

GESTIÓN DE DATOS DE INVESTIGACIÓN, DATOS FAIR Y PLANES DE GESTIÓN DE DATOS

Dra. Alicia Fátima Gómez Sánchez

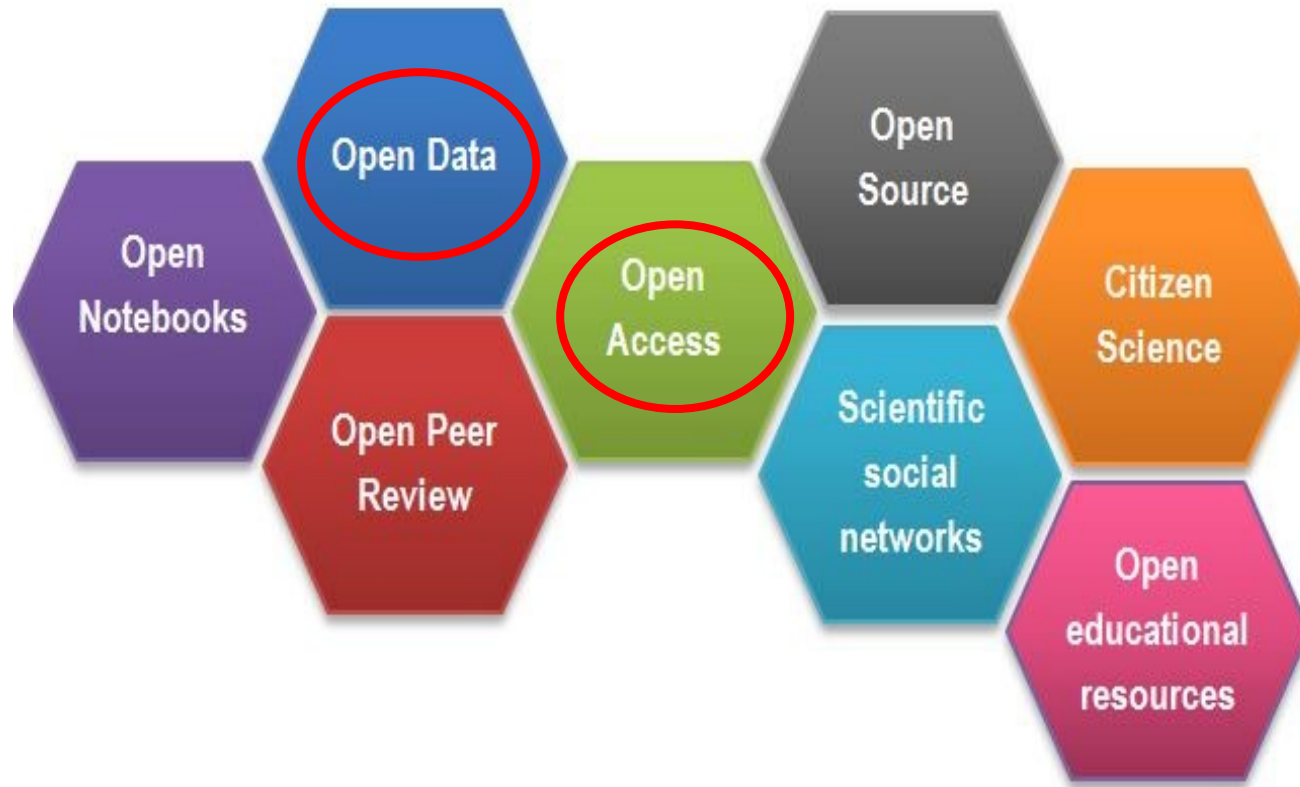
22 DE FEBRERO DE 2024

Objetivo general:

- Conocer el marco legal de la ciencia abierta y del acceso abierto en relación a los datos de investigación
- Comprender qué significan los datos FAIR
- Entender qué es un plan de gestión de datos.

Ciencia abierta y acceso abierto

Ciencia abierta



¿Para qué?

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

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>

La ciencia abierta implica un cambio:

Cambio de paradigma

Ciencia Cerrada	Ciencia Abierta
<ul style="list-style-type: none">• Basada en publicar artículos científicos	<ul style="list-style-type: none">• Basada en publicar cualquier resultado de investigación
<ul style="list-style-type: none">• Ciencia individualista	<ul style="list-style-type: none">• Ciencia colaborativa
<ul style="list-style-type: none">• Solo unos pocos tienen acceso a los resultados de investigación	<ul style="list-style-type: none">• Toda la sociedad tiene acceso a los resultados de investigación
<ul style="list-style-type: none">• Ciencia vertical / especializada	<ul style="list-style-type: none">• Ciencia horizontal, interdisciplinar
<ul style="list-style-type: none">• Sin participación ciudadana	<ul style="list-style-type: none">• Ciencia Ciudadana con y para la sociedad
<ul style="list-style-type: none">• Sistema de crédito basado en el impacto de las publicaciones	<ul style="list-style-type: none">• Sistema de crédito basado en el impacto del investigador
<ul style="list-style-type: none">• Factor de Impacto (FI – JCR)	<ul style="list-style-type: none">• 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/>

Resumiendo: 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.

Marco legal: Requisitos europeos y nacionales.

Búsqueda de políticas y mandatos.

Conclusions OS in Horizon Europe

Horizon Europe Regulation

Open
Science
*as modus
operandi*

- Promote OS: science communication and citizen science, among others, e.g. through a combination of obligations and incentives (possibly also in the evaluation of proposals)
- Sanctions for those grant beneficiaries (e.g. not necessarily the same as researchers) that repeatedly and consistently fail to provide the required OA
- Appropriate metrics for better assessing the impact of research output and engagement open science

https://www.slideshare.net/OpenAIRE_eu/horizon-2020-open-science-policies-and-beyond-with-emilie-hermans-openaire



Main elements of Open Science in Horizon Europe

Horizon Europe Regulation



Open access to publications ensured (=no way around this! ☺)



Open access to research data: 'as open as possible as closed as necessary'



Responsible management of research data: Data Management Planning, FAIR data, long-term preservation of data



Open science practices promoted and encouraged; may provide additional incentives or obligations to adhere to open science practices



May require additional obligations to use EOSC for storing and giving access to research data



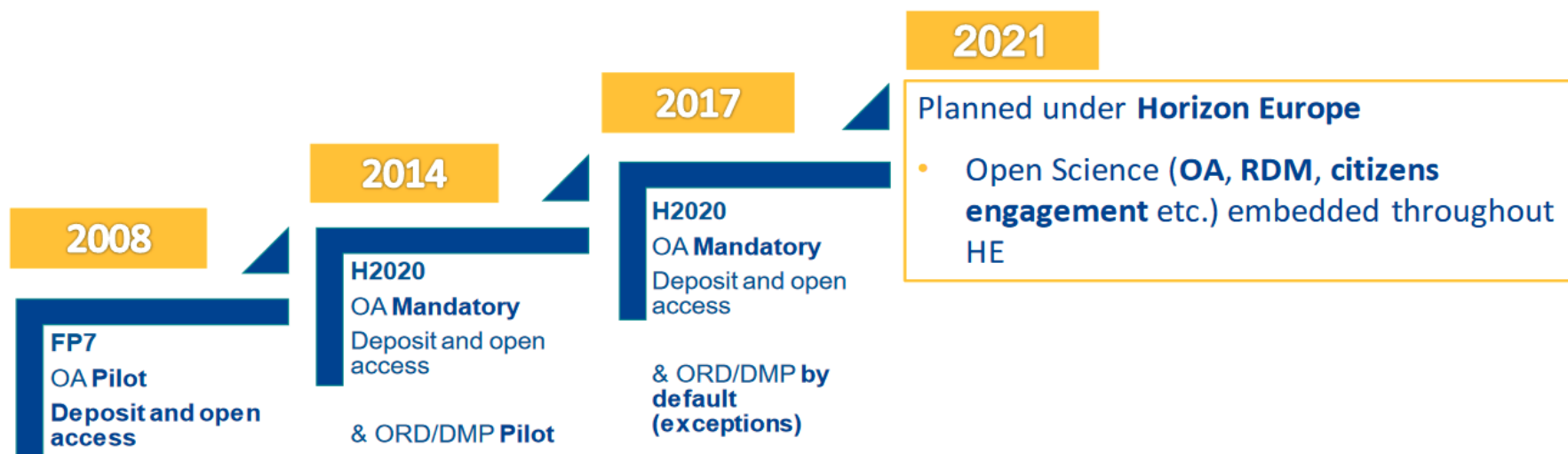
Authors/beneficiaries must retain enough rights for open access



Exceptions to open access for research data described

Del acceso abierto a la ciencia abierta en Horizonte Europa

Evolution of open science policy across FPs



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.



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

Qué requiere la normativa europea

Garantizar el acceso abierto a:

- ✓ Los **datos de investigación y los metadatos** necesarios para validar los resultados presentados en publicaciones científicas
- ✓ Otros datos y metadatos incluidos en el Plan de gestión de datos (PGD)

EXCEPCIONES:

- Protección derechos de propiedad industrial
- Protección datos personales
- Confidencialidad en relación a asuntos de seguridad
- El objetivo principal del proyecto se vea comprometido
- No se generen datos en el proyecto
- Otras razones a justificar

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

¡Indispensable!!



- El mandato de acceso abierto se extiende a libros y otras publicaciones de formato extenso.
- Acceso abierto inmediato, no más embargos, no se reembolsan las tarifas de las publicaciones en lugares 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.
- Open Research Europe: una nueva plataforma de publicación de acceso abierto para los beneficiarios de H2020 / Horizonte Europa.

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.

OPEN SCIENCE

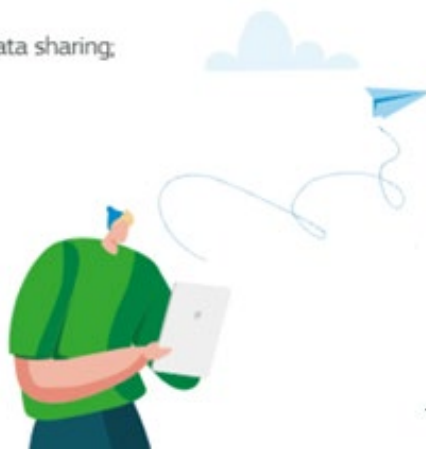
EARLY KNOWLEDGE AND DATA SHARING, AND OPEN COLLABORATION



The challenge is for Europe **to embrace open science as the modus operandi for all researchers**. Open science consists in the sharing of knowledge, data and tools as early as possible in the Research and Innovation (R&I) process, in open collaboration with all relevant knowledge actors, including academia, industry, public authorities, end users, citizens and society at large. Open science has the potential to increase the quality, efficiency and impact of R&I, lead to greater responsiveness to societal challenges, and increase trust of society in the science system.

What are open science practices?

- Open access to research outputs such as publications, data, software, models, algorithms, and workflows;
- Early and open sharing of research, for example through preregistration, registered reports, pre-prints, and crowd-sourcing of solutions to a specific problem;
- Use of open research infrastructures for knowledge and data sharing;
- Participation in open peer-review;
- Measures to ensure reproducibility of results; and
- Open collaboration within science and with other knowledge actors, including involving citizens, civil society and end-users, such as in citizen science.



