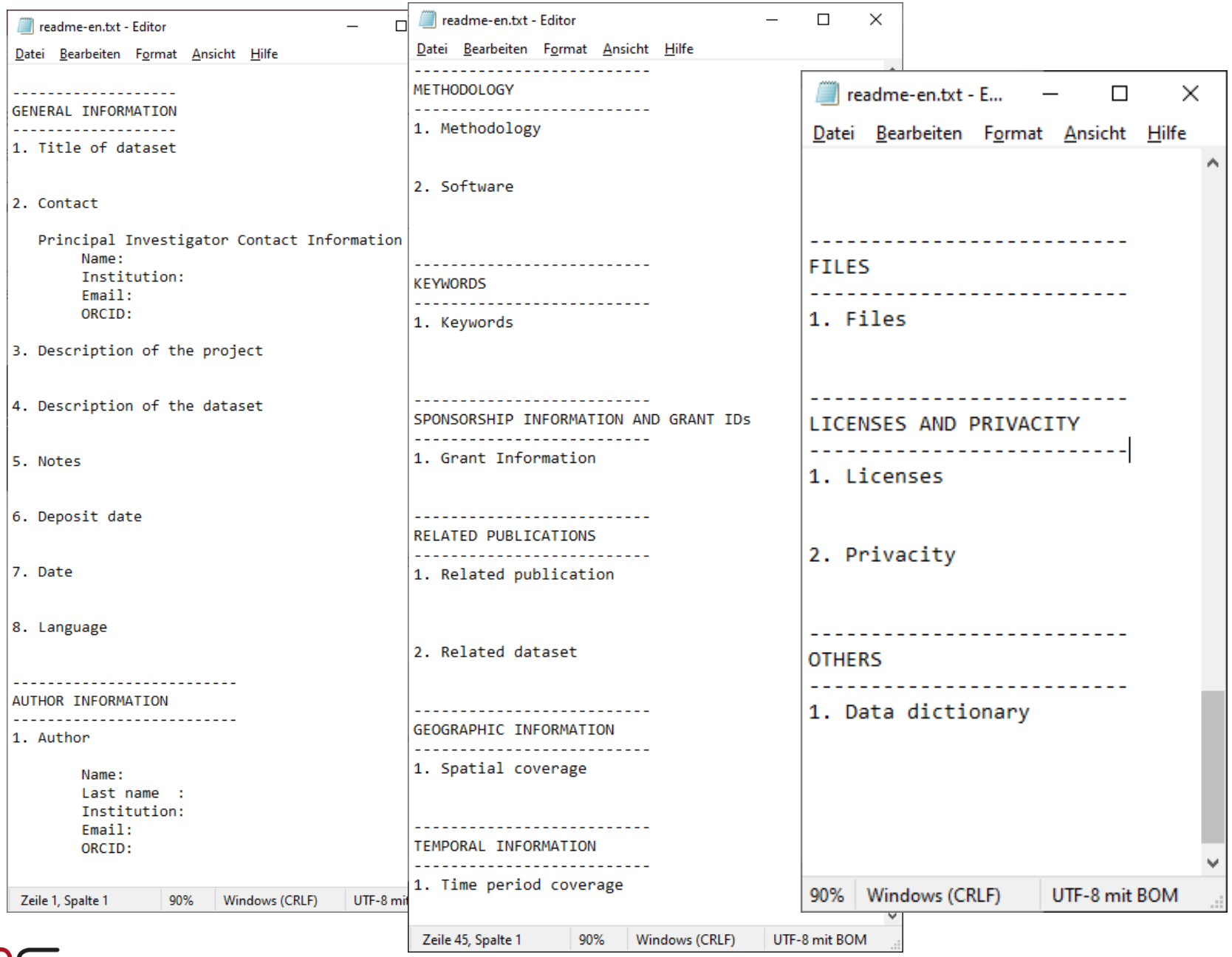
Zeile 29, Spalte 1 90% Windows (CRLF) UTF-8 mit BOM

AUTHOR_DATASET_ReadmeTemplate.txt

https://cornell.app.box.com/v/ReadmeTemplate



https://edatos.consorciomadrono.es/readme.xhtml



BIBLIO MAD SALUD

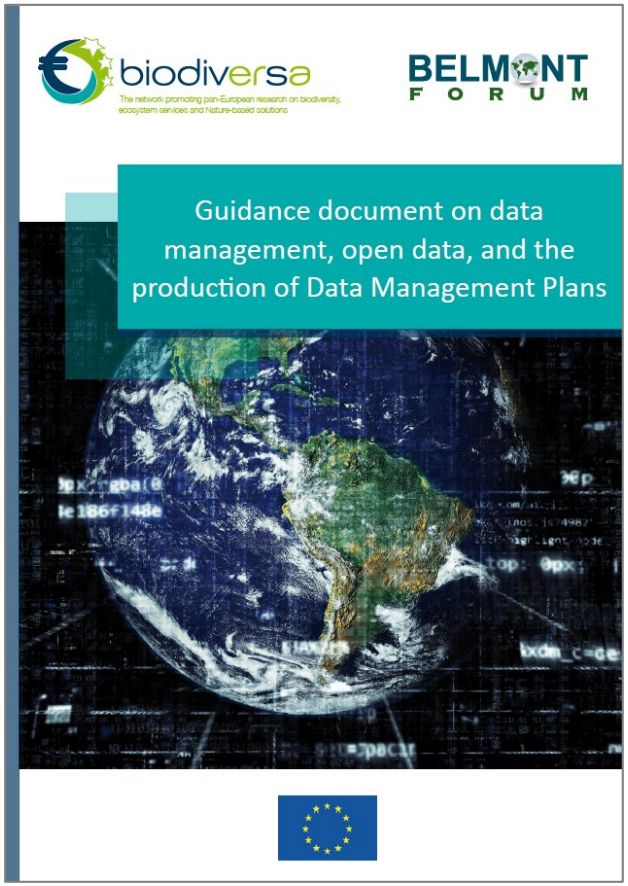
https://edatos.consorciomadrono.es/resources/txt/readme-en.txt

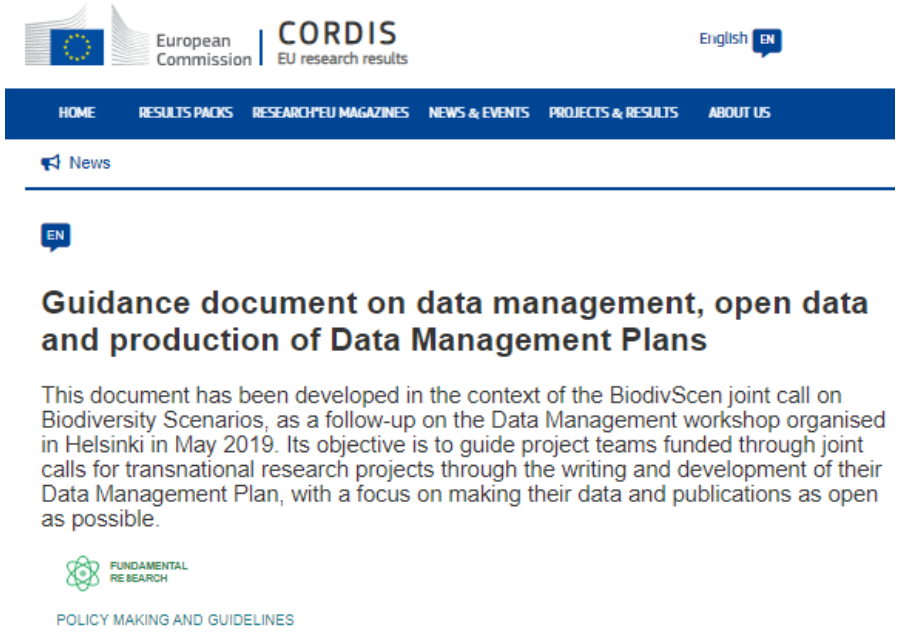
Recomendaciones para compartir datos

- Acompañar los datos de un Readme.txt file
- Publicar los datos en formatos abiertos, para que puedan ser reproducidos sin trabas
- Publicar los datos con una declaración explícita de los deseos del investigador para con los datos.
- Utilizar licencias apropiadas y que reconozcan los derechos de autor.
- Debería definirse de acuerdo a la definición de OKF. SI es posible, los datos no deben utilizar cláusulas restrictivas.
- Ante la fuerte recomendación de publicación de datos producidos con fondos públicos, se recomienda que se hagan públicos siguiendo los *Science Commons* *Protocol for Implementing Open Access Data* *and the* *Open Knowledge/Data Definition*.

Peter Murray-Rust, Cameron Neylon, Rufus Pollock y John Wilbanks. *Panton Principles. Principles for Open Data in Science*. http://pantonprinciples.org

PLANTILLAS Y HERRAMIENTAS PARA CREAR PGDs





Goudeseune Lise, Le Roux Xavier, Eggermont Hilde, Bishop Bradley, Bléry Claire, Brosens Dimitri, … Ungvári Judit. (2019, September 19). Guidance document for scientists on data management, open data, and the production of Data Management Plans. BiodivERsA report. Zenodo. http://doi.org/10.5281/zenodo.3448251

DCC (2013) Checklist for a Data Management Plan. v.4.0. Edinburgh: Digital Curation Centre. Disponible en: http://www.dcc.ac.uk/resources/data-management-plans

Se trata de una checklist de 4 páginas que enumera las principales preguntas o temas que se han de cubrir al escribir un PGD:

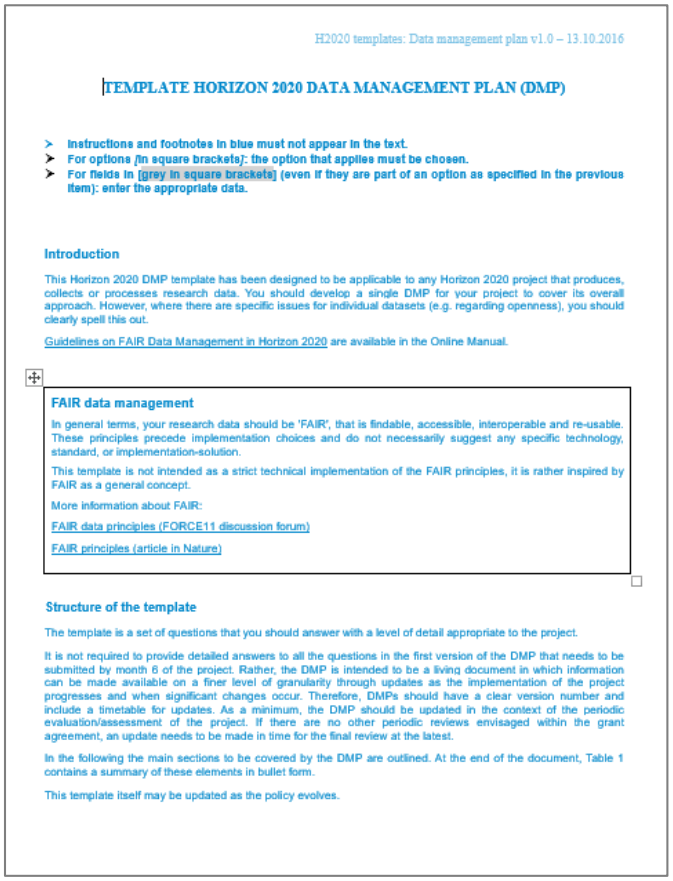
Selection and Preservation	
Which data should be retained, shared, and/or preserved?	**Questions to consider:** - What data must be retained/destroyed for contractual, legal, or regulatory purposes? - How will you decide what other data to keep? - What are the foreseeable research uses for the data? - How long will the data be retained and preserved? - **Guidance:** Consider how the data may be reused e.g. to validate your research findings, conduct new studies, or for teaching. Decide which data to keep and for how long. This could be based on any obligations to retain certain data, the potential reuse value, what is economically viable to keep, and any additional effort required to prepare the data for data sharing and preservation. Remember to consider any additional effort required to prepare the data for sharing and preservation, such as changing file formats.
What is the long-term preservation plan for the dataset?	**Questions to consider:** - Where e.g. in which repository or archive will the data be held? - What costs if any will your selected data repository or archive charge? - Have you costed in time and effort to prepare the data for sharing / preservation? **Guidance:** Consider how datasets that have long-term value will be preserved and curated beyond the lifetime of the grant. Also outline the plans for preparing and documenting data for sharing and archiving. If you do not propose to use an established repository, the data management plan should demonstrate that resources and systems will be in place to enable the data to be curated effectively beyond the lifetime of the grant.

Cruz M, van den Berg E. (2020) *NWO DMP assessment rubric.* Zenodo. http://doi.org/10.5281/zenodo.3629157

Se trata de una tabla que se divide en 'criterios a cumplir', donde cada criterio va explicado y acompañado de 'niveles de cumplimiento' que describen lo que el **NWO (Dutch Research Council)** considera que es un nivel de respuesta suficiente e insuficiente.
Ejemplo de preguntas, descripción y formas de valorarlo:

Performance Criteria		Performance Levels	
DMP Question	**DMP Guidance**	**Sufficiently Addressed**	**Insufficiently Addressed**
Will you re-use existing data for this research? ☐ Yes ☐ No **If yes**: explain which existing data you will re-use and under which terms of use.	NWO encourages the re-use of existing data wherever possible. Explain which existing data you will re-use and state any constraints on re- use of existing data if there are any.	One of the boxes checked. If 'Yes' is selected, it clearly indicates the source(s) of any existing data being re-used and explains the terms of re-use. Otherwise, if 'No' is selected, no explanation required.	None of the boxes checked and no explanation provided. 'Yes' is selected, but it doesn't specify source(s) of any existing data being re-used and/or the terms of re-use.
If new data will be produced: describe the data you expect your research will generate and the format and volumes to be collected or produced.	- Give details on the kind of data: for example numeric (databases, spreadsheets), textual (documents), image, audio, video, and/or mixed media. - Give details on the data format: the way in which the data is encoded for storage, often reflected by the filename extension (for example pdf, xls, doc, txt, or rdf). - Give preference to open and standard formats as they facilitate sharing and long- term re-use of data. Several repositories provide lists of such 'preferred formats' (see e.g. DANS File Formats and 4TU.ResearchData Preferred Formats).	Clearly describes or lists what data types will be generated (e.g. numeric, textual, audio, video, etc.) and their associated data formats. Alternatively, it clearly states that no new data will be produced or generated by the project. NB. Information derived from previously existing data sources (as listed in question 1.1) – namely, output, processed, analysed data – are to be considered new data under this question.	No information provided, without a valid reason to do so (e.g. a statement that no data will be produced or generated); or only lists/describes data types without specifying the data formats; alternatively, it only lists formats, without specifying the data types.

Plantillas



H2020 templates: Data management plan v1.0 – 13.10.2016

TEMPLATE HORIZON 2020 DATA MANAGEMENT PLAN (DMP)

- **Instructions and footnotes in blue must not appear in the text.**
- **For options [in square brackets]: the option that applies must be chosen.**
- **For fields in [grey in square brackets] (even if they are part of an option as specified in the previous item): enter the appropriate data.**

Introduction

This Horizon 2020 DMP template has been designed to be applicable to any Horizon 2020 project that produces, collects or processes research data. You should develop a single DMP for your project to cover its overall approach. However, where there are specific issues for individual datasets (e.g. regarding openness), you should clearly spell this out.

Guidelines on FAIR Data Management in Horizon 2020 are available in the Online Manual.

FAIR data management

In general terms, your research data should be 'FAIR', that is findable, accessible, interoperable and re-usable. These principles precede implementation choices and do not necessarily suggest any specific technology, standard, or implementation-solution.

This template is not intended as a strict technical implementation of the FAIR principles, it is rather inspired by FAIR as a general concept.

More information about FAIR:

FAIR data principles (FORCE11 discussion forum)

FAIR principles (article in Nature)

Structure of the template

The template is a set of questions that you should answer with a level of detail appropriate to the project.

It is not required to provide detailed answers to all the questions in the first version of the DMP that needs to be submitted by month 6 of the project. Rather, the DMP is intended to be a living document in which information can be made available on a finer level of granularity through updates as the implementation of the project progresses and when significant changes occur. Therefore, DMPs should have a clear version number and include a timetable for updates. As a minimum, the DMP should be updated in the context of the periodic evaluation/assessment of the project. If there are no other periodic reviews envisaged within the grant agreement, an update needs to be made in time for the final review at the latest.

In the following the main sections to be covered by the DMP are outlined. At the end of the document, Table 1 contains a summary of these elements in bullet form.

This template itself may be updated as the policy evolves.

H2020 templates: Data management plan v1.0 – 13.10.2016

1. Data Summary

What is the purpose of the data collection/generation and its relation to the objectives of the project?

What types and formats of data will the project generate/collect?

Will you re-use any existing data and how?

What is the origin of the data?

What is the expected size of the data?

To whom might it be useful ('data utility')?

2. FAIR data

2. 1. Making data findable, including provisions for metadata

Are the data produced and/or used in the project discoverable with metadata, identifiable and locatable by means of a standard identification mechanism (e.g. persistent and unique identifiers such as Digital Object Identifiers)?

What naming conventions do you follow?

Will search keywords be provided that optimize possibilities for re-use?

Do you provide clear version numbers?

What metadata will be created? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

2.2. Making data openly accessible

Which data produced and/or used in the project will be made openly available as the default? If certain datasets cannot be shared (or need to be shared under restrictions), explain why, clearly separating legal and contractual reasons from voluntary restrictions.

Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if relevant provisions are made in the consortium agreement and are in line with the reasons for opting out.

How will the data be made accessible (e.g. by deposition in a repository)?

What methods or software tools are needed to access the data?

Is documentation about the software needed to access the data included?

Is it possible to include the relevant software (e.g. in open source code)?

Where will the data and associated metadata, documentation and code be deposited? Preference should be given to certified repositories which support open access where possible.

Have you explored appropriate arrangements with the identified repository?

If there are restrictions on use, how will access be provided?

Is there a need for a data access committee?

Are there well described conditions for access (i.e. a machine readable license)?

How will the identity of the person accessing the data be ascertained?

2.3. Making data interoperable

Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating re-combinations with different datasets from different origins)?

What data and metadata vocabularies, standards or methodologies will you follow to make your data interoperable?

3

Template for the Horizon 2020 Data Management Plan

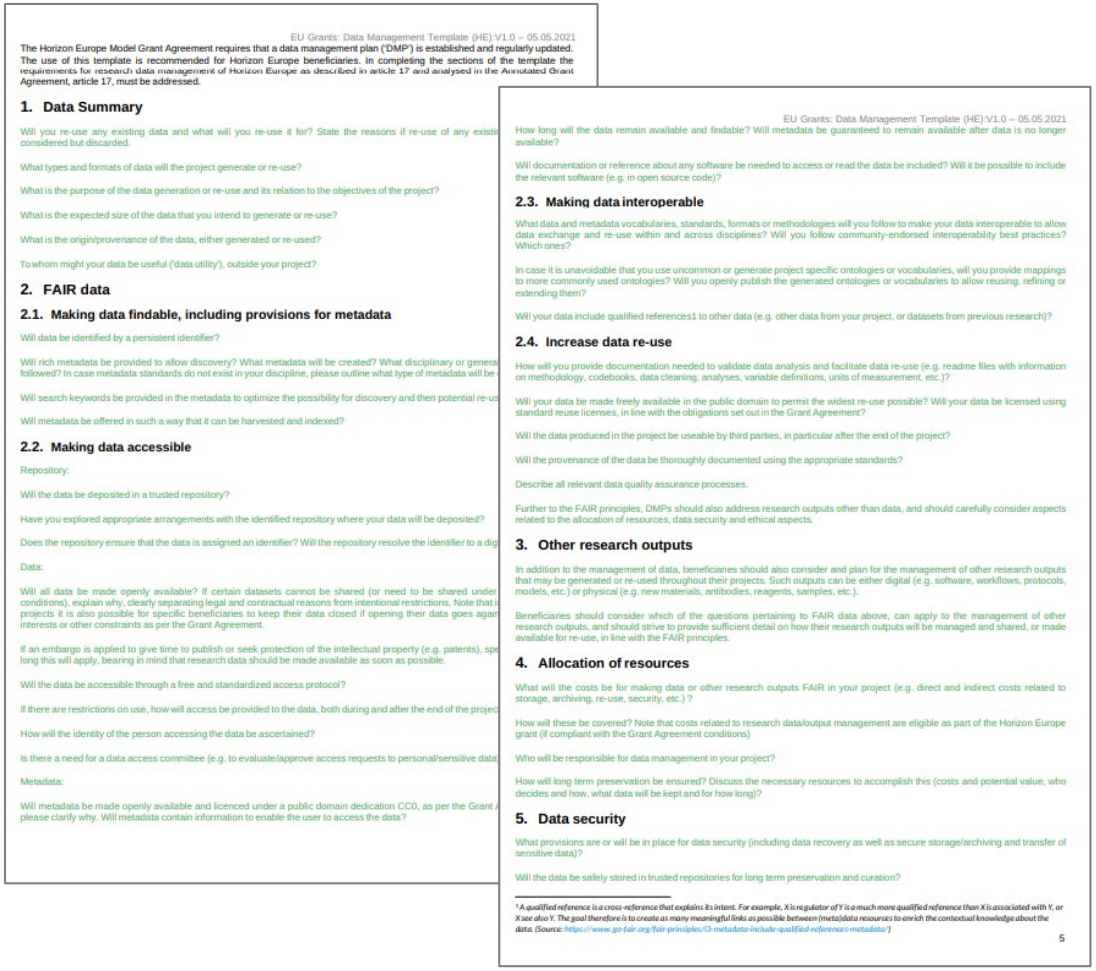
https://ec.europa.eu/research/participants/data/ref/h2020/gm/reporting/h2020-tpl-oa-data-mgt-plan_en.docx

Template for the Horizon Europe Data Management Plan

Horizon Europe

Data Management Plan Template

Version 1.0
05 May 2021



EU Grants: Data Management Template (HE):V1.0 – 05.05.2021

The Horizon Europe Model Grant Agreement requires that a data management plan ('DMP') is established and regularly updated. The use of this template is recommended for Horizon Europe beneficiaries. In completing the sections of the template the requirements for research data management of Horizon Europe as described in article 17 and analysed in the Annotated Grant Agreement, article 17, must be addressed.

1. Data Summary

Will you re-use any existing data and what will you re-use it for? State the reasons if re-use of any existi considered but discarded.

What types and formats of data will the project generate or re-use?

What is the purpose of the data generation or re-use and its relation to the objectives of the project?

What is the expected size of the data that you intend to generate or re-use?

What is the origin/provenance of the data, either generated or re-used?

To whom might your data be useful ('data utility'), outside your project?

2. FAIR data

2.1. Making data findable, including provisions for metadata

Will data be identified by a persistent identifier?

Will rich metadata be provided to allow discovery? What metadata will be created? What disciplinary or genera followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be

Will search keywords be provided in the metadata to optimize the possibility for discovery and then potential re-us

Will metadata be offered in such a way that it can be harvested and indexed?

2.2. Making data accessible

Repository:

Will the data be deposited in a trusted repository?

Have you explored appropriate arrangements with the identified repository where your data will be deposited?

Does the repository ensure that the data is assigned an identifier? Will the repository resolve the identifier to a dig

Data:

Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under conditions), explain why, clearly separating legal and contractual reasons from intentional restrictions. Note that projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes again interests or other constraints as per the Grant Agreement.

If an embargo is applied to give time to publish or seek protection of the intellectual property (e.g. patents), spe long this will apply, bearing in mind that research data should be made available as soon as possible.

Will the data be accessible through a free and standardized access protocol?

If there are restrictions on use, how will access be provided to the data, both during and after the end of the projec

How will the identity of the person accessing the data be ascertained?

Is there a need for a data access committee (e.g. to evaluate/approve access requests to personal/sensitive data

Metadata:

Will metadata be made openly available and licenced under a public domain dedication CC0, as per the Grant please clarify why. Will metadata contain information to enable the user to access the data?

EU Grants: Data Management Template (HE):V1.0 – 05.05.2021

How long will the data remain available and findable? Will metadata be guaranteed to remain available after data is no longer available?

Will documentation or reference about any software be needed to access or read the data be included? Will it be possible to include the relevant software (e.g. in open source code)?

2.3. Making data interoperable

What data and metadata vocabularies, standards, formats or methodologies will you follow to make your data interoperable to allow data exchange and re-use within and across disciplines? Will you follow community-endorsed interoperability best practices? Which ones?

In case it is unavoidable that you use uncommon or generate project specific ontologies or vocabularies, will you provide mappings to more commonly used ontologies? Will you openly publish the generated ontologies or vocabularies to allow reusing, refining or extending them?

Will your data include qualified references[1] to other data (e.g. other data from your project, or datasets from previous research)?

2.4. Increase data re-use

How will you provide documentation needed to validate data analysis and facilitate data re-use (e.g. readme files with information on methodology, codebooks, data cleaning, analyses, variable definitions, units of measurement, etc.)?

Will your data be made freely available in the public domain to permit the widest re-use possible? Will your data be licensed using standard reuse licenses, in line with the obligations set out in the Grant Agreement?

Will the data produced in the project be useable by third parties, in particular after the end of the project?

Will the provenance of the data be thoroughly documented using the appropriate standards?

Describe all relevant data quality assurance processes.

Further to the FAIR principles, DMPs should also address research outputs other than data, and should carefully consider aspects related to the allocation of resources, data security and ethical aspects.

3. Other research outputs

In addition to the management of data, beneficiaries should also consider and plan for the management of other research outputs that may be generated or re-used throughout their projects. Such outputs can be either digital (e.g. software, workflows, protocols, models, etc.) or physical (e.g. new materials, antibodies, reagents, samples, etc.).

Beneficiaries should consider which of the questions pertaining to FAIR data above, can apply to the management of other research outputs, and should strive to provide sufficient detail on how their research outputs will be managed and shared, or made available for re-use, in line with the FAIR principles.

4. Allocation of resources

What will the costs be for making data or other research outputs FAIR in your project (e.g. direct and indirect costs related to storage, archiving, re-use, security, etc.) ?

How will these be covered? Note that costs related to research data/output management are eligible as part of the Horizon Europe grant (if compliant with the Grant Agreement conditions)

Who will be responsible for data management in your project?

How will long term preservation be ensured? Discuss the necessary resources to accomplish this (costs and potential value, who decides and how, what data will be kept and for how long)?

5. Data security

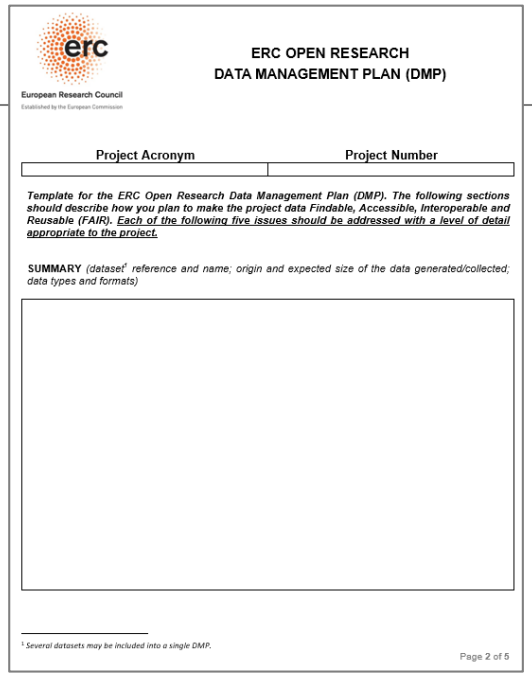
What provisions are or will be in place for data security (including data recovery as well as secure storage/archiving and transfer of sensitive data)?

Will the data be safely stored in trusted repositories for long term preservation and curation?