Open science across Horizon Europe

Legal provisions in the grant agreements will strengthen open access rights and obligations for beneficiaries. Horizon Europe will require immediate open access to all scientific publications and responsible research data management so that data are **Findable, Accessible, Interoperable and Re-usable (FAIR)**. Data will be made 'as open as possible, but will be allowed to stay as closed as necessary', safeguarding legitimate interests or constraints. However access to research outputs shall be provided for third parties to be able to verify or validate publications.

The **criteria for evaluating research proposals** under Horizon Europe will take into account the quality and appropriateness of the open science practices in the submitted proposals. This will be assessed as part of the project's methodology, under the excellence award criterion. This will provide a strong incentive for applicants and beneficiaries to practice open science.

<https://op.europa.eu/en/web/eu-law-and-publications/publication-detail/-/publication/9570017e-cd82-11eb-ac72-01aa75ed71a1#>

Ley 17/2022, de 5 de septiembre, por la que se modifica la Ley 14/2011, de 1 de junio, de la Ciencia, la Tecnología y la Innovación.

Uno de los objetivos generales de esta ley es precisamente “Impulsar la ciencia abierta al servicio de la sociedad y promover iniciativas **orientadas a facilitar el libre acceso a los datos, documentos y resultados generados por la investigación**, desarrollar infraestructuras y plataformas abiertas, y **fomentar la participación abierta de la sociedad civil en los procesos científicos**”.

“Se impulsará que se haga difusión de los resultados de la actividad científica, tecnológica y de innovación, y que los resultados de la investigación, incluidas las publicaciones científicas, **datos, códigos y metodologías, estén disponibles en acceso abierto**. El acceso gratuito y libre a los resultados se fomentará mediante el desarrollo de repositorios institucionales o temáticos de acceso abierto.”

https://boe.es/diario_boe/txt.php?id=BOE-A-2022-14581

- El personal de investigación del sector público o cuya actividad investigadora esté financiada mayoritariamente con fondos públicos y que opte por diseminar sus resultados de investigación en publicaciones científicas, **deberá depositar una copia de la versión final aceptada para publicación y los datos asociados a las mismas en repositorios institucionales o temáticos de acceso abierto, de forma simultánea a la fecha de publicación.**
- Además del acceso abierto, y siempre con el objetivo de hacer la ciencia más abierta, accesible, eficiente, transparente y beneficiosa para la sociedad, los Ministerios de Ciencia e Innovación y de Universidades, y las Comunidades Autónomas promoverán también otras **iniciativas orientadas a facilitar el libre acceso y gestión de los datos generados por la investigación (datos abiertos), de acuerdo a los principios internacionales FAIR (sencillos de encontrar, accesibles, interoperables y reutilizables)**, a desarrollar infraestructuras y plataformas abiertas, a fomentar la publicación de los resultados científicos en acceso abierto, y la participación abierta de la sociedad civil en los procesos científicos

Estrategia Española de Ciencia, Tecnología e Innovación 2021-2027

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.

EECTI

Estrategia Española de
Ciencia, Tecnología e
Innovación
2021-2027



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*. (p. 36)

<https://www.ciencia.gob.es/site-web/Estrategias-y-Planes/Estrategias/Estrategia-Espanola-de-Ciencia-Tecnologia-e-Innovacion-2021-2027.html>

Modelos de justificación científica de los Proyectos de Investigación en Salud (PI).

Programa Estatal de Generación de Conocimiento y Fortalecimiento del Sistema Español de I+D+I

<https://www.isciii.es/QueHacemos/Financiacion/Seguimiento/Paginas/DocumentosNormalizados.aspx>

https://www.isciii.es/QueHacemos/Financiacion/Seguimiento/Documents/PI_JUSTIF_MEMORIAS_INTERMEDAS_2023.pdf

JUSTIFICANTES	DESCRIPCIÓN	OBLIGATORIO S/N
Plan de gestión de datos	NUEVO <ul style="list-style-type: none">• Tipología y formato de los datos recogidos/generados en el marco del proyecto• Procedimiento desarrollado para el acceso a los mismos (quién, cómo y cuándo puede acceder a ellos)• Titularidad de los datos• Repositorio en el que se realizará su depósito,• Procedimiento para garantizar los requisitos éticos o legales específicos de aplicación. Además, se adjuntará un PGD como documento adjunto	SI
Perspectiva de Género	NUEVO <p>Cómo la perspectiva de género está integrada en su investigación y contenido técnico. Si considera que no es necesario, justifique por qué</p>	NO
Indicador FEDER RCO_06	NUEVO <p>Medirá la dedicación a I+d+i teniendo en cuenta el % de dedicación al proyecto, de cada miembro del equipo</p>	SI

**(Versión actualizada
en diciembre de 2023)**

PREGUNTAS FRECUENTES
PROYECTOS DE GENERACIÓN DE CONOCIMIENTO. CONVOCATORIA 2023

VERSIÓN	CAUSA DEL CAMBIO	FECHA DEL CAMBIO
1	Documento inicial	14/12/2023

SIGLAS	11
DENOMINACIÓN DE LA CONVOCATORIA	11
1. ¿Cómo se denomina la convocatoria de 2023?.....	11
MODALIDADES Y TIPOS DE PROYECTO	11
2. ¿Cuáles son las modalidades de proyectos?.....	11
3. ¿Cuáles son los tipos de proyectos?.....	12
4. Proyectos tipo A.....	12
4.1. Características y requisitos de los proyectos tipo A.....	12
4.2. ¿Existe una edad límite para ser IP de un proyecto de tipo A?.....	12
4.3. Un Centro de Apoyo Tecnológico ¿puede solicitar un proyecto de tipo A con un/a IP que no tiene doctorado ni ha sido IP en una convocatoria anterior?.....	12
5. Características y requisitos de los proyectos tipo B.....	13
6. Características y requisitos de los proyectos tipo RTA.....	13
7. La convocatoria menciona que se pretende fomentar tanto la multidisciplinariedad como la interdisciplinariedad. ¿A qué se refieren estos conceptos?.....	13
PROYECTOS COORDINADOS	13
8. Características de los proyectos coordinados.....	14
8.1. ¿En qué se distingue un proyecto coordinado (con dos IP y sendos equipos de investigación) de otro no coordinado con dos IP?.....	14
8.2. ¿Qué peculiaridades tiene en la aplicación de solicitud la presentación de un proyecto coordinado?.....	14
8.3. ¿Quién es el/a IP de un proyecto coordinado?.....	14
8.4. ¿Los subproyectos de un coordinado pueden ser de tipo A?.....	15
8.5. ¿Pueden pertenecer a la misma entidad o institución los distintos subproyectos de un coordinado?.....	15
8.6. ¿Cómo se financian los proyectos coordinados?.....	15
8.7. ¿Todos los subproyectos de un proyecto coordinado tienen que tener la misma duración?.....	15
8.8. ¿Pueden ser diferentes las áreas temáticas de los subproyectos?.....	15
8.9. ¿Cómo se evalúan los proyectos coordinados?.....	15

83.	¿Se permite la inclusión de EPO (Ente Promotor Observar) en la memoria científico-técnica?.....	39
84.	Los documentos adjuntos ¿pueden aportarse en cualquier formato digital?.....	39
85.	¿Qué características debe tener la memoria científico-técnica?.....	40
85.1.	¿Cuántas memorias científico-técnicas hay en los proyectos coordinados?.....	40
85.2.	Si no he tenido tiempo de rellenar algunos apartados de la memoria científico-técnica, ¿podré adjuntarla corregida en otro momento?.....	40
85.3.	Si no adjunto la memoria científico-técnica en el modelo normalizado ¿Es posible subsanar este defecto?.....	40
85.4.	¿Qué sucede si no se presenta la memoria en lengua inglesa cuando es obligatorio hacerlo?.....	40
85.5.	Datos de Investigación - ¿Que son los datos de investigación?.....	41
85.6.	Datos de Investigación - ¿Qué es un Plan de Gestión de Datos?.....	41
85.7.	Datos de Investigación - ¿Debo incluir en la solicitud de mi proyecto el Plan de Gestión de Datos (PGD)?.....	41
85.8.	Datos de Investigación - ¿Cuándo debo enviar el Plan de Gestión de Datos?.....	41
85.9.	Datos de Investigación - ¿Qué significa que los datos de investigación sean FAIR?.....	41
85.10.	Datos de Investigación - ¿Es lo mismo datos abiertos que datos "FAIR"?.....	42
85.11.	Datos de Investigación - ¿Qué datos de investigación se deben depositar?.....	42
85.12.	Datos de Investigación - ¿Qué repositorios se podrán utilizar para depositar los datos?.....	42
85.13.	¿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 <i>link</i> de ese trabajo en la web.....	42
85.14.	En las instrucciones para rellenar la memoria, en el apartado de capacidad formativa, se recomienda incluir una relación de tesis realizadas o en curso. ¿Hay que tener en cuenta todas las tesis del equipo investigador de la propuesta o solo aquellas del investigador que solicita al FPI?.....	43
85.15.	Si hay un gran número de tesis doctorales defendidas en el período de tiempo indicado, la inclusión de toda la información relativa a la misma implicaría un gran número de páginas de la memoria científica y la única solución posible es limitar la información que se incluye ¿Cuál sería la información más relevante a incluir para los evaluadores?.....	43
85.16.	Dentro de la información sobre las tesis realizadas o en curso, hay que incluir las "publicaciones en las que figura el contratado predoctoral". ¿A qué se refiere?.....	43
85.17.	En la parte en la que se pide una "Breve descripción del desarrollo científico o profesional de los doctores egresados del equipo de investigación durante los últimos 10 años", ¿a qué se refiere?.....	43
85.18.	La convocatoria obliga a que se publique en abierto, pero ¿a qué se refiere?.....	43
85.19.	La publicación de resultados en repositorios abiertos, ¿se refiere por ejemplo a las revistas científicas a las que accede la comunidad científica del campo científico correspondiente?.....	43
85.20.	¿Qué se debe incluir en el apartado de Condiciones Específicas en relación con los puntos de: "Una indicación de las instalaciones que son preceptivas y de las que dispone en su institución para la ejecución del proyecto. Y en su caso, su previsión para aquellas instalaciones de las que no dispone" y "La relación de autorizaciones preceptivas de las que ya dispone, o que se encuentren en tramitación, para las actividades previstas en el proyecto"?.....	44

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.

85.6. 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 este 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 2023
«PROYECTOS DE
GENERACION DE
CONOCIMIENTO»**

Instrucciones para la elaboración del Plan de Gestión de Datos (PGD)



Instrucciones para la elaboración del Plan de Gestión de Datos (PGD)

Para completar la justificación científico-técnica final es **indispensable** rellenar el Plan de Gestión

Un Plan de Gestión de Datos 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.

La Comisión Europea proporciona orientación sobre cómo elaborar un plan de gestión de datos en sus directrices para gestionar datos según el esquema FAIR.