[1] *A qualified reference is a cross-reference that explains its intent. For example, X is regulator of Y is a much more qualified reference than X is associated with Y, or X see also Y. The goal therefore is to create as many meaningful links as possible between (meta)data resources to enrich the contextual knowledge about the data. (Source: https://www.go-fair.org/fair-principles/i3-metadata-include-qualified-references-metadata/)*

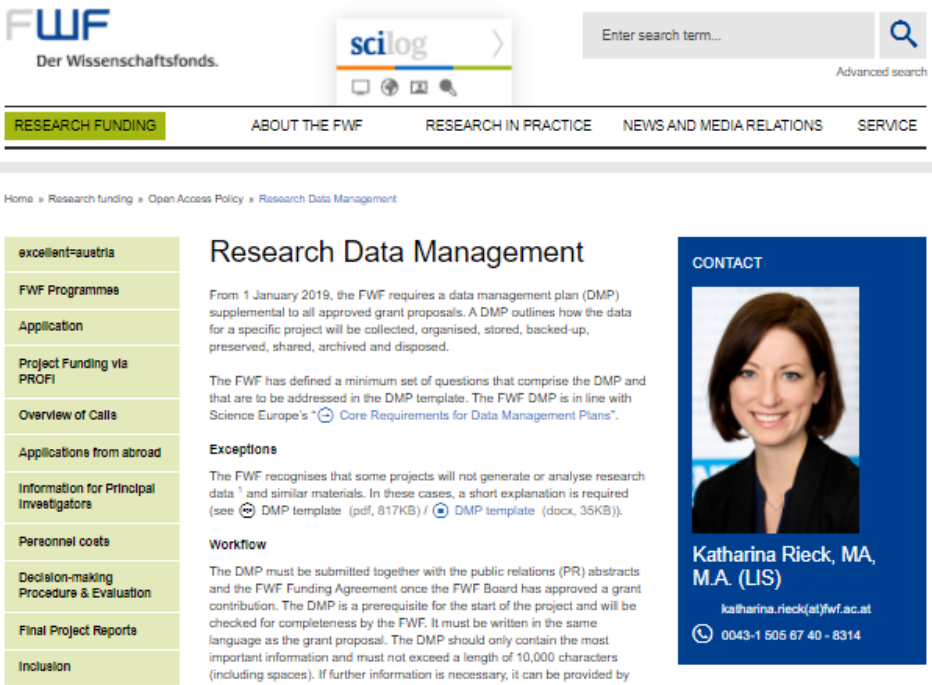
https://enspire.science/wp-content/uploads/2021/09/Horizon-Europe-Data-Management-Plan-Template.pdf

Template for the ERC Data Management Plan



1. MAKING DATA FINDABLE *(dataset description: metadata, persistent and unique identifiers e.g., DOI).*
2. MAKING DATA OPENLY ACCESSIBLE *(which data will be made openly available and if some datasets remain closed, the reasons for not giving access; where the data and associated metadata, documentation and code are deposited (repository?); how the data can be accessed (are relevant software tools/methods provided?).*
3. MAKING DATA INTEROPERABLE (which standard or field-specific data and metadata vocabularies and methods will be used).
4. INCREASE DATA RE-USE *(what data will remain re-usable and for how long, is embargo foreseen; how the data is licensed; data quality assurance procedures).*
5. ALLOCATION OF RESOURCES and DATA SECURITY *(estimated costs for making the project data open access and potential value of long-term data preservation; procedures for data backup and recovery; transfer of sensitive data and secure storage in repositories for long term preservation and curation).*

https://erc.europa.eu/content/erc-data-management-plan-template-0


FWF
Der Wissenschaftsfonds.

scilog

Enter search term...

Advanced search

RESEARCH FUNDING | ABOUT THE FWF | RESEARCH IN PRACTICE | NEWS AND MEDIA RELATIONS | SERVICE

Home » Research funding » Open Access Policy » Research Data Management

- excellent=austria
- FWF Programmes
- Application
- Project Funding via PROFI
- Overview of Calls
- Applications from abroad
- Information for Principal Investigators
- Personnel costs
- Decision-making Procedure & Evaluation
- Final Project Reports
- Inclusion
- Research Integrity & Research Ethics
- Coaching Workshops & Information Events
- Open Access Policy
  - » Open Access to Peer-reviewed Publications
  - » Open Access to Research Data
  - » Research Data Management
- FAQ

Research Data Management

From 1 January 2019, the FWF requires a data management plan (DMP) supplemental to all approved grant proposals. A DMP outlines how the data for a specific project will be collected, organised, stored, backed-up, preserved, shared, archived and disposed.

The FWF has defined a minimum set of questions that comprise the DMP and that are to be addressed in the DMP template. The FWF DMP is in line with Science Europe's "Core Requirements for Data Management Plans".

Exceptions

The FWF recognises that some projects will not generate or analyse research data [1] and similar materials. In these cases, a short explanation is required (see DMP template (pdf, 817KB) / DMP template (docx, 35KB)).

Workflow

The DMP must be submitted together with the public relations (PR) abstracts and the FWF Funding Agreement once the FWF Board has approved a grant contribution. The DMP is a prerequisite for the start of the project and will be checked for completeness by the FWF. It must be written in the same language as the grant proposal. The DMP should only contain the most important information and must not exceed a length of 10,000 characters (including spaces). If further information is necessary, it can be provided by links to external sources.

The DMP is to be viewed as a living document that can be modified throughout the project. Any changes made to the DMP should be documented, and its final version must be included in the final grant report.

Financing

Within grant proposals, costs for the preparation, archiving, open access and later use of research data in repositories can be requested. For further information, see the application guidelines for grant proposals.

DMP Template

The Data Management template can be downloaded here:
https://www.fwf.ac.at/fileadmin/files/Dokumente/Open_Access/FWF_DMPTemplate_e.pdf
https://www.fwf.ac.at/fileadmin/files/Dokumente/Open_Access/FWF_DMPTemplate_e.docx

Sample DMPs

Digital Curation Centre (DCC): http://www.dcc.ac.uk/resources/data-management-plans/guidance-examples

FWF-Kontakt

» Katharina Rieck

CONTACT

Katharina Rieck, MA, M.A. (LIS)

katharina.rieck(at)fwf.ac.at

0043-1 505 67 40 - 8314

FWF Data Management Plan Template (DMP) - Guide

	Data Officer	*Who is responsible for the data management and the DMP of the project (name/email address)?*
I	Data Characteristics	
I.1	Description of the data	*What kinds of data/source code will be generated or reused (type, format, volume)?* *How will the research data be generated and which methods will be used?* *How will you structure the data and handle versioning?* *Who is the target audience?*
II	Documentation and Metadata	
II.1	Metadata standards	*What metadata standards (if any) will be in use and why? (see Digital Curation Centre)*
II.2	Documentation of data	*What information is needed for the data to be findable, accessible, interoperable and re-usable (FAIR) in the future?* *Is the data machine-readable?* *How are you planning to document this information?*
II.3	Data quality control	*What quality assurance processes will you adopt?* *How will the consistency and quality of data collection be controlled and documented? (This may include processes such as repeat samples or measurements, standardised data capture, peer review of data or representation with controlled vocabularies.)*
III	Data Availability and Storage	
III.1	Data sharing strategy	*How and when will the data be shared and made accessible?* *What repository will you be using?* *What persistent identifier will be used?*
III.2	Data storage strategy	*What data are to be preserved for the long-term, and what data will not be stored? How and where will the data be stored and backed up during the research?* *How and where will the data be stored after the project ends?* *For how long will the data be stored?* *Are there any costs that need to be covered for storage?* *At what point during or after the project will the data be stored?* *Are there any technical barriers to making the research data fully or partially accessible?*

FWF Data Management Plan Template (DMP)

	Data Officer	
I	Data Characteristics	
I.1	Description of the data	
II	Documentation and Metadata	
II.1	Metadata standards	

https://www.fwf.ac.at/en/research-funding/open-access-policy/research-data-management

Blog News Events Contact Subscribe

Software Sustainability Institute

About Programmes and Events Resources search

- Get up to speed!
- Online sustainability evaluation
- Open Evidence Bank
- REF 2021: Guidance for Software Outputs
- Guides
- Top tips
- Publications
- Case studies
- T-shirts
- Videos

Software Management Plans

Why write a Software Management Plan?

Research software can take many guises. It can be a 50 line bash shell script for manipulating and filtering files, running a bioinformatics analysis, 10,000 lines of Java for medical image analysis or 100,000 lines of Fortran fo be written in scripting languages such as Unix shell, Python, R or MATLAB or in "traditional" programming lang But, whatever guise it takes, research software is an integral part of the modern research ecosystem.

When developing research software, it is easy to focus only on goals and activities such as collaborating with o attending conferences and applying for funding. Together, the demands of daily research practice can all consp development of research software.

A Software Management Plan (SMP) can help you to define a set of structures and goals to understand your re going to develop; who the software is for (even if it is just for yourself); how you will deliver your software to its intended users; how it will help them; and how you will assess whether it has helped them, and contributed to research, in the ways that you intended. An SMP also helps you to understand how you can support those who wish to, or do, use your research software; how your software relates to other artefacts in your research and how you will ensure that your software remains available beyond the lifetime of your current project.

Though an SMP can be of most benefit when starting a project to develop research software, there are benefits to adopting one on a proj already underway. An SMP provides a way to draw together and summarise research software-related aspects that have already been de doing so can reveal additional aspects or options that weren't considered, or weren't applicable, when the project began.

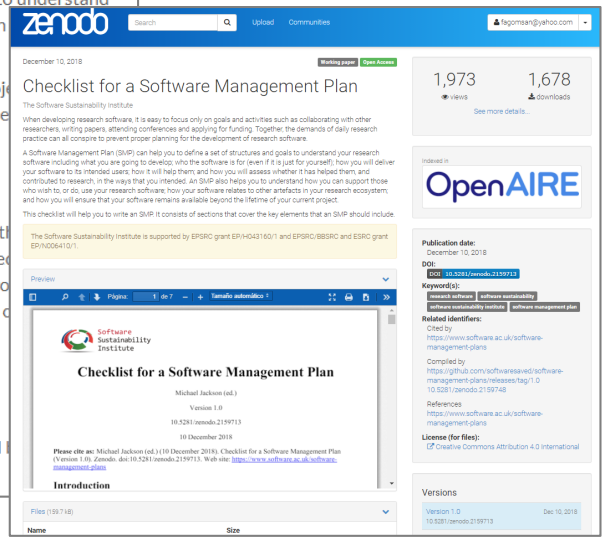
Write a Software Management Plan

We have written a checklist to help you write an SMP. It consists of sections that cover the key elements that an SMP should include. Wit section questions are posed to help you to complete that section. Complementary guidance and links to other resources are also provide questions are relevant to all projects and the extent to which a specific question can be answered may depend both upon the nature of yo software and its current state of development. Consider your SMP as a living document, to be reviewed and revised as the development research software progresses.

The checklist, and templates in Markdown, Microsoft Word, and OpenOffice/Libre office formats, is available within Zenodo:

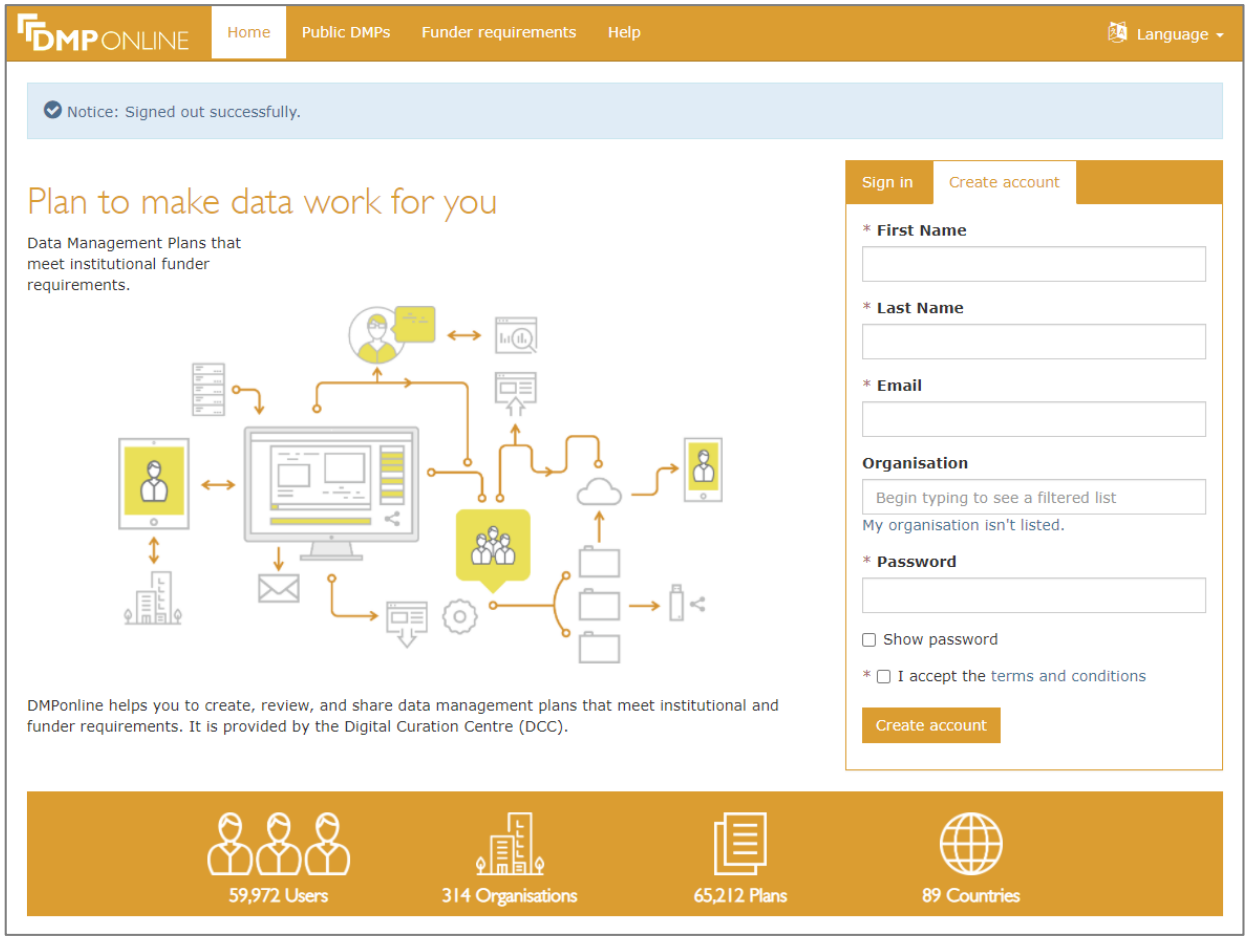
The Software Sustainability Institute. (2018). **Checklist for a Software Management Plan. v1.0**. doi: DOI 10.5281/zenodo.2159713

The Markdown template can be shared within a source code repository or pasted into a Markdown-compliant issue tracker (as provided GitLab and BitBucket, for example).



https://www.software.ac.uk/software-management-plans

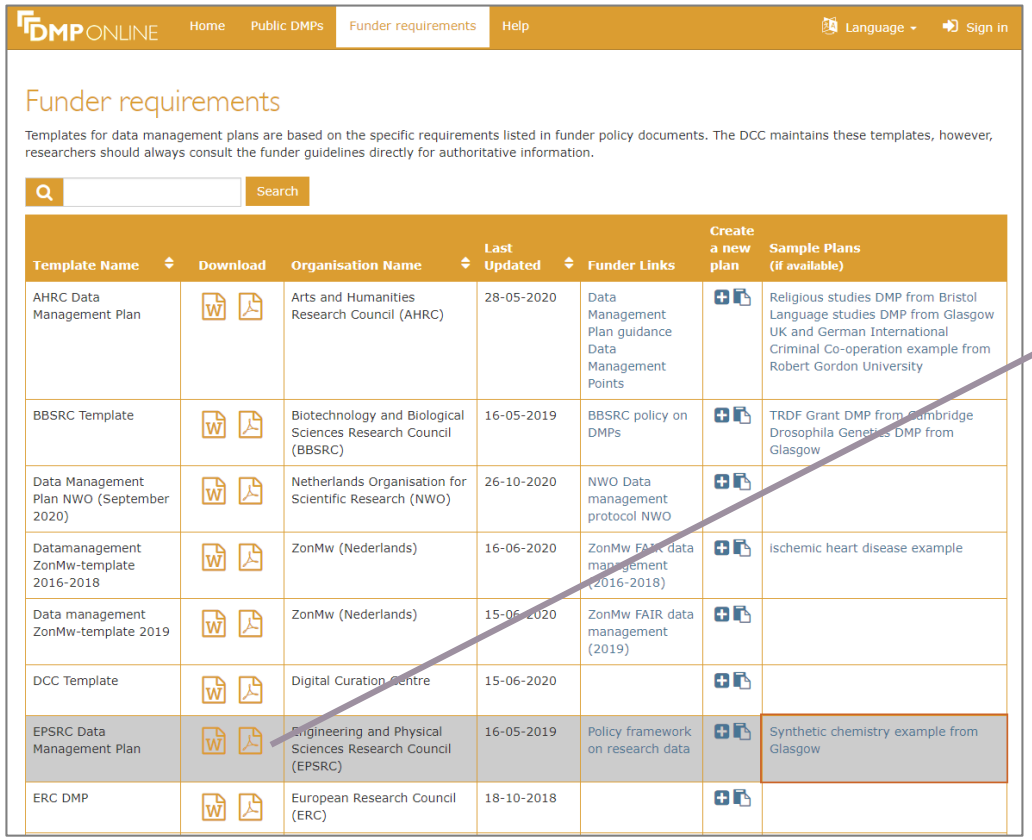
DMP online

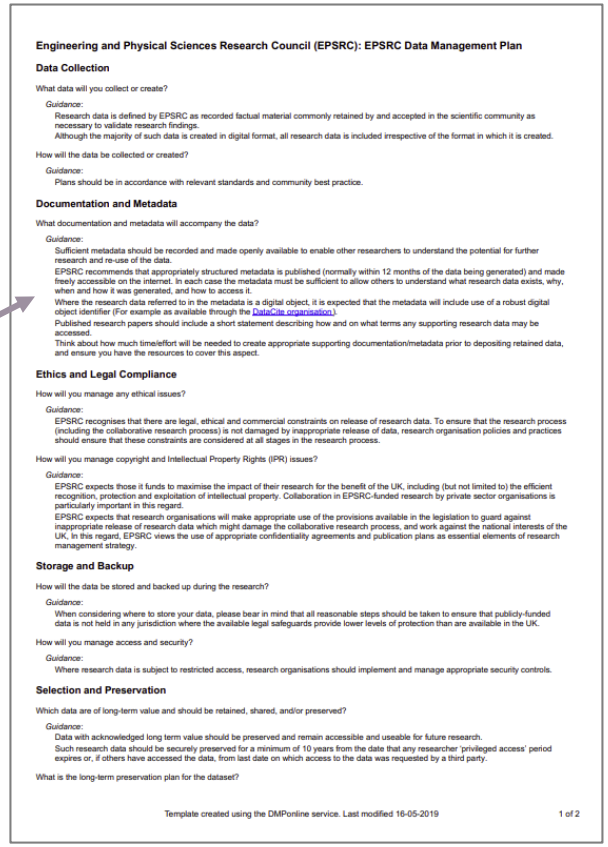


https://dmponline.dcc.ac.uk/

- El DMP Online fue desarrollado de manera conjunta por el Digital Curation Centre (DCC) y el Curation Center de la Universidad de California (UC3).
- Es una herramienta en línea para crear, mantener y exportar DMPs, y se desarrolló para ayudar a los equipos de investigación a cumplir con los requisitos de los financiadores y responder a la recomendación u obligación de que los proyectos de investigación financiados presenten un plan de gestión de datos estructurado como parte integral del proyecto.
- Se desarrolló a partir de la checklist del DCC para crear PGDs, para agregar el valor añadido de que los planes fueran además legibles e intercambiables por máquinas. Esta característica de los PGDs hace que tanto la información que contienen como sus metadatos (vocabularios y estándares utilizados en la descripción del contenido) sean interoperables y localizables.
- Se lanzó hace 10 años, y la semana del 16 al 20 de noviembre de 2020 se hicieron diferentes actividades para celebrar el aniversario, con webinars, juegos, actividades en Twitter con el hashtag #10yearsDMPonline, así como para considerar su futuro y nuevas colaboraciones.
  DMPonline 10th year anniversary celebration week!
- Además de la herramienta propiamente dicha, la web del DMPonline incluye interesantes recursos y guías para la gestión de los datos de investigación, junto con ejemplos de PGDs, a través de la url https://dmponline.dcc.ac.uk/help#PlanningHelp.

Modelos según financiadores





Engineering and Physical Sciences Research Council (EPSRC): EPSRC Data Management Plan

Data Collection

What data will you collect or create?

Guidance:
Research data is defined by EPSRC as recorded factual material commonly retained by and accepted in the scientific community as necessary to validate research findings.
Although the majority of such data is created in digital format, all research data is included irrespective of the format in which it is created.

How will the data be collected or created?

Guidance:
Plans should be in accordance with relevant standards and community best practice.

Documentation and Metadata

What documentation and metadata will accompany the data?

Guidance:
Sufficient metadata should be recorded and made openly available to enable other researchers to understand the potential for further research and re-use of the data.
EPSRC recommends that appropriately structured metadata is published (normally within 12 months of the data being generated) and made freely accessible on the internet. In each case the metadata must be sufficient to allow others to understand what research data exists, why, when and how it was generated, and how to access it.
Where the research data referred to in the metadata is a digital object, it is expected that the metadata will include use of a robust digital object identifier (For example as available through the DataCite organisation.)
Published research papers should include a short statement describing how and on what terms any supporting research data may be accessed.
Think about how much time/effort will be needed to create appropriate supporting documentation/metadata prior to depositing retained data, and ensure you have the resources to cover this aspect.

Ethics and Legal Compliance

How will you manage any ethical issues?

Guidance:
EPSRC recognises that there are legal, ethical and commercial constraints on release of research data. To ensure that the research process (including the collaborative research process) is not damaged by inappropriate release of data, research organisation policies and practices should ensure that these constraints are considered at all stages in the research process.

How will you manage copyright and Intellectual Property Rights (IPR) issues?

Guidance:
EPSRC expects those it funds to maximise the impact of their research for the benefit of the UK, including (but not limited to) the efficient recognition, protection and exploitation of intellectual property. Collaboration in EPSRC-funded research by private sector organisations is particularly important in this regard.
EPSRC expects that research organisations will make appropriate use of the provisions available in the legislation to guard against inappropriate release of research data which might damage the collaborative research process, and work against the national interests of the UK. In this regard, EPSRC views the use of appropriate confidentiality agreements and publication plans as essential elements of research management strategy.

Storage and Backup

How will the data be stored and backed up during the research?

Guidance:
When considering where to store your data, please bear in mind that all reasonable steps should be taken to ensure that publicly-funded data is not held in any jurisdiction where the available legal safeguards provide lower levels of protection than are available in the UK.

How will you manage access and security?

Guidance:
Where research data is subject to restricted access, research organisations should implement and manage appropriate security controls.

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

Guidance:
Data with acknowledged long term value should be preserved and remain accessible and useable for future research.
Such research data should be securely preserved for a minimum of 10 years from the date that any researcher 'privileged access' period expires or, if others have accessed the data, from last date on which access to the data was requested by a third party.

What is the long-term preservation plan for the dataset?

Template created using the DMPonline service. Last modified 16-05-2019 1 of 2

https://dmponline.dcc.ac.uk/public_templates

Una vez que se han ido completando los distintos apartados, la herramienta da la opción de exportar y descargar el plan en diversas plantillas, según formatos de diferentes organismos financiadores.

Ejemplos de DMPs

DMPONLINE Home | Public DMPs | Funder requirements | Help | Language | Sign in

Public DMPs

Public DMPs are plans created using the DMPonline service and shared publicly by their owners. They are not vetted for quality, completeness, or adherence to funder guidelines.

Search

Project Title	Template	Organisation	Owner	Download
Wellbeing Project During COVID-19	EUR Data Management Plan	Erasmus University Rotterdam	Sophie Sweijen	
Fieldlab Besmettingsrisicoanalyse	TU Delft Data Management Questions	Delft University of Technology	Daniel Brus	
Validation Studies of a Questionnaire for Evaluating Human Interaction with An Artificial Social Agent	TU Delft Data Management Questions	Delft University of Technology	Siska Fitrianie	
DEVELOPMENT OF A PROJECT RISK MANAGEMENT FRAMEWORK A STUDY OF AFRICAN MAJOR ECONOMIES	University of Manchester Generic Template	University of Manchester	BABATUNDE DOSUMU	
Quantum-accelerated algorithmic feature learning	TU Delft Data Management Questions	Delft University of Technology	Aritra Sarkar	
Radiocarbon geochronology of the Southern Brazilian upper margin: Data revision and new information	DCC Template	Other	Michel Michaelovitch de Mahiques	
Long Distance Accessibility By Air Transportation Focus Group Meeting	TU Delft Data Management Questions	Delft University of Technology	Sihyun Yoo	
Uncertainty, Ambivalence and Doubt: 'Indo-Guyanese' futures in the context of oil, flooding, and COVID-19	ESRC Template	London School of Economics and Political Science	Rhys Madden	
Administrar o espaço, fiscalizar o mercado: agentes e conflitos na administração de São Paulo (1828-1841)	DCC Template	Other	Claudia de Andrade de Rezende	
Translating Machine Learning into Finance: The Role of 'Modelling Entrepreneurs' in the OTC Derivatives Markets	UoE Data Management Plan	University of Edinburgh	Taylor Spears	

Uncertainty, Ambivalence and Doubt: 'Indo-Guyanese' futures in the context of oil, flooding, and COVID-19

A Data Management Plan created using DMPonline

Creator: Rhys Madden

Affiliation: London School of Economics and Political Science

Funder: Economic and Social Research Council (ESRC)

Template: ESRC Template

ORCID iD: 0000-0003-1759-6210

Grant number: ES/P000622/1 2301163

Project abstract:

This project will explore how individuals of 'Indo-Guyanese' heritage are thinking about the future of Guyana given three major changes: 1) the discovery and production of offshore oil; 2) worsening flooding caused by climate change and deteriorating infrastructure; 3) the COVID-19 pandemic. More specifically, many scholars have written about 'Indo-Guyanese' history and identity in terms of uncertainty and ambivalence. This includes a focus on histories of indentured labour, contemporary relationships to India, and interaction with other racial/ethnic groups. I aim to investigate if these categories are relevant for how 'Indo-Guyanese' individuals view the future, and how these categories may shape, or be shaped by, the potential changes which Guyana faces. Methodologically, this project will be comprised of initial interviews, conducted over Zoom video conferencing software, and more informal follow-up conversations conducted over social media messaging, email, and further video interviews. Participants will be drawn from within Guyana and the Guyanese diaspora abroad, and will be recruited via outreach on online groups and forums, as well as prior participant recommendations.

Last modified: 08-11-2020

Copyright information:

The above plan creator(s) have agreed that others may use as much of the text of this plan as they would like in their own plans, and customise it as necessary. You do not need to credit the creator(s) as the source of the language used, but using any of the plan's text does not imply that the creator(s) endorse, or have any relationship to, your project or proposal

Difficulties in data sharing and measures to overcome these

There are two kinds of data that would be difficult to share. The first relates to things said in confidence by participants. There are topics relating to this research project that may generate answers which could cause harm to participants if openly shared. An example would be reasons given by individuals against returning to Guyana, where this may include knowledge of corruption, drug smuggling, or other illegal activity.
This kind of confidential information may still be curated and ethically shared in cases where anonymization is possible. Anonymization strategies include utilizing pseudonyms for people and places. Where individuals are still likely to be identifiable, even given attempted anonymization (through uniqueness of narratives, for example), then data will only be archived with strict access restrictions allowing vetting of third-party access. Consent for data sharing will be gained through informed consent forms that the researcher will talk through with participants who agree to have interviews recorded.
The second kind of data that would be difficult to share is researcher fieldnotes in response to ongoing informal conversations with participants. This is because fieldnotes are work-in-progress documents involving general observation and partial interpretation, which could be easily misconstrued. No fieldnotes will be made accessible in a data archive. However, meta-data on the extent and general topics of fieldnotes will be shared.

Consent, anonymisation and strategies to enable further re-use of data

Consent for data sharing will be gained through informed consent forms. In line with common anthropological practice, and responding to the risks of over-professionalization for building deep and ongoing relationships with participants, these forms will not be presented at initial interviews. Instead, individuals who verbally consent to having interviews recorded will be contacted at a later date, before the close of the project, to ask for explicit consent to store and share interview transcripts. This will likely take the form of a digitized eConsent form that carefully explains the ways in which data will be handled, as well as the mechanisms by which they can retract their consent in the future. Transcripts from individuals who do not consent to have their data stored or shared will be destroyed.
Anonymization of participants will be adaptable to local frameworks of meaning, responsibility and harm. A flexible approach seems most appropriate as some may want to remain anonymous, particularly if there is a risk of statements or actions reflecting back on them in the future. However, others may expect me to name them as a recognition of their contribution, for example. What is locally expected will become clearer throughout the research process. If I judge there to be a risk to participants that they themselves are not aware of, I will act in accordance with my ethical obligation to anonymize individuals. Strategies may include the use of pseudonyms, collapsing multiple stories together (and clearly explaining where this has been done), or erasing narratives entirely.