Se puede utilizar cualquier **plantilla o herramientas** para crear el Plan de Gestión de Datos. Se trata de un conjunto de preguntas que debe responderse con un nivel de detalle apropiado al proyecto, para garantizar que **no se pierdan los datos**, sobre todo al finalizar el proyecto de investigación, facilitar la **correcta custodia** desde su producción, permitir su preservación y validar y reutilizar los resultados obtenidos. Como recomendaciones el PGD debe incluir la siguiente información:

- **Resumen de los datos**, donde se describa de forma breve y concisa: qué datos se han generado/recogido, qué formatos y estándares se han utilizado, qué valor tienen los datos para otros investigadores, qué datos no se pueden compartir y por qué motivo.
- **Responsabilidades**, indicando las personas responsables y especificando quién determina el acceso a los datos y, en el caso de que exista propiedad intelectual, quién es el titular de los derechos del plan de seguimiento, de la gestión de los datos y de su preservación y conservación.
- **Datos FAIR (de Findable, Accessible, Interoperable y Reusable)***. Donde se describan las actuaciones realizadas a cabo para que los datos sean localizables, accesibles, interoperables y reutilizables
 - **Datos localizables**, se debe indicar que metadatos se han utilizado, especificando el identificador utilizado para los datos y si se ha utilizado algún control de versión.
 - **Accesibilidad**, se debe indicar cuándo y quién puede acceder a los datos, especificando su disponibilidad y/o los métodos o el software necesarios para acceder a los datos. Indicar si se incluye el software necesario y su documentación, especificando dónde se encuentran depositados los datos, los metadatos, la documentación y el código, y como se gestiona el acceso en los casos en los que existen restricciones.



- **Interoperabilidad**, describiendo como se facilita la interoperabilidad de los datos mediante el uso de vocabularios controlados.
- **Reutilización**, se debe especificar el grado de reutilización permitido para los datos puestos al alcance de otros investigadores y del público en general, y a partir de qué momento los datos son accesibles para ser reutilizados. Si procede, se debe explicar los motivos y el periodo durante el cual los datos están embargados. Especificar si los datos generados y/o recogidos en el proyecto son reutilizables por terceras partes. Si la reutilización de los datos es restringida, deben explicarse los motivos.

* FAIR son las siglas en inglés de Findable, Accessible, Interoperable and Reusable, y responden a la aplicación de estos cuatro principios que harán que los datos de investigación sean localizables, accesibles, interoperables y que se puedan reutilizar. Desde un punto de vista práctico, la implementación de los principios FAIR implica la aplicación de estándares de metadatos y vocabularios asociados al dominio científico de los datos, la utilización de identificadores persistentes, la correcta identificación de licencias y procedencia de los datos, así como el establecimiento de mecanismos de almacenamiento y preservación.

- **Seguridad de los datos***, indicando el tratamiento de la recuperación de los datos, así como el almacenamiento seguro y el traspaso de datos sensibles y las medidas adoptadas para la seguridad de los datos. Esto incluye el almacenamiento y las copias de seguridad, la preservación a largo plazo y los procedimientos para mantener los datos seguros y protegidos, la cobertura del contexto de la revisión ética (describir, si fuera de aplicación, los problemas éticos en la recogida, almacenamiento, procesamiento y archivo de los datos, así como en los procedimientos de aprobación ética relacionados con el proyecto).

* Si las actividades de investigación incluyen a niños, pacientes, población vulnerable, el uso de células madre embrionarias, cuestiones de privacidad y protección de datos o investigación en animales y primates, deben cumplirse los principios éticos y la legislación nacional pertinente y la de la Unión Europea y debe tratarse en este apartado del plan de gestión de datos.

**Datos de investigación.
Datos abiertos. Datos FAIR**



¿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.

¿Qué son los datos abiertos?

“Los datos de investigación abiertos son datos de **acceso gratuito** que pueden ser **reutilizados, remezclados y redistribuidos** para la investigación académica y la docencia, entre otros usos. Idealmente, los datos abiertos no tienen restricciones para su reutilización y redistribución y cuentan con licencias acordes a ello. En casos excepcionales (por ejemplo, para proteger la identidad de sujetos humanos), se pueden incluir restricciones de acceso especiales o limitadas.”



Datamarts Update Calendar Glossary API More

32 databases - 60 million records

Other UNSD Databases:



Popular statistical tables, country (area) and regional profiles

Population

- Population, surface area and density
PDF | CSV Updated: 11-May-2021
- International migrants and refugees
PDF | CSV Updated: 5-Nov-2020
- Population growth, fertility, life expectancy and mortality
PDF | CSV Updated: 20-Aug-2019
- Population in the capital city, urban and rural areas
PDF | CSV

- Afghanistan
- Albania
- Algeria
- American Samoa
- Andorra
- Angola
- Anguilla
- Antigua and Barbuda
- Argentina
- Armenia
- Aruba
- World
- Africa
- Northern Africa
- Sub-Saharan Africa
- Eastern Africa
- Middle Africa
- Southern Africa
- Western Africa
- Americas
- Northern America
- Latin America & the Caribbean



<http://data.un.org/>

data.europa.eu

El portal oficial de datos europeos

Español (es)

Search site content

Buscar

Discover the datasets from the former EU Open Data Portal [here](#).

Datos Impacto y estudios Formación Noticias y eventos Acerca de

data.europa.eu

Español



Buscar conjuntos de datos



<https://data.europa.eu/es>

datos.gob.es
reutiliza la información pública

Español



INICIATIVA APORTA CATALOGO DE DATOS IMPACTO SECTORES ACTUALIDAD INNOVACION INTERACTUA

CONJUNTOS DE DATOS

API

PUNTO SPARQL

Inicio | Catálogo de datos | Conjuntos de datos

Catálogo de datos

Categoría

- Sector público (8353)
- Sociedad y bienestar (6639)
- Medio ambiente (6388)
- Demografía (4574)
- Economía (4414)

Mostrar más

Formato

- JSON (19104)
- CSV (18252)
- HTML (15051)
- XLS (12074)
- PC-Axis (11153)

Mostrar más

Buscar conjuntos de datos...

BUSCAR

41,137 conjuntos de datos encontrados

Ordenar por resultados: Publicado Descendente

Puntos de recarga eléctrica para vehículos

Publicador: Ayuntamiento de Alcobendas

Ayuntamiento del Alcobendas. Ubicación de los puntos de recarga para vehículos eléctricos.

XLS CSV

Registro de contratos de Ayuntamiento de Asparrena del 2021

Publicador: Comunidad Autónoma de País Vasco

El Registro de Contratos del Sector Público constituye el sistema oficial central de información sobre la contratación pública y, como tal, el soporte para el conocimiento...

XLSX

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

El portal oficial de datos europeos

170 Catálogos

36 Países

Trending datasets

- Lista consolidada de entidades sujetas a la UE
- Número de personas por edad
- CORDIS — Proyecto UE en el marco de H 2020)
- Subastas Diario Elec (subconjunto csv) — contratación pública
- Principalii indicatori

data.europa academy

Filtrar por ubicación

covid

Conjuntos de datos Catálogos Editorial Content

51 conjuntos de datos hallados para

Categories: Salud

Datos de COVID por región

Datos día a día y todos los días relativos a la infección por COVID-19 (positivos recuperados, hospitalizados, cuidados intensivos) para los países europeos, divididos por sus regiones, cuando estén disponibles

Actualizado: 28.02.2020 Creado: 28.02.2020

Datos de COVID en todo el mundo

Datos día a día y todos los días relacionados con la infección por COVID (positivos recuperados, hospitalizados, cuidados intensivos) para los países del mundo proceden de agencias nacionales o del ECDC

Actualizado: 01.01.2020 Creado: 01.01.2020

Datos del coronavirus de la enfermedad de COVID

Este conjunto de datos contiene los últimos datos públicos disponibles sobre el brote de COVID-19, incluida una actualización diaria de la situación, la curva epidemiológica y la distribución geográfica mundial (UE/EEE y Reino Unido, y en todo el mundo). El 12 de febrero de 2020, el nuevo coronavirus se denominó coronavirus del síndrome respiratorio agudo...

Creado: 01.02.2020

European Centre for Disease Prevention and Control

Ensayos clínicos sobre la COVID-19 en el Espacio Económico Europeo (EEE)

European Centre for Disease Prevention and Control

Editor: European Centre for Disease Prevention and Control

Actualizado:

La información sobre los ensayos clínicos en curso relacionados con la COVID-19 en el Espacio Económico Europeo (EEE) está disponible en el registro de ensayos clínicos de la UE. Los usuarios pueden encontrar información sobre:

- ensayos clínicos en adultos de la fase II a la fase IV en los que el centro de investigación se encuentre en el EEE;
- ensayos clínicos en niños en los que el centro de investigación se encuentre en la UE y cualquier ensayo que forme parte de un plan de investigación pediátrica, incluidos aquellos en los que el centro investigador se encuentre fuera de la UE. Los ensayos clínicos de la fase I realizados únicamente en adultos no se ponen a disposición del público en el registro de ensayos clínicos de la UE.

Más información sobre el registro de ensayos clínicos de la UE, incluidas la fuente de información y la base jurídica.

[Download All](#)

Distribuciones (4)

Datos sobre la vacunación contra la COVID-19 en la UE/EEE

Fichero de datos descargable que contiene información sobre la vacunación contra la COVID-19 en la UE/EEE.

Opciones Descargar Linked Data

Datos sobre la vacunación contra la COVID-19 en la UE/EEE

Fichero de datos descargable que contiene información sobre la vacunación contra la COVID-19 en la UE/EEE.

Licencia Creative Commons Attribution 4.0 International

Datos sobre la vacunación contra la COVID-19 en la UE/EEE

Fichero de datos descargable que contiene información sobre la vacunación contra la COVID-19 en la UE/EEE.

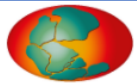
Licencia Creative Commons Attribution 4.0 International

Datos sobre la vacunación contra la COVID-19 en la UE/EEE

Fichero de datos descargable que contiene información sobre la vacunación contra la COVID-19 en la UE/EEE.

Licencia Creative Commons Attribution 4.0 International

YearWeek	Country	Denominator	Number Doses Received	Number Doses Exported	First Dose	First Dose Refused	Second Dose	Dose Additional	Unknown Dose	Region	TargetGroup	Vaccine	Population
1	2020-W53	AT	8901064	61425	0	0	1	0	0	AT	ALL	UNK	8901064
2	2020-W53	AT	8901064	0	0	0	0	0	0	AT	ALL	UNK	8901064
3	2020-W53	AT	8901064	0	0	0	0	0	0	AT	ALL	AZ	8901064
4	2020-W53	AT	8901064	0	0	0	0	0	0	AT	ALL	MOD	8901064
5	2020-W53	AT	8901064	0	0	0	0	0	0	AT	ALL	JANSS	8901064
6	2020-W53	AT	435835	61425	0	1	0	0	0	AT	Age0_4	COM	8901064
7	2020-W53	AT	435835	0	0	0	0	0	0	AT	Age0_4	JANSS	8901064
8	2020-W53	AT	435835	0	0	0	0	0	0	AT	Age0_4	AZ	8901064
9	2020-W53	AT	435835	0	0	0	0	0	0	AT	Age0_4	UNK	8901064
10	2020-W53	AT	435835	0	0	0	0	0	0	AT	Age0_4	MOD	8901064
11	2020-W53	AT	435835	0	0	0	0	0	0	AT	Age0_4	MOD	8901064
12	2020-W53	AT	421998	0	0	0	0	0	0	AT	Age10_14	MOD	8901064
13	2020-W53	AT	421998	0	0	0	0	0	0	AT	Age10_14	UNK	8901064
14	2020-W53	AT	421998	61425	0	0	0	0	0	AT	Age10_14	COM	8901064
15	2020-W53	AT	421998	0	0	0	0	0	0	AT	Age10_14	AZ	8901064



PANGAEA.