Copyright and intellectual property ownership

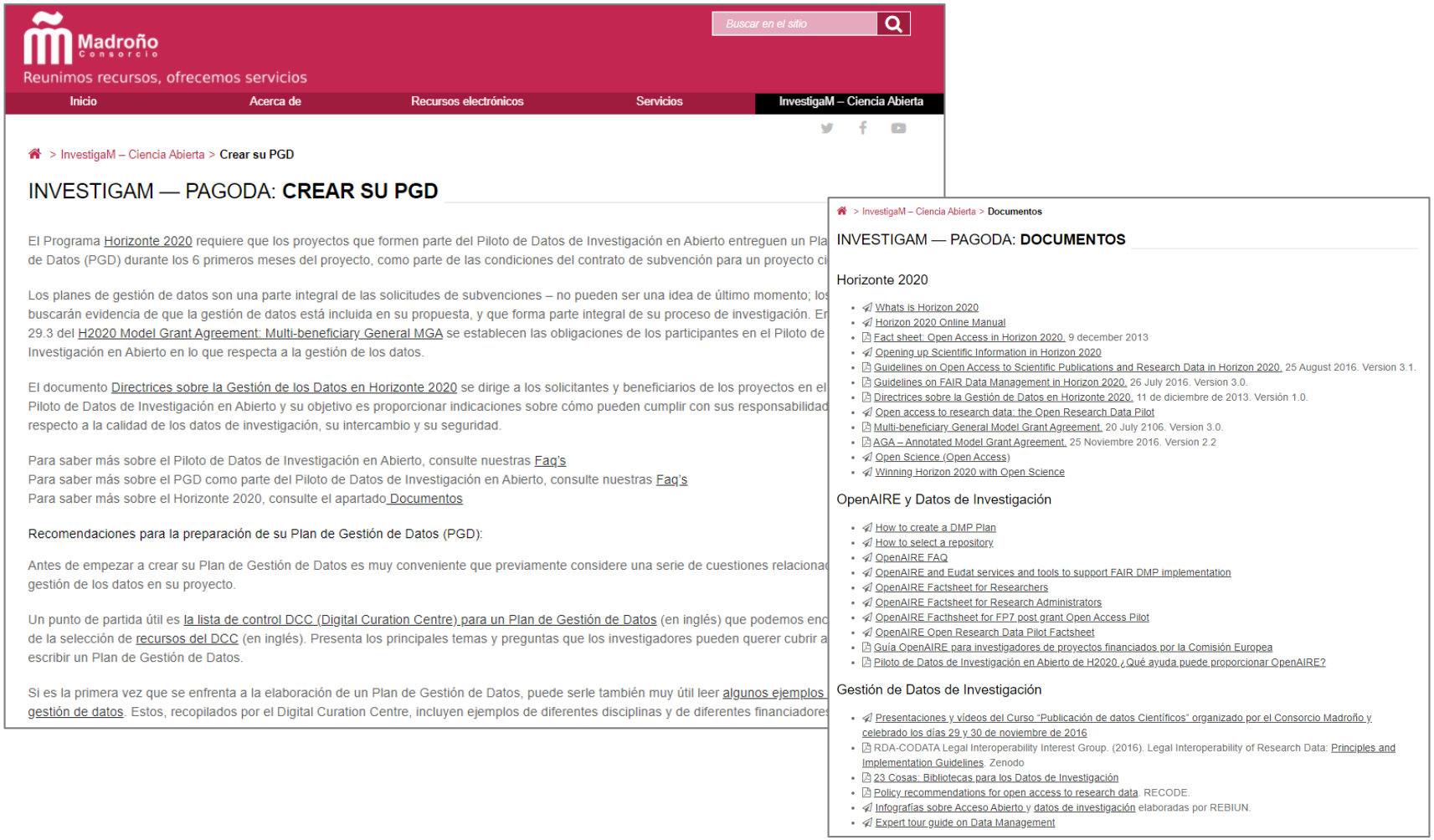
The copyright of recordings and transcripts generated through this project will belong to the researcher.

Responsibilities

The principal investigator, Rhys Madden, will be responsible for all aspects of data management outlined in this document, including garnering informed consent, securely storing data, ensuring data access and futurity through data archiving and sharing, and producing and sharing relevant metadata.

PGDonline

- En el año 2014, el Consorcio de Universidades de la Comunidad de Madrid y de la UNED para la Cooperación Bibliotecaria (Consorcio Madroño) desarrolló un portal para dar apoyo a los investigadores en la creación de un Plan de Gestión de Datos.
- El proyecto se denominó **PaGoDa**.
- Como parte de esa ayuda se llevó a cabo la traducción y adaptación de la herramienta de gestión PGDonline del Digital Curation Centre del Reino Unido.
- En 2016 el proyecto lanzón además eCienciaDatos https://edatos.consorciomadrono.es/, el primer repositorio de datos de las universidades españolas. Junto con estos dos servicios, Madroño ofrece material de apoyo y un servicio de asesoramiento en cada una de las bibliotecas miembro del consorcio.
- Además enlaza con ejemplos de planes de gestión de datos recopilados por el Digital Curation Centre, de diferentes disciplinas y de diferentes financiadores [en inglés].



Madroño Consorcio

Reunimos recursos, ofrecemos servicios

Buscar en el sitio

Inicio | Acerca de | Recursos electrónicos | Servicios | InvestigaM – Ciencia Abierta

> InvestigaM – Ciencia Abierta > Crear su PGD

INVESTIGAM — PAGODA: **CREAR SU PGD**

El Programa Horizonte 2020 requiere que los proyectos que formen parte del Piloto de Datos de Investigación en Abierto entreguen un Pla
de Datos (PGD) durante los 6 primeros meses del proyecto, como parte de las condiciones del contrato de subvención para un proyecto ci

Los planes de gestión de datos son una parte integral de las solicitudes de subvenciones – no pueden ser una idea de último momento; los
buscarán evidencia de que la gestión de datos está incluida en su propuesta, y que forma parte integral de su proceso de investigación. En
29.3 del H2020 Model Grant Agreement: Multi-beneficiary General MGA se establecen las obligaciones de los participantes en el Piloto de
Investigación en Abierto en lo que respecta a la gestión de los datos.

El documento Directrices sobre la Gestión de los Datos en Horizonte 2020 se dirige a los solicitantes y beneficiarios de los proyectos en el
Piloto de Datos de Investigación en Abierto y su objetivo es proporcionar indicaciones sobre cómo pueden cumplir con sus responsabilidad
respecto a la calidad de los datos de investigación, su intercambio y su seguridad.

Para saber más sobre el Piloto de Datos de Investigación en Abierto, consulte nuestras Faq's
Para saber más sobre el PGD como parte del Piloto de Datos de Investigación en Abierto, consulte nuestras Faq's
Para saber más sobre el Horizonte 2020, consulte el apartado Documentos

Recomendaciones para la preparación de su Plan de Gestión de Datos (PGD):

Antes de empezar a crear su Plan de Gestión de Datos es muy conveniente que previamente considere una serie de cuestiones relacionad
gestión de los datos en su proyecto.

Un punto de partida útil es la lista de control DCC (Digital Curation Centre) para un Plan de Gestión de Datos (en inglés) que podemos enc
de la selección de recursos del DCC (en inglés). Presenta los principales temas y preguntas que los investigadores pueden querer cubrir a
escribir un Plan de Gestión de Datos.

Si es la primera vez que se enfrenta a la elaboración de un Plan de Gestión de Datos, puede serle también muy útil leer algunos ejemplos
gestión de datos. Estos, recopilados por el Digital Curation Centre, incluyen ejemplos de diferentes disciplinas y de diferentes financiadores

> InvestigaM – Ciencia Abierta > Documentos

INVESTIGAM — PAGODA: **DOCUMENTOS**

Horizonte 2020

- Whats is Horizon 2020
- Horizon 2020 Online Manual
- Fact sheet: Open Access in Horizon 2020. 9 december 2013
- Opening up Scientific Information in Horizon 2020
- Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020. 25 August 2016. Version 3.1.
- Guidelines on FAIR Data Management in Horizon 2020. 26 July 2016. Version 3.0.
- Directrices sobre la Gestión de Datos en Horizonte 2020. 11 de diciembre de 2013. Versión 1.0.
- Open access to research data: the Open Research Data Pilot
- Multi-beneficiary General Model Grant Agreement. 20 July 2106. Version 3.0.
- AGA – Annotated Model Grant Agreement. 25 Noviembre 2016. Version 2.2
- Open Science (Open Access)
- Winning Horizon 2020 with Open Science

OpenAIRE y Datos de Investigación

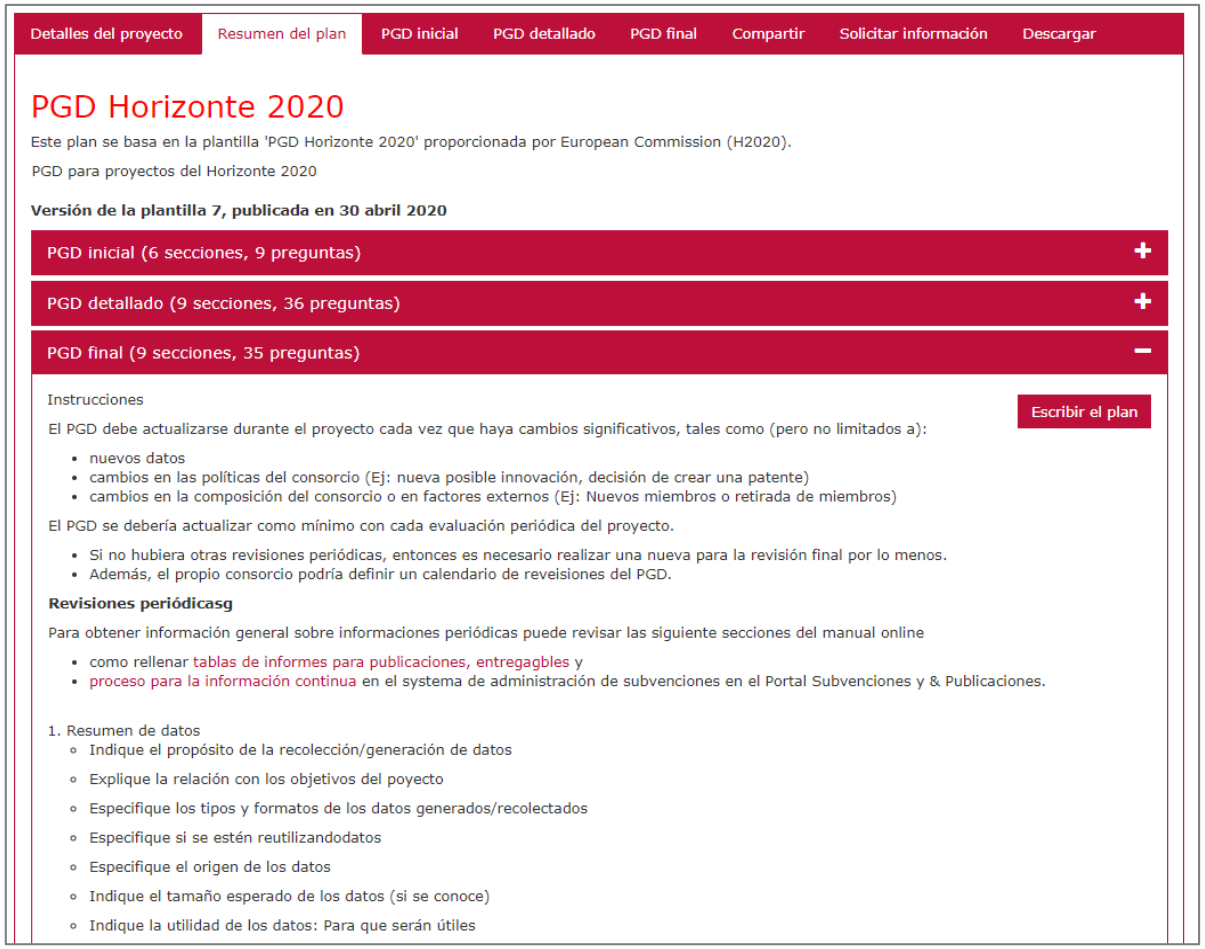
- How to create a DMP Plan
- How to select a repository
- OpenAIRE FAQ
- OpenAIRE and Eudat services and tools to support FAIR DMP implementation
- OpenAIRE Factsheet for Researchers
- OpenAIRE Factsheet for Research Administrators
- OpenAIRE Factsheet for FP7 post grant Open Access Pilot
- OpenAIRE Open Research Data Pilot Factsheet
- Guía OpenAIRE para investigadores de proyectos financiados por la Comisión Europea
- Piloto de Datos de Investigación en Abierto de H2020 ¿Qué ayuda puede proporcionar OpenAIRE?

Gestión de Datos de Investigación

- Presentaciones y vídeos del Curso "Publicación de datos Científicos" organizado por el Consorcio Madroño y celebrado los días 29 y 30 de noviembre de 2016
- RDA-CODATA Legal Interoperability Interest Group. (2016). Legal Interoperability of Research Data: Principles and Implementation Guidelines. Zenodo
- 23 Cosas: Bibliotecas para los Datos de Investigación
- Policy recommendations for open access to research data. RECODE.
- Infografías sobre Acceso Abierto y datos de investigación elaboradas por REBIUN.
- Expert tour guide on Data Management