Data Publisher for Earth & Environmental Science

SEARCH SUBMIT HELP ABOUT CONTACT

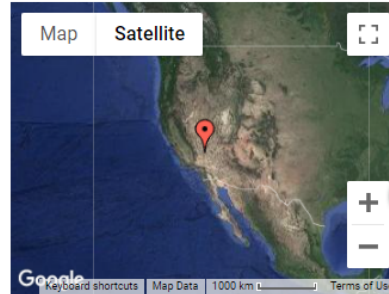
Citation:

Lasser, Jana; Goehring, Lucas (2020): Grain size distributions of sand samples from Badwater Basin site P5 in central California, collected in 2016. PANGAEA, doi: <https://doi.org/10.1594/PANGAEA.910992>,

In: Lasser, J; Goehring, L (2020): Grain size distributions of sand samples from Owens Lake and Badwater Basin in central California, collected in 2016 and 2018. *Max Planck Institute for Dynamics and Self-Organization*, PANGAEA, doi: <https://doi.org/10.1594/PANGAEA.910996>

Always quote citation above when using data! You can download the citation in several formats below.

RIS Citation BibTeX Citation Copy Citation Facebook Twitter Show Map Google Earth



Related to:

Lasser, Jana (2019): Geophysical Pattern Formation of Salt Playa (PhD thesis). *Georg-August Universität Göttingen*, hdl:11858/00-1735-0000-002E-E5DB-2 [Q](#)

Lasser, Jana; Nield, Joanna M; Ernst, Marcel; Karius, Volker; Wiggs, Giles F S; Goehring, Lucas (2019): Salt Polygons are Caused by Convection. *Nonlinear Sciences > Pattern Formation and Solitons*, <https://arxiv.org/abs/1902.03600> [Q](#)

Coverage:

Latitude: 36.227570 * Longitude: -116.783930

Minimum DEPTH, sediment/rock: 0.18 m * Maximum DEPTH, sediment/rock: 0.35 m

Event(s):

Badwater_basin_P5 [Q](#) * Latitude: 36.227570 * Longitude: -116.783930 * Location: California, USA [Q](#) * Method/Device: Sampling by hand (HAND) [Q](#)

Parameter(s):

#	Name	Short Name	Unit	Principal Investigator	Method/Device	Comment
1	DEPTH, sediment/rock Q	Depth	m	Lasser, Jana Q		Geocode
2	Percentile 10 Q	Perc10	µm	Lasser, Jana Q	Laser Particle Sizer (LPS) (LPSIZ) Q	
3	Median, grain size Q	D50	µm	Lasser, Jana Q	Laser Particle Sizer (LPS) (LPSIZ) Q	
4	Percentile 90 Q	Perc90	µm	Lasser, Jana Q	Laser Particle Sizer (LPS) (LPSIZ) Q	
5	Soil composition Q	Soil comp		Lasser, Jana Q		
6	Soil water content Q	SWC	%	Lasser, Jana Q		wt%
7	Diameter Q	Ø	µm	Lasser, Jana Q	Laser Particle Sizer (LPS) (LPSIZ) Q	lower channel diameter
8	Difference Q	Diff	%	Lasser, Jana Q	Laser Particle Sizer (LPS) (LPSIZ) Q	volume
9	Difference Q	Diff	%	Lasser, Jana Q	Laser Particle Sizer (LPS) (LPSIZ) Q	in volume, -2SD
10	Difference Q	Diff	%	Lasser, Jana Q	Laser Particle Sizer (LPS) (LPSIZ) Q	in volume, +2SD
11	Diameter Q	Ø	µm	Lasser, Jana Q	Laser Particle Sizer (LPS) (LPSIZ) Q	middle channel diameter

Datos de investigación abiertos

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

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



¿Qué significa FAIR?

RESEARCH DATA - OPEN BY DEFAULT



'Localizables'

se puede encontrar a través de metadatos, y son identificables y localizables mediante un mecanismo de identificación estándar.

'Accesibles'

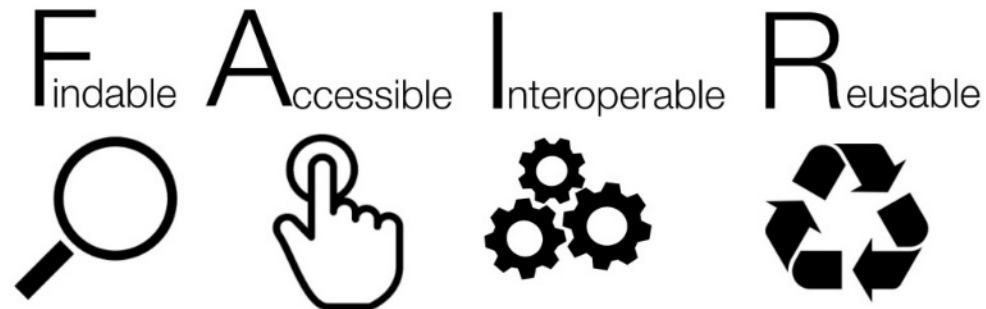
siempre disponibles; incluso si los datos están restringidos, los metadatos están abiertos.

'Interoperables'

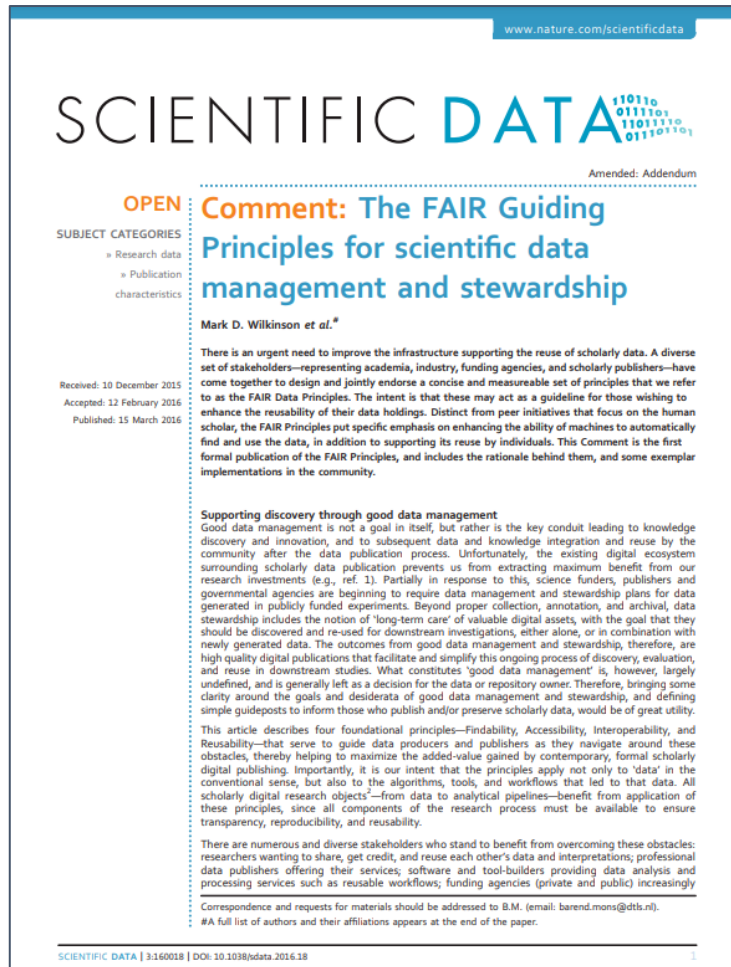
sintácticamente analizables y semánticamente comprensibles, lo que permite el intercambio y la reutilización de datos entre investigadores, instituciones, organizaciones o países.

'Reutilizables'

suficientemente descritos y compartidos con licencias mínimamente restrictivas, para permitir así una reutilización lo más amplia posible.



¿Qué son los datos FAIR?



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

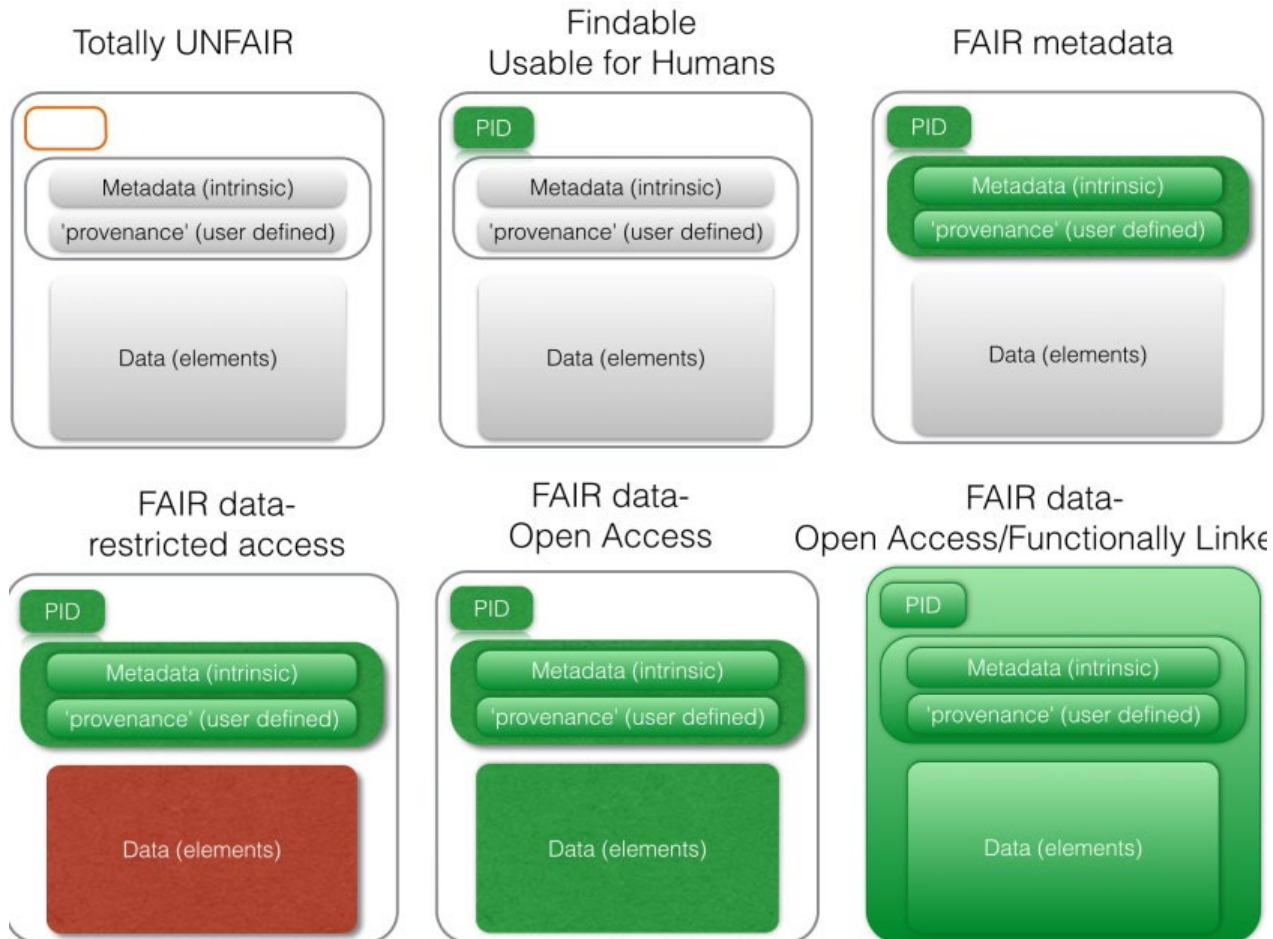
Los principios enfatizan la capacidad de los sistemas computacionales para encontrar, acceder, interoperar y reutilizar datos con ninguna o mínima intervención humana).

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>

- F1. A los (meta)datos se les debe asignar un identificador único y persistente.
- F2. Los datos se describen con metadatos enriquecidos.
- F3. (meta)datos se registran o indexan en un sistema que disponga de motor de búsqueda.
- F4. Los metadatos contienen un campo específico para el identificador del dataset.
- A1. Los (meta)datos se pueden recuperar mediante su identificador utilizando un protocolo de comunicación estandarizado.
 - A1.1 El protocolo es abierto, gratuito y universalmente aplicable.
 - A1.2 El protocolo permite un procedimiento de autenticación y autorización, cuando sea necesario.
- A2. Los metadatos son accesibles, incluso cuando los datos ya no están disponibles.
- I1. Los (meta)datos utilizan un lenguaje formal, accesible, compartido y ampliamente aplicable para la representación del conocimiento.
- I2. Los (meta)datos usan vocabularios que siguen principios FAIR.
- I3. Los (meta)datos incluyen referencias calificadas a otros (meta)datos.
- R1. Los (meta)datos tienen disponen de atributos precisos y pertinentes.
 - R1.1. Los (meta)datos se publican con una licencia de reutilización clara y accesible.
 - R1.2. Los (meta)datos están asociados con su procedencia.
 - R1.3. Los (meta)datos cumplen con los estándares propios de una área de conocimiento determinada.

<https://force11.org/info/the-fair-data-principles/>

Data as increasingly FAIR Digital Objects



Nivel 1: Cada objeto de datos tiene un PID y metadatos FAIR intrínsecos (en esencia, "estáticos").

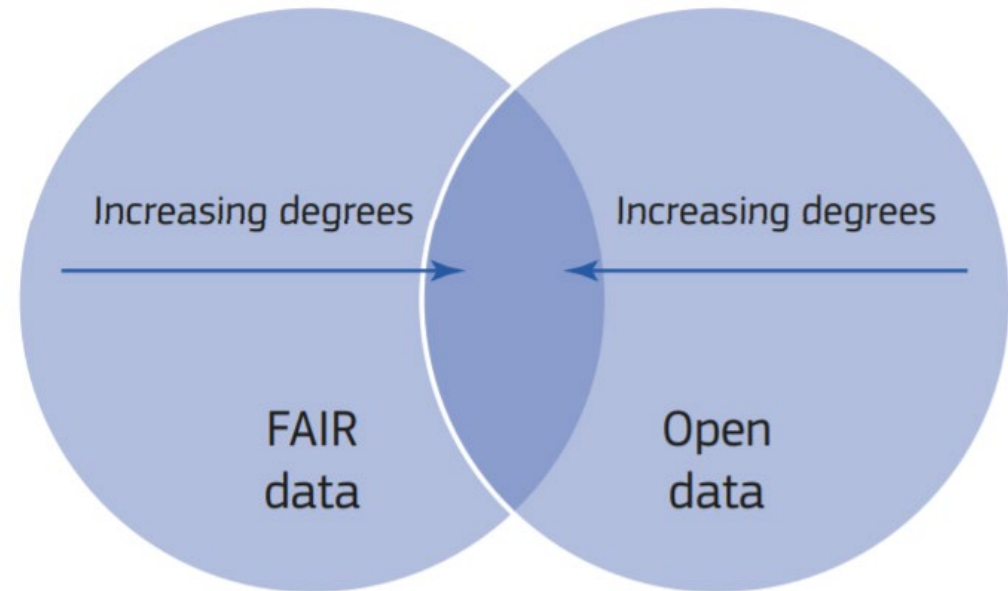
Nivel 2: Cada objeto de datos tiene metadatos "definidos por el usuario" (y actualizados) para proporcionar una procedencia rica en formato FAIR de los datos, qué ha pasado con ellos, para qué se han utilizado, para qué pueden utilizarse, etc..

Nivel 3. Los propios datos son "técnicamente" también FAIR, pero no totalmente de libre acceso y no reutilizables sin restricciones (por ejemplo, datos de pacientes o datos protegidos).

Nivel 4: Los metadatos, así como los propios datos, son totalmente FAIR y completamente públicos, bajo una licencia bien definida. (Los datos sin licencia considerados "públicos" por su propietario seguirán estando excluidos de los proyectos de integración por parte, por ejemplo, de las empresas farmacéuticas).



Open ≠ FAIR



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

AS OPEN AS POSSIBLE, AS CLOSED AS NECESSARY

Grantees have the right to opt-out, but need to say **why**



Top three reasons for opt-out:

privacy

intellectual
property rights

might jeopardise
project's main
objective

**Tan abierto como sea posible,
tan cerrado como sea
necesario**



**¿Qué herramientas podemos usar
para hacer que nuestro datos
cumplan con los principios FAIR?**

Australian Research Data Commons' [FAIR](https://ardc.edu.au/resources/aboutdata/fair-data/fair-self-assessment-tool/) data self assessment tool

<https://ardc.edu.au/resources/aboutdata/fair-data/fair-self-assessment-tool/>

The screenshot shows the landing page of the FAIR Data Self Assessment Tool. At the top is the ARDC logo and navigation links. The main heading is "FAIR Data Self Assessment Tool". Below it, a description states: "Use our FAIR data self assessment tool to assess how FAIR your research dataset is and get practical tips on how to enhance its FAIRness." A subtext explains: "This handy tool helps you assess the FAIRness of a dataset and determine how to enhance its FAIRness (where applicable)." There is a section for "FAQs" with three questions: "How does the FAIR data self assessment tool work?", "Who can use the FAIR data self assessment tool?", and "What are the FAIR principles?". A "Last updated" box shows "12 May 2022". A "Research Topic" box lists "Tools and Services" and "All". A "Share & Print" box contains social media icons. At the bottom, a progress bar shows "Total across FAIR" with a green bar indicating progress, and a row of 12 questions (Q1-Q12) with Q1-Q12 marked as "Answered".



The screenshot shows the "Findable" section of the assessment tool. It starts with a definition: "The data has sufficiently rich metadata and a unique and persistent Identifier to be easily discovered by others. This includes assigning a persistent Identifier (like a DOI or Handle), having rich metadata to describe the data and making sure it is findable through disciplinary local or international discovery portals." The section contains three questions, each with a green checkmark icon and a "What is this?" link:

- Does the dataset have any identifiers assigned?**
 - ☒ Globally unique, citable, and persistent (e.g. DOI, PURL, ARK or Handle)
 - ☐ Web Address (URL)
 - ☐ Local Identifier
 - ☐ No Identifier
- Is the dataset identifier included in all metadata records/files describing the data?**
 - ☒ Yes
 - ☐ No
- How is the data described with metadata?**
 - ☒ Comprehensively using a formal machine-readable metadata schema
 - ☐ Comprehensively, but in a text-based, non-standard format
 - ☐ Brief title and description
 - ☐ The data is not described

Below these questions is another question:

- What type of repository or registry is the metadata record in?**
 - ☐ Data is in one place but discoverable through several registries
 - ☐ Generalist public repository
 - ☒ Domain-specific repository
 - ☐ Local institutional repository
 - ☐ The data is not described in any repository

At the bottom of the section is a green progress bar.

F-UJI tool to assess FAIRness of datasets

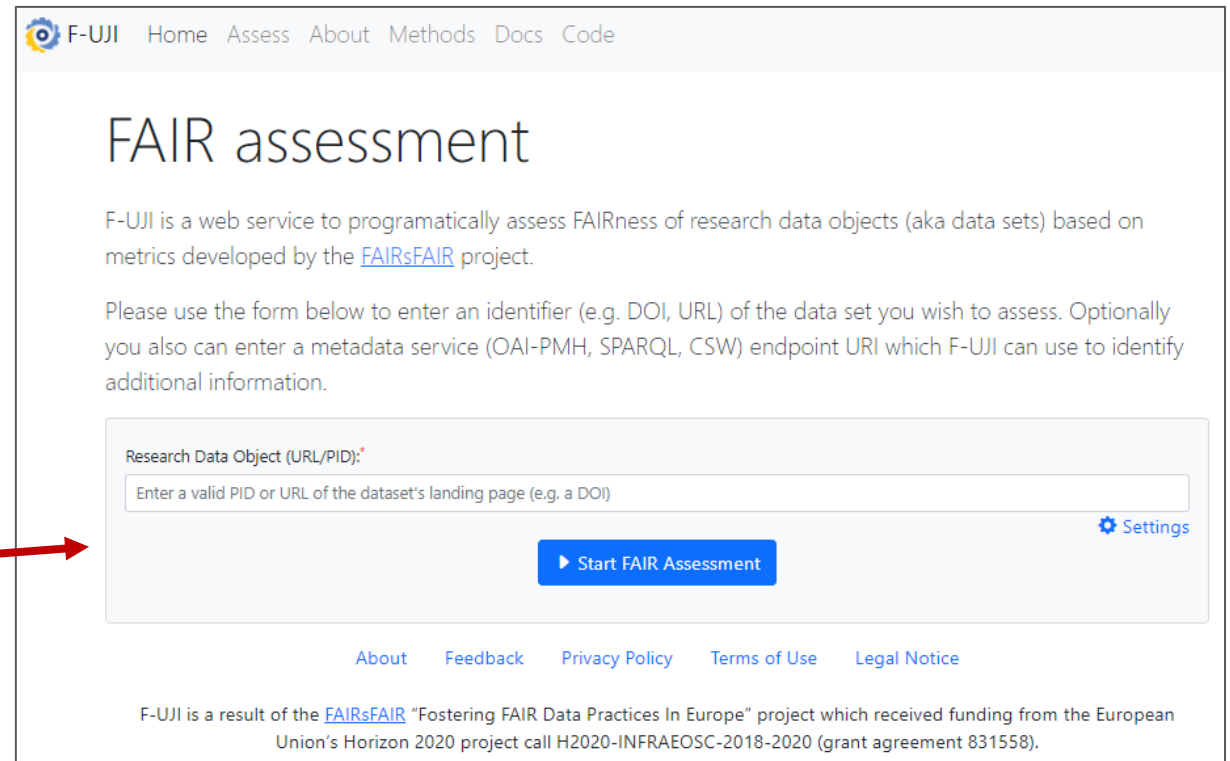
Un servicio web para evaluar el grado FAIR de los objetos de datos de investigación a nivel de conjunto de datos.