http://www.consorciomadrono.es/investigam/crear-su-pgd/

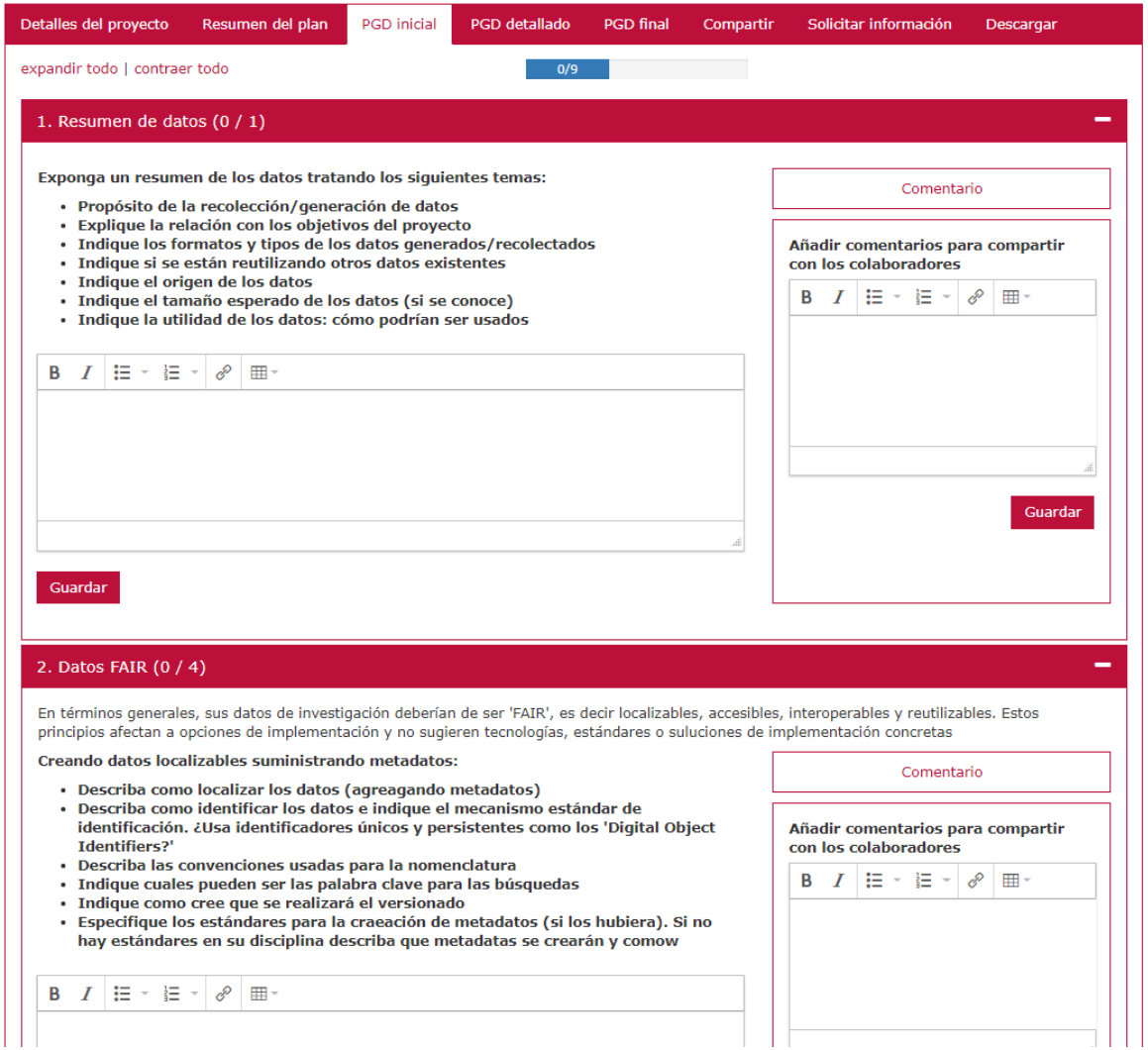
En la sección de ‘Resumen del plan’ se pueden desplegar las preguntas que sirven de guía para rellenar cada sección en los correspondientes apartados de ‘Escribir el plan’.



Las explicaciones que acompañan a las preguntas sirven de guía para no olvidar información importante en cada uno de los apartados, permite añadir comentarios para los colaboradores con los que se vaya a compartir la redacción del plan.

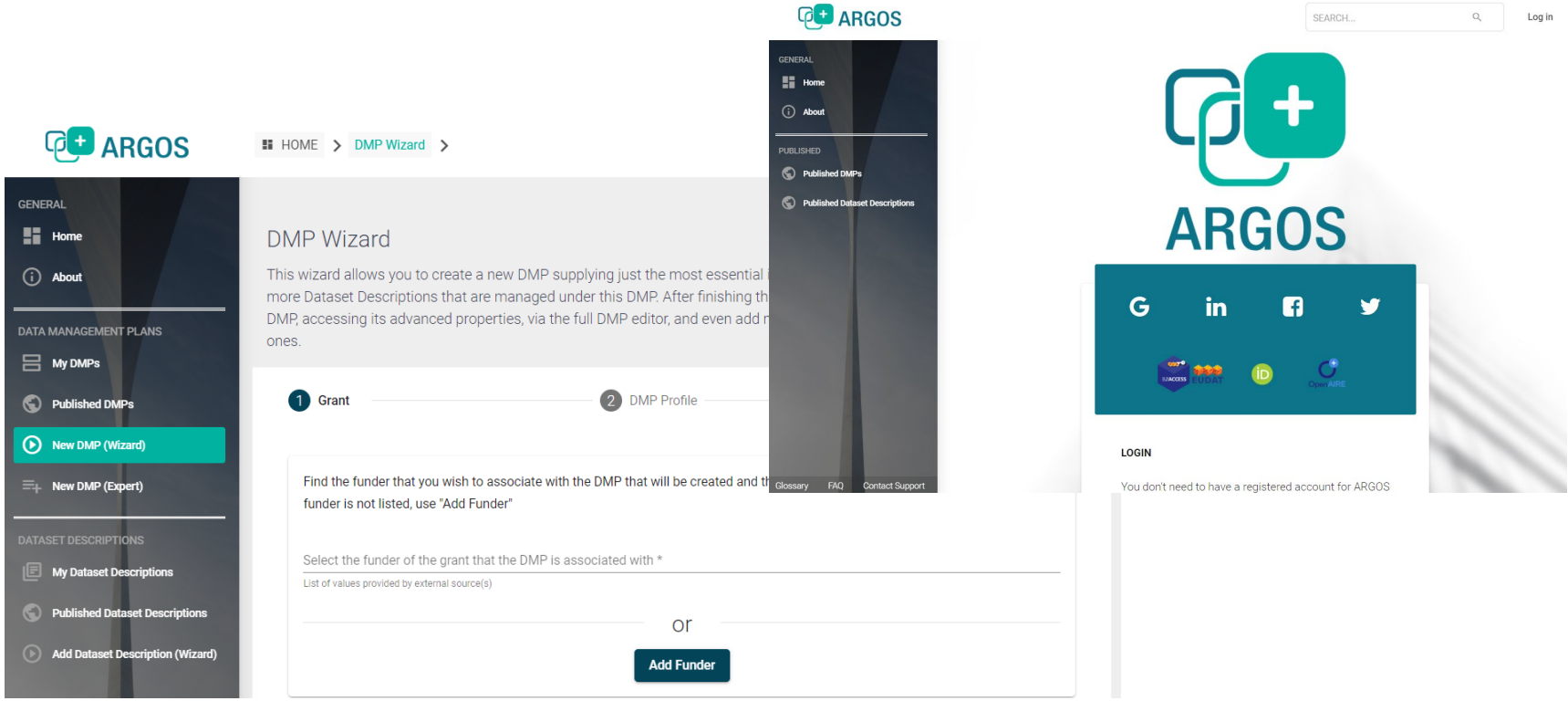
Los planes se pueden compartir a través de correo electrónico con otros colaboradores, a los que se pueden otorgar permisos de lectura o edición.

La descarga de planes se puede hacer en varios formatos, y permite ajustar tipos de fuente, tamaño y márgenes).



ARGOS - Create, Link, Share Data

https://argos.openaire.eu

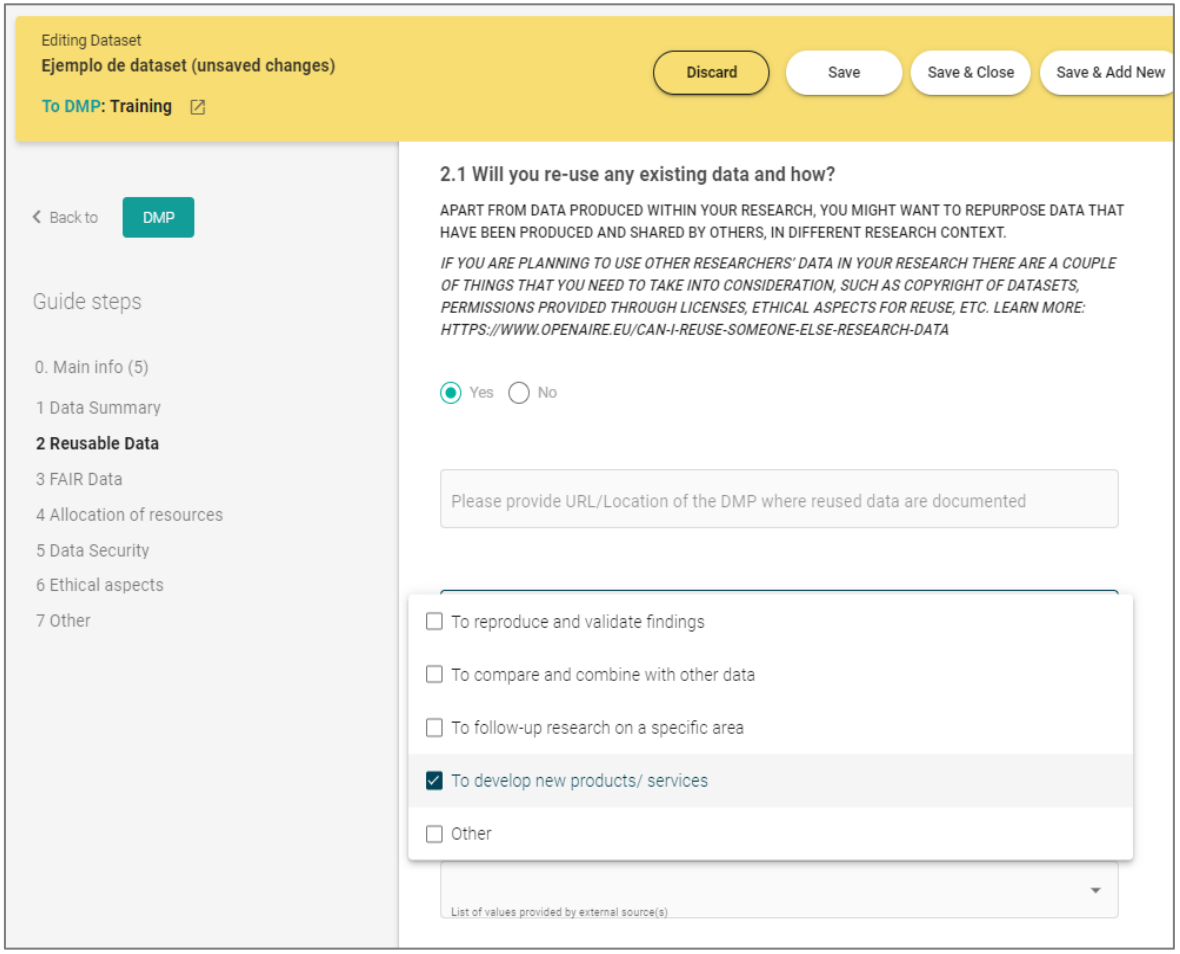


- Combina los campos imprescindibles de la plantilla del DMP con la plantilla de la descripción de los distintos conjuntos de datos.

ARGOS

- ARGOS es una herramienta de código abierto, configurable y extensible para planificar actividades de gestión de datos de investigación (RDM) de acuerdo con las políticas de datos de acceso abierto y FAIR.
- De uso gratuito para investigadores
- Proceso completo de generación y publicación de DMP de acuerdo con los principios FAIR.
- DMP procesables por máquina que hacen que los datos sean interoperables.
- Los PGD generados en ARGOS se gestionan a su vez como objetos resultado de investigación a los que se pueden asignar DOI, licencias, etc. Además, permite que los PGDs se hagan de forma colaborativa entre colegas del proyecto.

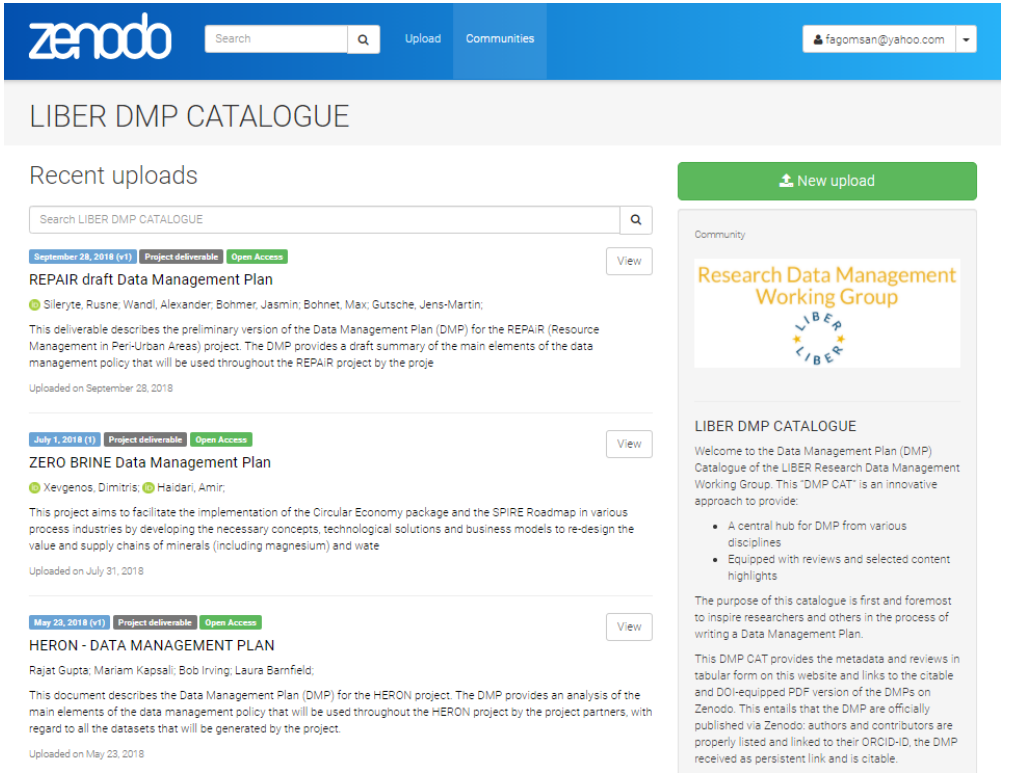
Al igual que el DMPonline tiene 7 apartados (información principal, resumen de los datos, datos reutilizables, datos FAIR, asignación de recursos, seguridad de los datos, aspectos éticos y otros).



La herramienta guía al usuario por cada una de las preguntas, y en función de las respuestas que el usuario vaya respondiendo, va generando opciones.

- Una vez completadas las preguntas de todos los apartados, el PGD podrá ser exportado a distintos formatos, compartido y publicado.
- Argos permite buscar ejemplos de planes de gestión de datos o de descripción de datasets entre los desarrollados por otros usuarios hayan decidido compartir de manera pública.

- En caso de que se busquen ejemplos concretos de una institución o de un organismo y que no haya ninguno disponible, remite a la comunidad de LIBER en Zenodo LIBER DMP Catalog.