F-UJI | Automated FAIR Data Assessment Tool

F-UJI is a web service to programatically assess FAIRness of research data objects at the dataset level based on the FAIRsFAIR Data Object Assessment Metrics [↗](#)

[Click here to assess a dataset](#)



F-UJI Home Assess About Methods Docs Code

FAIR assessment

F-UJI is a web service to programatically assess FAIRness of research data objects (aka data sets) based on metrics developed by the [FAIRsFAIR](#) project.

Please use the form below to enter an identifier (e.g. DOI, URL) of the data set you wish to assess. Optionally you also can enter a metadata service (OAI-PMH, SPARQL, CSW) endpoint URI which F-UJI can use to identify additional information.

Research Data Object (URL/PID):*

[Settings](#)

[▶ Start FAIR Assessment](#)

[About](#) [Feedback](#) [Privacy Policy](#) [Terms of Use](#) [Legal Notice](#)

F-UJI is a result of the [FAIRsFAIR](#) "Fostering FAIR Data Practices In Europe" project which received funding from the European Union's Horizon 2020 project call H2020-INFRAEOSC-2018-2020 (grant agreement 831558).

Data supporting the "Health Libraries Sharing Game".

[10.5281/zenodo.3360451](https://doi.org/10.5281/zenodo.3360451)

Assessment Results:

Evaluated Resource:

Data supporting the "Health Libraries Sharing Game"

✓ Save

↓ (JSON)

📄 New

FAIR level: ?

advanced

Resource PID/URL: 10.5281/zenodo.3360451

DataCite support: enabled

Metric Version: metrics_v0.4

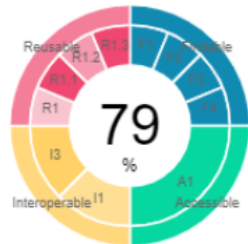
Metric Specification: <https://doi.org/10.5281/zenodo.4081213>

Software version: 1.4.3

Download assessment results: [\(JSON\)](#)

Save and share assessment results:

Summary:



	Score earned:		Fair level:
Findable:	7 of 7	🟢	advanced
Accessible:	3 of 3	🟢	advanced
Interoperable:	3 of 4	🟡	moderate
Reusable:	6 of 10	🟡	moderate

Report:

Findable

FsF-F1-01D - Data is assigned a globally unique identifier.



FsF-F1-02D - Data is assigned a persistent identifier.



FsF-F2-01M - Metadata includes descriptive core elements (creator, title, data identifier, publisher, publication date, summary and keywords) to support data findability.



FsF-F3-01M - Metadata includes the identifier of the data it describes.



FsF-F4-01M - Metadata is offered in such a way that it can be retrieved programmatically.



Accessible

FsF-A1-01M - Metadata contains access level and access conditions of the data.



FsF-A1-03D - Data is accessible through a standardized communication protocol.



FsF-A1-02M - Metadata is accessible through a standardized communication protocol.



Interoperable

FsF-I1-01M - Metadata is represented using a formal knowledge representation language.



FsF-I1-02M - Metadata uses semantic resources



FsF-I3-01M - Metadata includes links between the data and its related entities.



Los planes de gestión de datos: ¿Qué son?

Recursos.

QUÉ ES UN PLAN DE GESTIÓN DE DATOS O PGD (DMP)

- 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.
- Normalmente 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, Accessible, Interoperable and Reusable), es decir, localizables, accesibles, interoperables y reutilizables.
- El PGD es un elemento requerido cada vez por más organismos financiadores. A menudo se pide un resumen en la propuesta del proyecto.

CUÁNDO HAY QUE ENTREGAR VERSIONES DEL PGD

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**.

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 Times New Roman, Calibri o Arial de un tamaño mínimo de 11 puntos; márgenes laterales de 2,5 cm; márgenes superior e inferior de 2,5 cm; interlineado mínimo sencillo.**
3. Las memorias pueden rellenarse en español o en inglés, pero hacerlo en inglés cuando la cantidad económica solicitada en el subproyecto es igual o superior a 100.000,00 € (artículo 14.2.a) de la convocatoria.
4. Se recomienda rellenar la memoria empleando un ordenador con Windows y usando como procesador de textos MS Word (MS Office).
5. Una vez terminada la memoria en Word, deberá convertir el archivo a PDF (de no más de 4Mb) y aportarlo en la aplicación informática de solicitud en el apartado Añadir documentos > Memoria científico-técnica. Verifique que ha aportado el documento correcto y que no se supera la extensión.

Recuerde que:

- Debe rellenar el título del proyecto coordinado y los datos de los subproyectos.
- Los proyectos coordinados están compuestos por un mínimo de 2 subproyectos.
- El subproyecto 1 será liderado por el/la investigador/a principal o investigador/a principal de la investigación, que será responsable de su seguimiento y de su coordinación. Si el subproyecto coordinador tiene dos investigadores/as principales, ambos serán considerados/as investigadores/as coordinadores/as.
- Los proyectos tipo RTA solo podrán formar parte de proyectos coordinados si todos los subproyectos, incluido el subproyecto coordinador, sean de esta modalidad.
- Los proyectos tipo JIN no pueden formar parte de un proyecto coordinado.

TÍTULO DEL PROYECTO COORDINADO

Tal y como se indica en el artículo 14.1.d) de la convocatoria, en el caso de los proyectos 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.

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.

QUÉ HAY QUE INCLUIR EN EL PGD

- ¿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?
- ¿Cómo se van a almacenar los datos, cómo se van a compartir y preservar?
- ¿Qué aspectos éticos se van a considerar?
- ¿Quién se va a encargar de la gestión de los datos?

Información mínima requerida en un PGD

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

Plantillas: Template for the Horizon 2020 Data Management Plan

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

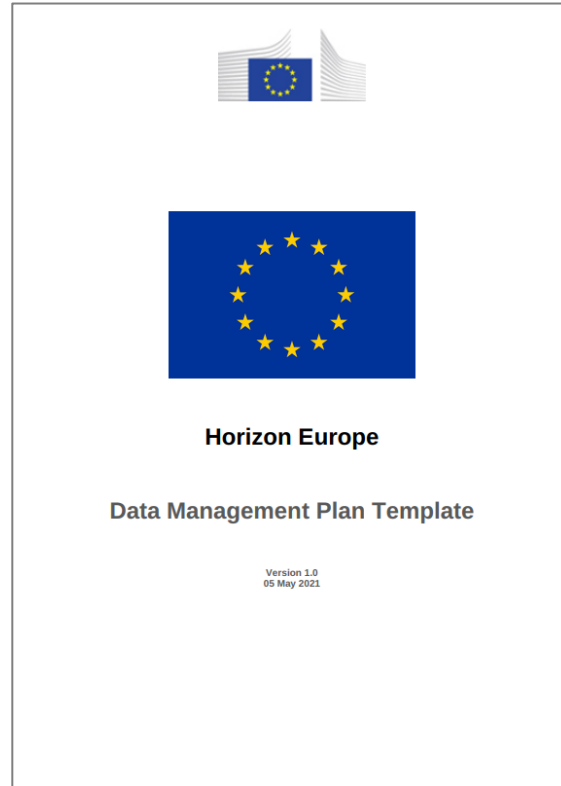
Are the data produced in the project interoperable, that is allowing data exchange and re-use between researchers, institutions, organisations, countries, etc. (i.e. adhering to standards for formats, as much as possible compliant with available (open) software applications, and in particular facilitating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

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

Plantillas: Template for the Horizon Europe Data Management Plan



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 existing data has been 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 general standards will be followed? In case metadata standards do not exist in your discipline, please outline what type of metadata will be created and how.

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

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 digital object?

Data:

Will all data be made openly available? If certain datasets cannot be shared (or need to be shared under restricted access conditions), explain why, clearly separating legal and contractual reasons from intentional restrictions. Note that in multi-beneficiary projects it is also possible for specific beneficiaries to keep their data closed if opening their data goes against their legitimate 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), specify why and how 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 project?

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 Agreement? If not, please clarify why. Will metadata contain information to enable the user to access the data?

4

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 references¹ 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?

¹A 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 uses 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/3-metadata-include-qualified-references-metadata/>)

5

ERC DMP Template

For ERC grantees, “as from 2021 it is no longer possible for applicants to opt out of the submission of Research Data Management plans.” ((European Research Council, 2021, p. 4)) In order to receive funding at European level, a data management plan has to be written.



ERC OPEN RESEARCH DATA MANAGEMENT PLAN (DMP)

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?))

Page 3 of 5

ERC OPEN RESEARCH DATA MANAGEMENT PLAN (DMP)

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)

Page 4 of 5

ERC OPEN RESEARCH DATA MANAGEMENT PLAN (DMP)

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)

DISCLAIMER. Please note that the ERC Data Management Plan is not a part of the Ethics Review. It is the responsibility of the Principal Investigator to inform the ERCEA Ethics Team of any ethics issues/concerns regarding the collection, processing, sharing and storage of data in relation to the project.

Page 5 of 5

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 data¹ and similar materials. In these cases, a short explanation is required (see [DMP template \(pdf, 617KB\)](#) / [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](mailto: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](#)

[About](#)
[Programmes and Events](#)

Resources

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 for a climate model. 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 other researchers, attending conferences and applying for funding. Together, the demands of daily research practice can all conspire to prevent proper planning for the development of research software.


A Software Management Plan (SMP) can help you to define a set of structures and goals to understand your research software including what you are 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 ecosystem; 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 project that is already underway. An SMP provides a way to draw together and summarise research software-related aspects that have already been decided and 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. Within each section questions are posed to help you to complete that section. Complementary guidance and links to other resources are also provided. Not all questions are relevant to all projects and the extent to which a specific question can be answered may depend both upon the nature of your research software and its current state of development. Consider your SMP as a living document, to be reviewed and revised as the development of your 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](https://doi.org/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 by GitHub, GitLab and BitBucket, for example).



[Upload](#)
[Communities](#)

[fagomsan@yahoo.com](#)

December 10, 2018

Working paper
Open Access

Checklist for a Software Management Plan

The Software Sustainability Institute

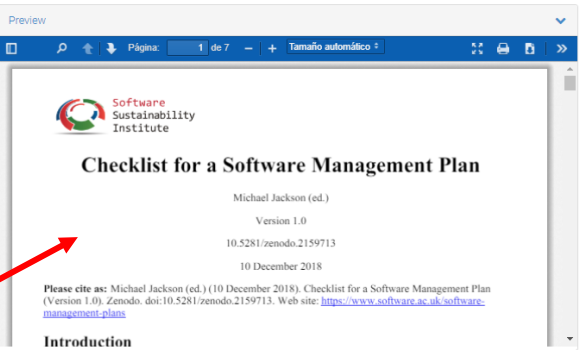
When developing research software, it is easy to focus only on goals and activities such as collaborating with other researchers, writing papers, attending conferences and applying for funding. Together, the demands of daily research practice can all conspire to prevent proper planning for the development of research software.

A Software Management Plan (SMP) can help you to define a set of structures and goals to understand your research software including what you are 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 ecosystem; and how you will ensure that your software remains available beyond the lifetime of your current project.

This checklist will help you to write an SMP. It consists of sections that cover the key elements that an SMP should include.

The Software Sustainability Institute is supported by EPSRC grant EP/H043160/1 and EPSRC/BBSRC and ESRC grant EP/N006410/1.


Preview




1,973
views

1,678
downloads

[See more details...](#)

Indexed in



Publication date:
December 10, 2018
DOI:
[DOI: 10.5281/zenodo.2159713](https://doi.org/10.5281/zenodo.2159713)
Keyword(s):
research software software sustainability software sustainability institute software management plan
Related identifiers:
Cited by
<https://www.software.ac.uk/software-management-plans>
Compiled by
<https://github.com/software-saved/software-management-plans/releases/tag/v1.0>
10.5281/zenodo.2159748
References
<https://www.software.ac.uk/software-management-plans>
License (for files):
 Creative Commons Attribution 4.0 International
Versions

Version 1.0	Dec 10, 2018
10.5281/zenodo.2159713	

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

DMP online


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

DMP ONLINE Home Public DMPs Funder requirements Help  Language ▾

✓ Notice: Signed out successfully.

Plan to make data work for you

Data Management Plans that meet institutional funder requirements.



DMPonline helps you to create, review, and share data management plans that meet institutional and funder requirements. It is provided by the Digital Curation Centre (DCC).

Sign in Create account

* **First Name**

* **Last Name**

* **Email**

Organisation


My organisation isn't listed.


* **Password**


☐ Show password


* ☐ I accept the terms and conditions

Create account


59,972 Users












314 Organisations


65,212 Plans


89 Countries

- Es una herramienta en línea que se lanzó en 2010 para crear, mantener y exportar DMPs, desarrollada de manera conjunta por el Digital Curation Centre (DCC) y el Curation Center de la Universidad de California (UC3).
- Se creó 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.
- 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>.

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.					
Q 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 of the language used, but using any of the plan's text 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 Consent 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 any 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 gathering informed consent, securely storing data, ensuring data access and future through data archiving and sharing, and producing and sharing relevant metadata.

Project Details
Contributors
Plan overview
Write Plan
Share
Download

expand all | collapse all
0/13

Data Collection (0 / 2)
Documentation and Metadata (0 / 1)
Ethics and Legal Compliance (0 / 2)
Storage and Backup (0 / 2)
Selection and Preservation (0 / 2)
Data Sharing (0 / 2)
Responsibilities and Resources (0 / 2)

Who will be responsible for data management?

B
I
List
List
Link
Table

Save

Guidance	Comments
DCC <p>Questions to consider:</p> <ul style="list-style-type: none"> Who is responsible for implementing the DMP, and ensuring it is reviewed and revised? Who will be responsible for each data management activity? How will responsibilities be split across partner sites in collaborative research projects? Will data ownership and responsibilities for RDM be part of any consortium agreement or contract agreed between partners? <p>Guidance:</p> <p>Outline the roles and responsibilities for all activities e.g. data capture, metadata production, data quality, storage and backup, data archiving & data sharing. Consider who will be responsible for ensuring relevant policies will be respected. Individuals should be named where possible.</p>	

How will you manage any ethical issues?

B *I*
☰
☰
🔗
📄

Save

Guidance

Comments

DCC

Questions to consider:

- Have you gained consent for data preservation and sharing?
- How will you protect the identity of participants if required? e.g. via anonymisation
- How will sensitive data be handled to ensure it is stored and transferred securely?

Guidance:

Ethical issues affect how you store data, who can see/use it and how long it is kept. Managing ethical concerns may include: anonymisation of data; referral to departmental or institutional ethics committees; and formal consent agreements. You should show that you are aware of any issues and have planned accordingly. If you are carrying out research involving human participants, you must also ensure that consent is requested to allow data to be shared and reused.

Ethics & privacy

- Investigators carrying out research involving human participants should request consent to preserve and share the data. Do not just ask for permission to use the data in your study or make unnecessary promises to delete it at the end.
- Consider how you will protect the identity of participants, e.g., via anonymisation or using managed access procedures.
- Ethical issues may affect how you store and transfer data, who can see/use it and how long it is kept. You should demonstrate that you are aware of this and have planned accordingly.
- See UK Data Service guidance on [consent for data sharing](#).
- See ICPSR approach to [confidentiality](#) and Health Insurance Portability and Accountability Act (HIPAA) [regulations for health research](#).

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

B	<i>I</i>				

Save

Intellectual Property Rights

- State who will own the copyright and IPR of any existing data as well as new data that you will generate. For multi-partner projects, IPR ownership should be covered in the consortium agreement.
- Outline any restrictions needed on data sharing, e.g., to protect proprietary or patentable data.
- Explain how the data will be licensed for reuse. See the DCC guide on [How to license research data](#) and EUDAT's [data and software licensing wizard](#).

Guidance

Comments

DCC

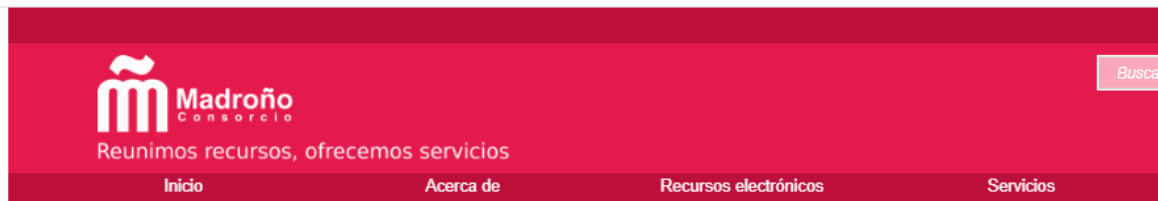
Questions to consider:

- Who owns the data?
- How will the data be licensed for reuse?
- Are there any restrictions on the reuse of third-party data?
- Will data sharing be postponed / restricted e.g. to publish or seek patents?

Guidance:

State who will own the copyright and IPR of any data that you will collect or create, along with the licence(s) for its use and reuse. For multi-partner projects, IPR ownership may be worth covering in a consortium agreement. Consider any relevant funder, institutional, departmental or group policies on copyright or IPR. Also consider permissions to reuse third-party data and any restrictions needed on data sharing.

PGDonline



[🏠](#) > [InvestigaM – Ciencia Abierta](#) > [PaGoDa](#)

INVESTIGAM: PAGODA

Un Plan de Gestión de Datos es el documento que describe cómo se van a recoger, organizar, analizar, conservar y compartir investigación recopilados o generados en el curso de un proyecto de investigación. Garantiza el cumplimiento de los requisitos

El plan de gestión de datos debe elaborarse al principio del proyecto de investigación para que la gestión de datos sea más eficiente.

En el Programa Marco de Investigación e Innovación de la Unión Europea [Horizonte 2020](#) la elaboración del plan de gestión de Datos de Investigación en Acceso Abierto, cuyo objetivo era garantizar el acceso y la reutilización de los datos generados en los participantes.

En el actual [Programa Horizonte Europa \(2021-2027\)](#), es obligatorio elaborar el plan de gestión de datos para todos los proyectos que **reutilicen datos**, dentro de los 6 primeros meses del proyecto, con el fin de:

- Realizar una gestión responsable de los datos de la investigación de acuerdo con los principios FAIR de "localización", "accesibilidad", "interoperabilidad" y "reutilización".
- Garantizar el acceso abierto a los datos de investigación, según el principio "tan abierto como sea posible, tan cerrado como sea necesario" depositando los datos en un repositorio de confianza, bajo una licencia Creative Commons (CC BY), de dominio público. Si no se proporciona acceso abierto (a algunos o a todos los datos), deberá justificarse en el plan de gestión de datos ([Horizonte Europa Grant Agreement](#)).

<http://www.consorcio-madrono.es/investigam/pagoda/>

CREAR EL PLAN

El Consorcio Madroño ofrece la herramienta [PGDonline](#), (adaptada y traducida al español de la herramienta Argos, desarrollada por OpenAIRE), que ayuda a redactar, compartir y exportar un plan de gestión de datos.

La elaboración de un plan de gestión de datos ahorrará mucho tiempo y recursos durante el transcurso del proyecto. Se deberá actualizar y ajustar periódicamente, de acuerdo con el progreso del proyecto.

El plan de gestión de datos se debe describir cómo:

- Organizar y describir los datos:** qué tipos de datos, formatos, normas y métodos de captura se utilizarán.
- Gestionar los requisitos legales y éticos:** cómo se protegerán los datos sensibles y si se restringirá su acceso y uso.
- Almacenar los datos:** cómo se almacenarán los datos a corto y largo plazo.
- Compartir los datos:** cómo se compartirán los datos y qué posibilidades de reutilización existen.
- Gestionar los recursos de personal y la formación:** quién será el responsable de la gestión de los datos, qué recursos necesita y si necesita formación.

Fuente: https://library.leeds.ac.uk/info/14062/research_data_management/62/data_management_planning.

En las plantillas utilizadas para elaborar el plan de gestión de datos se deben explicar los siguientes aspectos:

- Resumen de los datos:** se debe describir de forma general los datos que se utilizarán en el proyecto. Incluye el tipo y el formato, la finalidad, el tamaño y el origen de los datos.
- Datos FAIR:** se debe asegurar que los datos sean:
 - Localizables:** se incluyen los identificadores, palabras clave, normas de metadatos que los describen para optimizar su localización.
 - Accesibles:** se indica el repositorio donde se depositan los datos y los metadatos en acceso abierto y se justifica si hay datos cerrados.
 - Interoperables:** se exponen los vocabularios de datos, las normas, formatos o metodologías que se utilizarán para permitir el intercambio y la interoperabilidad de los datos.
 - Reutilizables:** se documenta la procedencia de los datos y se proporciona la información necesaria para validar los datos y facilitar su reutilización, ejemplo con los archivos [readme](#). Se indican las licencias de uso de los datos.
- Otros resultados de la investigación:** se consideran qué cuestiones relativas a los datos FAIR pueden aplicarse a la gestión de otros productos de la investigación generados o reutilizados en el proyecto como por ejemplo: softwares, flujos de trabajo, protocolos, nuevos materiales, muestras, etc.
- Asignación de recursos:** se indican cuáles serán los costes para hacer que los datos sean FAIR (por ejemplo, costes directos e indirectos relacionados con el almacenamiento, el archivo, la reutilización, la seguridad, etc.) y quién será responsable de la gestión de los datos.
- Seguridad de los datos:** se garantiza que los datos se almacenan de forma segura en repositorios de confianza para su preservación y conservación a largo plazo.
- Ética:** se presentan todas las cuestiones éticas o legales que puedan tener un impacto en el intercambio de datos. Además, cuando la investigación utilice datos personales, deberá hacerse referencia a aspectos como el consentimiento informado y la conservación a largo plazo.
- Otras cuestiones:** se describe si se van a utilizar, otros procedimientos (nacionales, de proveedores, sectoriales, departamentales) para la gestión de datos.