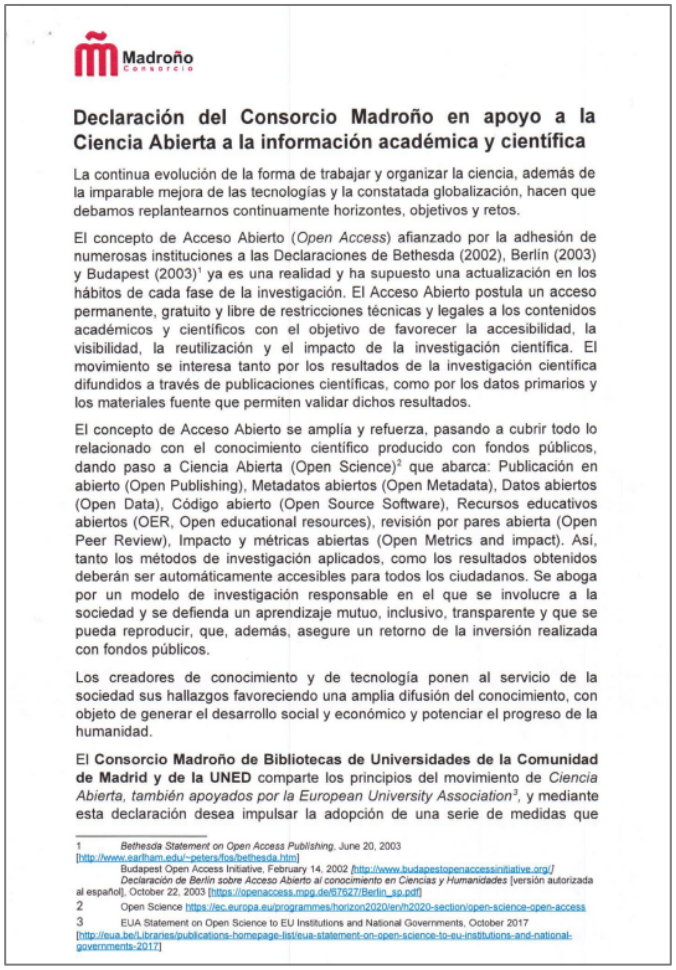


https://zenodo.org/communities/liber-dmp-cat/?page=1&size=20

¿QUÉ PUEDEN HACER LA INSTITUCIÓN Y LA BIBLIOTECA?

Desarrollo de políticas, buenas prácticas, etc.

- Política institucionales:
  - *De acceso abierto o, mejor, de ciencia abierta*
  - *De gestión de datos de investigación*
  - *De asignación de dois*
- Políticas de depósito en repositorios
- Firma o redacción de declaraciones
- Buenas prácticas en la gestión de los datos
  - *Fomentar el uso de formatos abiertos*
  - *Usar estándares, protocolos y vocabularios controlados ampliamente aceptados*
  - *Seleccionar los conjuntos de datos de interés*



Madroño Consorcio

Declaración del Consorcio Madroño en apoyo a la Ciencia Abierta a la información académica y científica

La continua evolución de la forma de trabajar y organizar la ciencia, además de la imparable mejora de las tecnologías y la constatada globalización, hacen que debamos replantearnos continuamente horizontes, objetivos y retos.

El concepto de Acceso Abierto (*Open Access*) afianzado por la adhesión de numerosas instituciones a las Declaraciones de Bethesda (2002), Berlín (2003) y Budapest (2003)[1] ya es una realidad y ha supuesto una actualización en los hábitos de cada fase de la investigación. El Acceso Abierto postula un acceso permanente, gratuito y libre de restricciones técnicas y legales a los contenidos académicos y científicos con el objetivo de favorecer la accesibilidad, la visibilidad, la reutilización y el impacto de la investigación científica. El movimiento se interesa tanto por los resultados de la investigación científica difundidos a través de publicaciones científicas, como por los datos primarios y los materiales fuente que permiten validar dichos resultados.

El concepto de Acceso Abierto se amplía y refuerza, pasando a cubrir todo lo relacionado con el conocimiento científico producido con fondos públicos, dando paso a Ciencia Abierta (Open Science)[2] que abarca: Publicación en abierto (Open Publishing), Metadatos abiertos (Open Metadata), Datos abiertos (Open Data), Código abierto (Open Source Software), Recursos educativos abiertos (OER, Open educational resources), revisión por pares abierta (Open Peer Review), Impacto y métricas abiertas (Open Metrics and impact). Así, tanto los métodos de investigación aplicados, como los resultados obtenidos deberán ser automáticamente accesibles para todos los ciudadanos. Se aboga por un modelo de investigación responsable en el que se involucre a la sociedad y se defienda un aprendizaje mutuo, inclusivo, transparente y que se pueda reproducir, que, además, asegure un retorno de la inversión realizada con fondos públicos.

Los creadores de conocimiento y de tecnología ponen al servicio de la sociedad sus hallazgos favoreciendo una amplia difusión del conocimiento, con objeto de generar el desarrollo social y económico y potenciar el progreso de la humanidad.

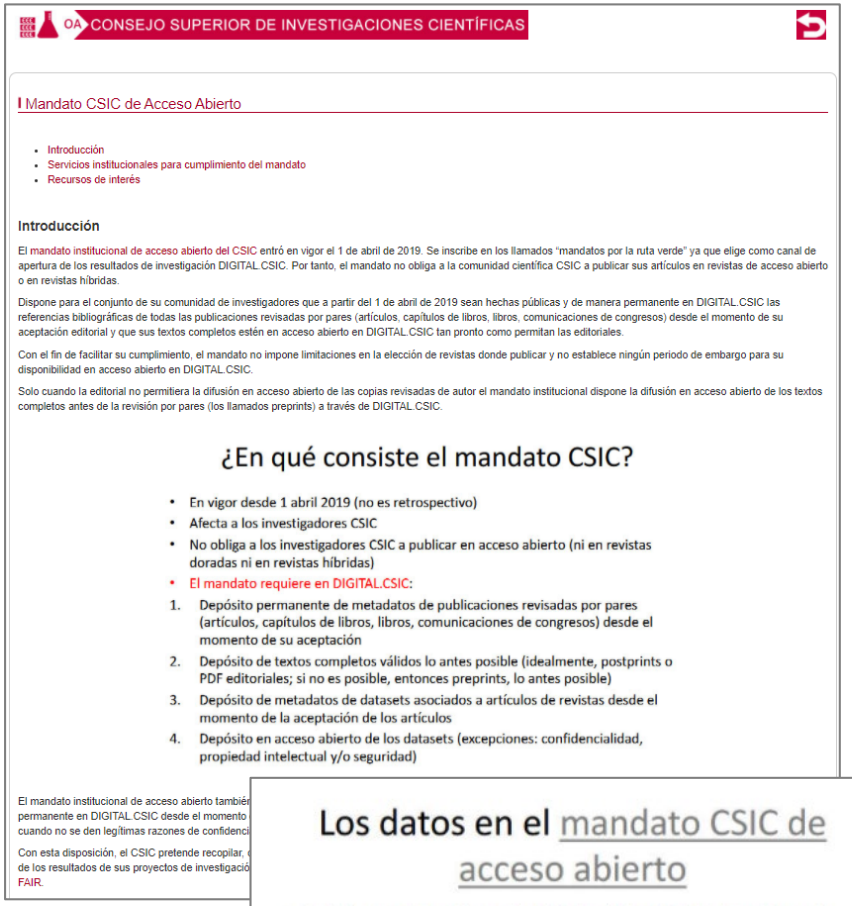
El **Consorcio Madroño de Bibliotecas de Universidades de la Comunidad de Madrid y de la UNED** comparte los principios del movimiento de *Ciencia Abierta, también apoyados por la European University Association*[3], y mediante esta declaración desea impulsar la adopción de una serie de medidas que

1 *Bethesda Statement on Open Access Publishing*, June 20, 2003 [http://www.earlham.edu/~peters/fos/bethesda.htm]
Budapest Open Access Initiative, February 14, 2002 [http://www.budapestopenaccessinitiative.org/]
Declaración de Berlín sobre Acceso Abierto al conocimiento en Ciencias y Humanidades [versión autorizada al español], October 22, 2003 [https://openaccess.mpg.de/67627/Berlin_sp.pdf]

2 Open Science https://ec.europa.eu/programmes/horizon2020/en/h2020-section/open-science-open-access

3 EUA Statement on Open Science to EU Institutions and National Governments, October 2017 [http://eua.be/Libraries/publications-homepage-list/eua-statement-on-open-science-to-eu-institutions-and-national-governments-2017]

http://www.consorciomadrono.es/docs/declaracion_ciencia_abierta.pdf



OA CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS

Mandato CSIC de Acceso Abierto

- Introducción
- Servicios institucionales para cumplimiento del mandato
- Recursos de interés

Introducción

El mandato institucional de acceso abierto del CSIC entró en vigor el 1 de abril de 2019. Se inscribe en los llamados "mandatos por la ruta verde" ya que elige como canal de apertura de los resultados de investigación DIGITAL.CSIC. Por tanto, el mandato no obliga a la comunidad científica CSIC a publicar sus artículos en revistas de acceso abierto o en revistas híbridas.

Dispone para el conjunto de su comunidad de investigadores que a partir del 1 de abril de 2019 sean hechas públicas y de manera permanente en DIGITAL.CSIC las referencias bibliográficas de todas las publicaciones revisadas por pares (artículos, capítulos de libros, libros, comunicaciones de congresos) desde el momento de su aceptación editorial y que sus textos completos estén en acceso abierto en DIGITAL.CSIC tan pronto como permitan las editoriales.

Con el fin de facilitar su cumplimiento, el mandato no impone limitaciones en la elección de revistas donde publicar y no establece ningún periodo de embargo para su disponibilidad en acceso abierto en DIGITAL.CSIC.

Solo cuando la editorial no permitiera la difusión en acceso abierto de las copias revisadas de autor el mandato institucional dispone la difusión en acceso abierto de los textos completos antes de la revisión por pares (los llamados preprints) a través de DIGITAL.CSIC.

¿En qué consiste el mandato CSIC?

- En vigor desde 1 abril 2019 (no es retrospectivo)
- Afecta a los investigadores CSIC
- No obliga a los investigadores CSIC a publicar en acceso abierto (ni en revistas doradas ni en revistas híbridas)
- El mandato requiere en DIGITAL.CSIC:

1. Depósito permanente de metadatos de publicaciones revisadas por pares (artículos, capítulos de libros, libros, comunicaciones de congresos) desde el momento de su aceptación
2. Depósito de textos completos válidos lo antes posible (idealmente, postprints o PDF editoriales; si no es posible, entonces preprints, lo antes posible)
3. Depósito de metadatos de datasets asociados a artículos de revistas desde el momento de la aceptación de los artículos
4. Depósito en acceso abierto de los datasets (excepciones: confidencialidad, propiedad intelectual y/o seguridad)

El mandato institucional de acceso abierto también
permanente en DIGITAL.CSIC desde el momento
cuando no se den legítimas razones de confidenci

Con esta disposición, el CSIC pretende recopilar,
de los resultados de sus proyectos de investigació
FAIR.

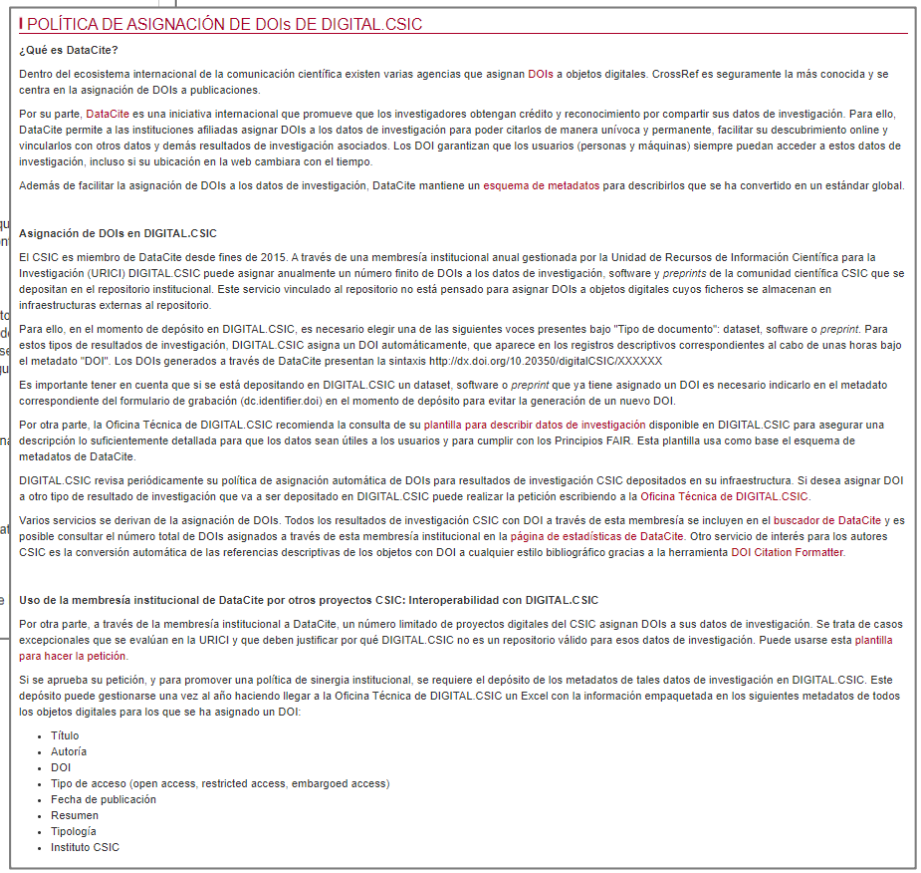
Los datos en el mandato CSIC de acceso abierto

- Solicita que **las referencias bibliográficas (…) de los *datasets* asociados a artículos de revistas** sean hechas públicas **de manera permanente en DIGITAL.CSIC desde el momento de la aceptación para su publicación de los artículos asociados**.

- Solicita que se ofrezcan en modalidad de **acceso abierto en DIGITAL.CSIC aquellos *datasets* asociados a publicaciones siempre y cuando no se den legítimas razones de confidencialidad, propiedad intelectual y/o seguridad. Estos *datasets* en acceso abierto deben ser FAIR e ir acompañados de una licencia estándar** que explícitamente indique las condiciones de uso **y favorezca la reproducibilidad científica (por ejemplo, Creative Commons y Open Data Commons)**

http://digital.csic.es/dc/mandato-oa-csic.jsp

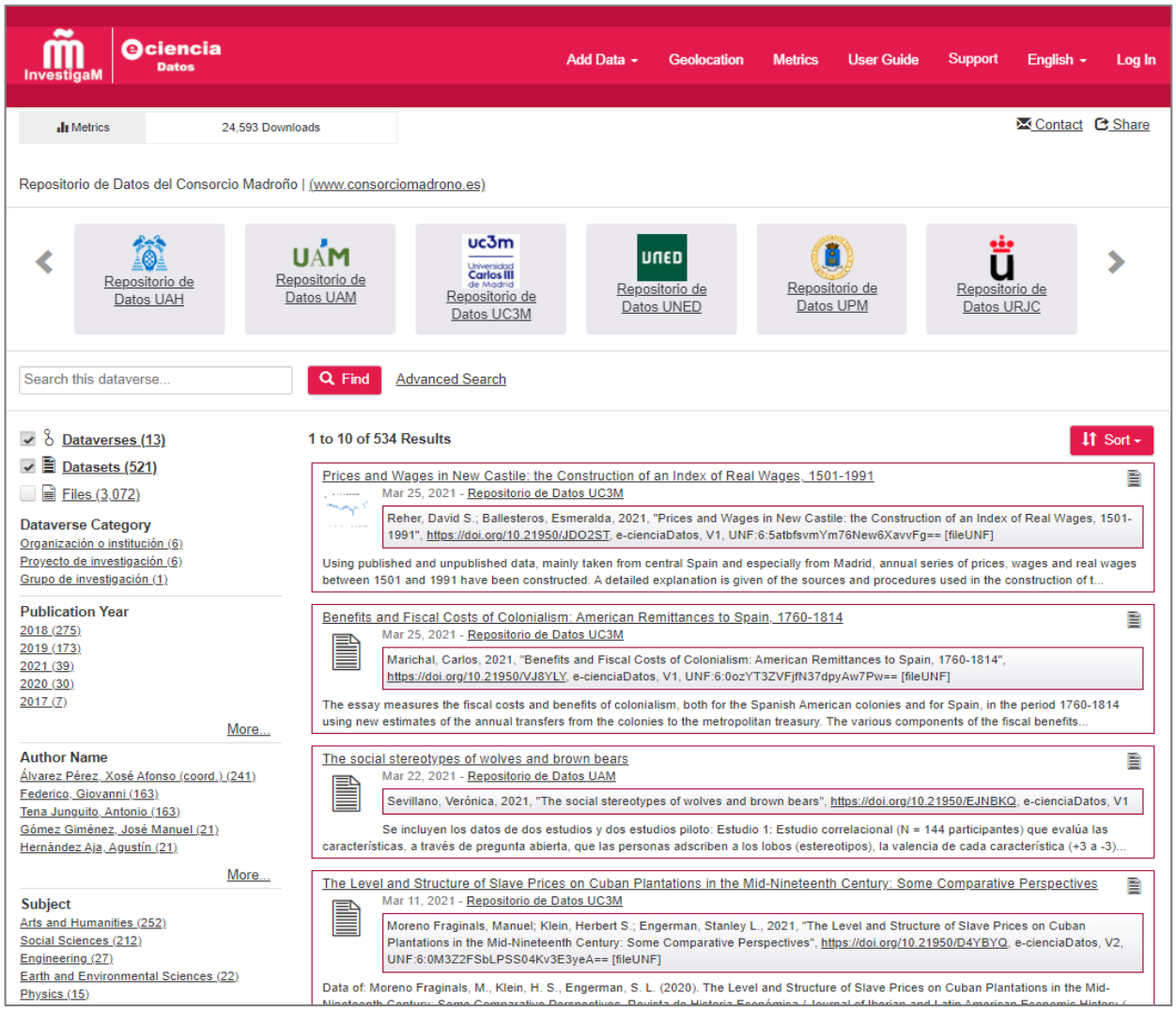
http://digital.csic.es/dc/politicas/politica-datos-digital-csic.jsp



http://digital.csic.es/dc/politicas/politica-asignacion-dois.jsp

Ejemplo de adaptación del repositorio para depositar datos

Ejemplo de repositorio de Datos del Consorcio Madroño: e-cienciaDatos



https://edatos.consorciomadrono.es/

Creación de plantillas

RECUERDE: Los datos con (*) son obligatorios

Palabras clave*:

Palabras o frases que describen el tema o el contenido de los datos. Se recomienda usar entradas procedentes de un vocabulario controlado. *Ejemplos: entradas de Library of Congress Subject Headings (LCSH), Library of Congress Classification (LCC), Medical Subject Headings (MESH), Dewey Decimal Classification (DDC) o el Catálogo de Autoridades de la Red de Bibliotecas y Archivos del CSIC.*

Resumen*:

Resuma de los objetivos y conclusiones de la investigación realizada. Suele recoger la motivación del proyecto, el resumen del contenido del dataset y su estructura, así como su metodología y proceso de validación. Se aceptan hasta 2.000 caracteres.

Está basado en:

Indique si el conjunto de datos es la versión de un conjunto de datos precedente. Debe incluirse la URI del dataset previo o en su defecto, la citación bibliográfica de un dataset asociado precedente.

Es parte de:

Indique si el conjunto de datos forma pa

Es referenciado por:

Referencie los trabajos (por ejemplo, art
mínimo los siguientes datos: Apellido, N

Patrocinadores*:

Escriba los nombres de los patrocinador

Descripción:

Ingrese cualquier otra descripción o com

RECUERDE: Los datos con (*) son obligatorios

Twitter:

¿Desea que se publique en Twitter de RIUMA? Escriba Sí o No.

Embargo hasta una fecha específica:

Debe consignar el primer día en que se permite el acceso. En formato: aaaa-mm-dd.

Motivo del embargo:

La razón del embargo, para uso interno normalmente. Opcional.

Licencie su trabajo*:

Escriba si desea asignarle a su ítem una licencia Creative Commons. Las licencias Creative Commons determinan qué pueden hacer los lectores con su trabajo. La licencia asignada debe ser la menos restrictiva, se recomienda CC BY (permitir uso comercial y permitir modificaciones), CC BY-NC-SA (No comercial, compartir igual) o CC BY-SA (Compartir igual).

Indique también la licencia en caso de depositar código informático o software (Licencia Apache, GNU, etc.)

RECUERDE: Los datos con (*) son obligatorios

Licencia de distribución UMA*:

Marque con una X, si acepta y da autorización al Servicio de Automatización de la Biblioteca de la Universidad de Málaga de aceptar, en su nombre, la licencia de distribución de la Universidad de Málaga para permitir reproducir, traducir y distribuir su envío a través del mundo. (Siga el enlace para más información sobre la licencia de distribución).

Sí, acepto y doy autorización al Servicio de Automatización de la Biblioteca de la Universidad de Málaga de aceptar, en mi nombre, la licencia de distribución de la Universidad de Málaga.

¡IMPORTANTE!: es imprescindible aceptar la licencia de distribución de la Universidad de Málaga para poder depositar el dataset.

Plantilla de descripción de datos de investigación

RiUMA

UNIVERSIDAD DE MÁLAGA | BUMA BIBLIOTECA UNIVERSITARIA UNIVERSIDAD DE MÁLAGA

28/04/2021

RECUERDE: Los datos con (*) son obligatorios

Autor*:

Apellido y Nombre(s)

Escriba el nombre completo de los autores de este ítem. Obras en colaboración: Se debe de contar con la autorización de todos los autores, se incluirán separados por (;).

Ej. Pérez Rodríguez, Manuel; Gómez Martín, Carmen

Título*:

Escriba el título principal del ítem. (Es muy importante asignar un título unívoco al dataset que se está describiendo y evitar, por ejemplo, asignarle el mismo título que el del artículo científico con el que se asocia, o con el proyecto al que pertenece)

Título Alternativo:

Si el ítem tiene títulos alternativos, por favor escríbalos. (también a nivel de idioma)

Fecha de publicación*:

Escriba la fecha de depósito en RIUMA completa: Año, Mes y Día. Ej. 2021/04/26

Fecha de Creación del Recurso:

Escriba la fecha de creación del recurso. Puede dejar en blanco el día y/o el mes si no aplican.

Grupos de Investigación*:

Rellene con el nombre de los grupos de investigación.

Departamentos*:

Escriba el departamento al que pertenece el material (Puede añadir ser varios).

Centro:

Escriba la Facultad o Escuela Técnica Superior de origen.

RECUERDE: Los datos con (*) son obligatorios

Editorial:

Indique el nombre de la editorial o el editor de este ítem, p.ej. Universidad de Málaga.

Cobertura espacial:

Hace referencia al lugar de recogida de datos o al lugar al que hacen referencia los datos.

Cobertura temporal:

Hace referencia al periodo que cubre el dataset o el periodo de recogida de datos.

Cita:

Escriba la cita estándar para la publicación previa de este ítem (cuando haya sido publicado en otro repositorio previamente). Debe incluir como mínimo los siguientes datos: Apellido, Nombre. Título. Handle y/o DOI.

Identificadores:

Escriba el código de identificación DOI, ISSN, ISBN, ISMN u otro, según el ejemplo:
DOI (código)
ISSN (código)...

Nº de versión:

Introduzca el número de versión, *ej. 1.0.*

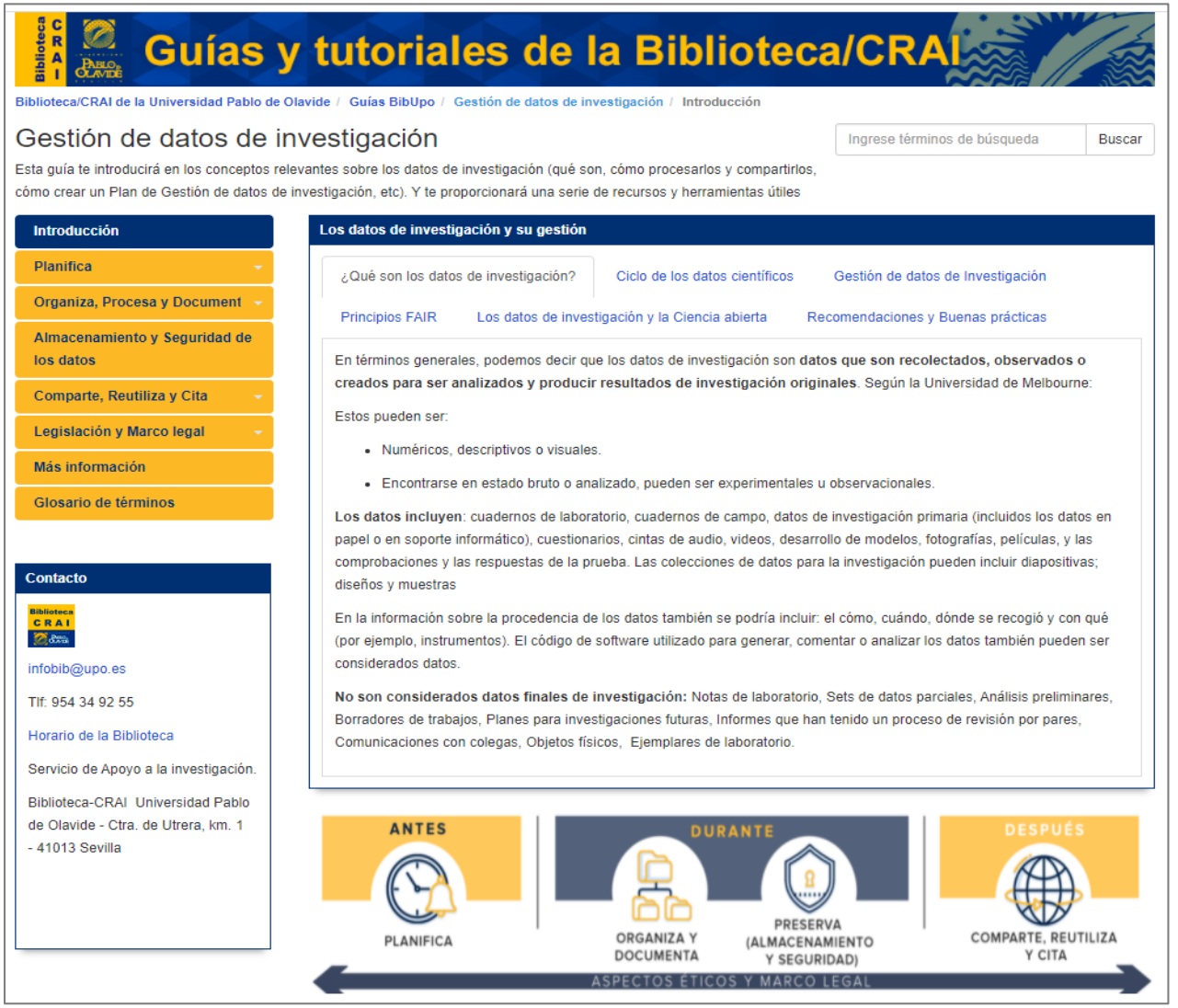
Tipo*:

DATASET

Idioma*:

Escriba el lenguaje del contenido principal del ítem. Si el contenido no tiene un idioma (por ejemplo, si es una imagen) ponga 'N/A'.

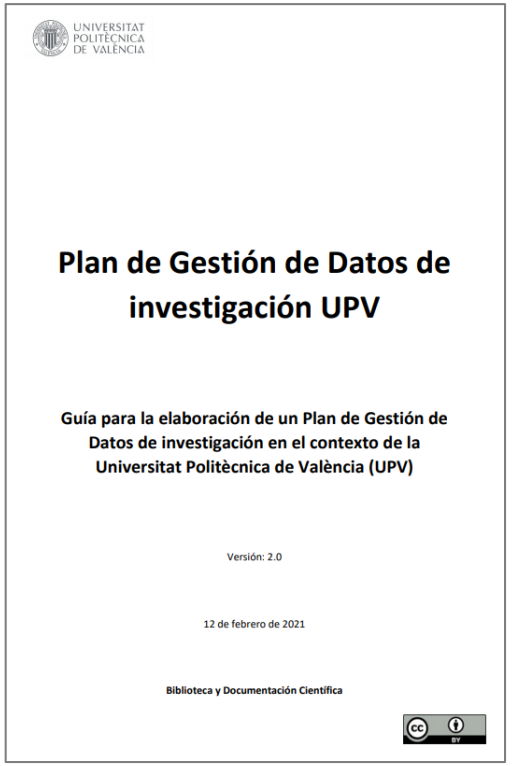
Creación de guías y tutoriales



https://guiasbib.upo.es/gestion_datos_de_investigacion

https://biblioguias.webs.upv.es/bg/index.php/es/datos-de-investigacion

Rubio Montero, FJ. (2021). Plan de Gestión de Datos de investigación UPV: guía para la elaboración de un Plan de Gestión de Datos de investigación en el contexto de la Universitat Politècnica de València (UPV). Universitat Politècnica de València. http://hdl.handle.net/10251/143184



¡Muchas Gracias!

info@aliciafgomez.com

https://twitter.com/fagomsan

http://es.linkedin.com/in/aliciafgomez

https://orcid.org/0000-0003-4898-1680

CC BY SA