Plantillas para elaborar el plan de datos:

[Horizon Europe Data Management Plan Template](#)

[Horizon 2020 Template for the Data Management Plan](#)

Ejemplos de planes de datos:

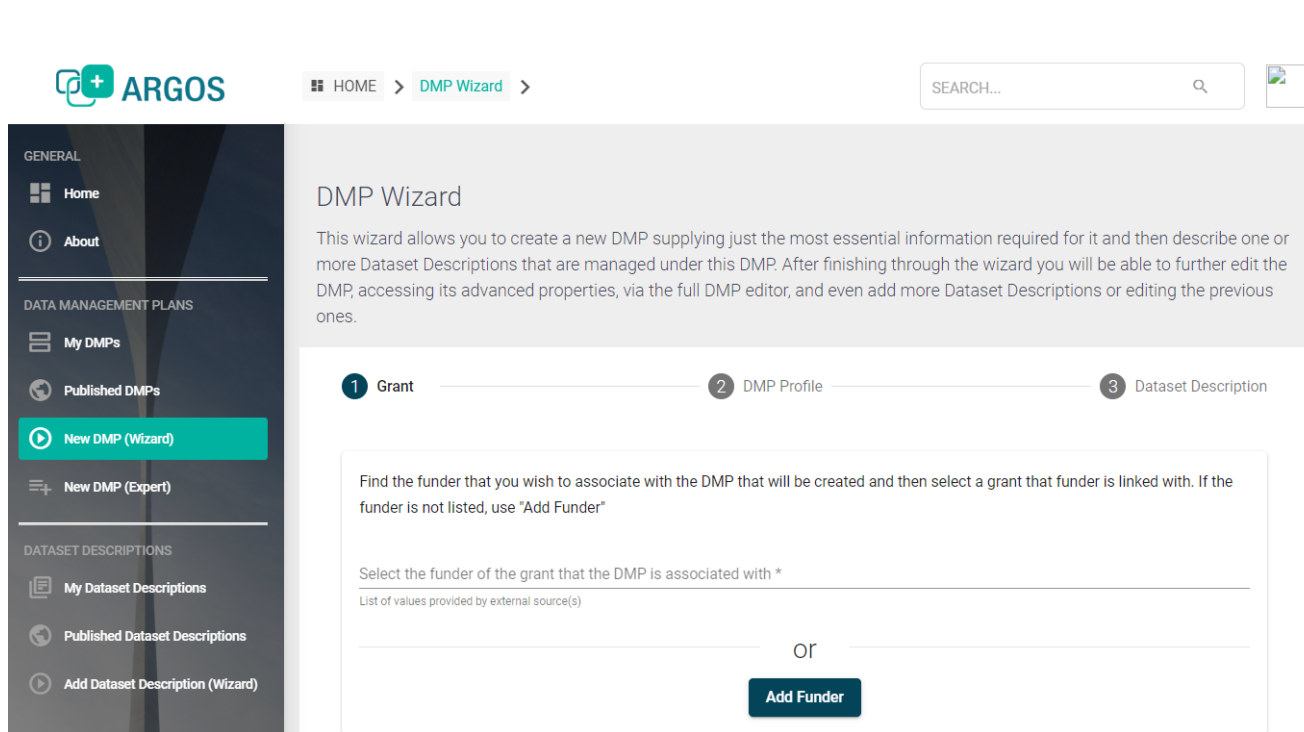
[ECePS ERA Chair in e-Governance and Digital Public Services](#)

[Neuronal circuits for perceptual inference](#)

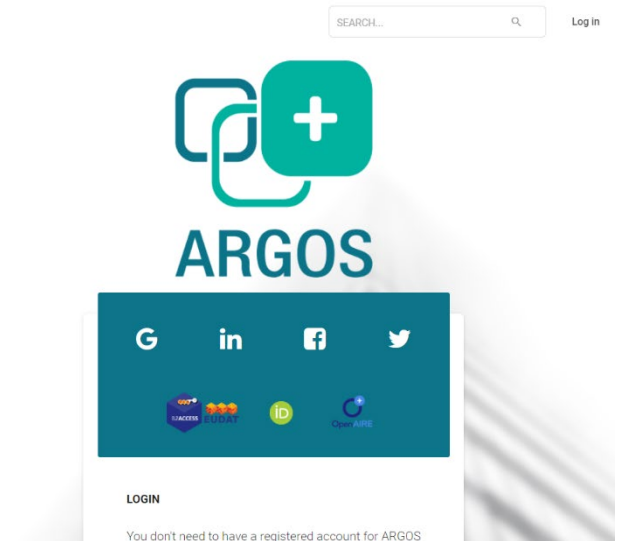
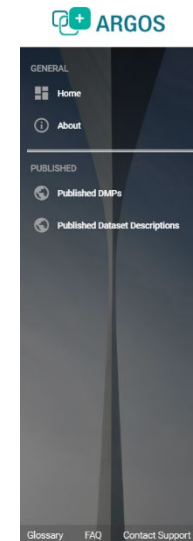
[Health of young Looked After Children in Scotland](#)

[Ethical Issues in 21st Century International Development](#)

ARGOS - Create, Link, Share Data



The screenshot shows the ARGOS DMP Wizard interface. The top navigation bar includes the ARGOS logo, a search bar, and a 'Log in' link. The left sidebar contains a menu with sections: GENERAL (Home, About), DATA MANAGEMENT PLANS (My DMPs, Published DMPs, New DMP (Wizard), New DMP (Expert)), and DATASET DESCRIPTIONS (My Dataset Descriptions, Published Dataset Descriptions, Add Dataset Description (Wizard)). The main content area is titled 'DMP Wizard' and contains a progress bar with three steps: 1. Grant, 2. DMP Profile, and 3. Dataset Description. The first step, 'Grant', is active and contains the following text: 'Find the funder that you wish to associate with the DMP that will be created and then select a grant that funder is linked with. If the funder is not listed, use "Add Funder"'. Below this text is a form with a label 'Select the funder of the grant that the DMP is associated with *' and a placeholder 'List of values provided by external source(s)'. There is an 'or' separator and an 'Add Funder' button.

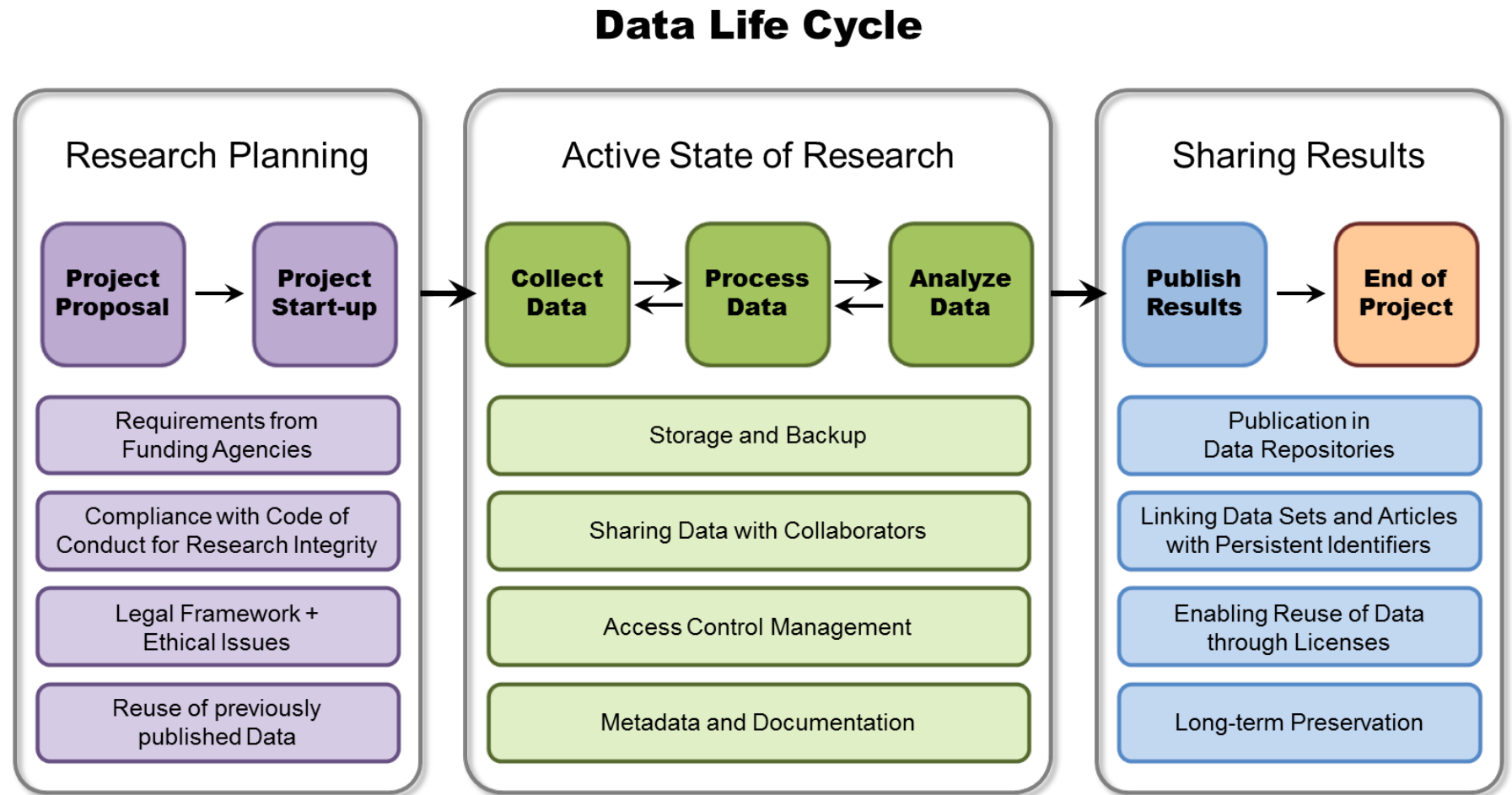


<https://argos.openaire.eu>

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

La gestión de los datos de investigación: el ciclo de los 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>

**DATOS DE INVESTIGACIÓN, DATOS FAIR Y
PLAN DE GESTIÓN DE DATOS**

Versión	Fecha	Modificaciones
2	Diciembre 2022	Consideraciones éticas y protección de datos Formatos de archivo Denominación de archivo Guía para realizar un PGD en Argos

ÍNDICE

1. PLAN DE GESTIÓN DE DATOS: VISIÓN GENERAL.....	2
¿Qué es un Plan de Gestión de Datos (Data Management Plan, DMP)?	2
¿Para qué sirve un DMP?.....	2
¿Por qué se consideran útiles los DMP?	2
¿Dónde encontrar Planes de Gestión de Datos reales?	3
2. PRINCIPIOS FAIR PARA LA GESTIÓN Y ADMINISTRACIÓN DE DATOS CIENTÍFICOS.....	4
3. CONSIDERACIONES ÉTICAS Y DE PROTECCIÓN DE DATOS	6
4. CONSIDERACIONES AL MOMENTO DE ELEGIR FORMATOS DE ARCHIVO.....	7
5. DENOMINACIÓN DE ARCHIVOS	9
6. GENERADORES DE PLANES DE GESTIÓN DE DATOS.....	11
7. GUÍA PARA PGD BASADA EN ARGOS.....	13
8. BIBLIOGRAFÍA.....	22

Descripción y organización de los datos: estructura de carpetas

- Comprobar si hay procedimientos internos 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 por áreas de trabajo, sin nombres de investigadores.
- Una vez decidido un método, todos los miembros participantes han de seguirlo.
- 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.

Descripción y organización de los datos:

Denominación de archivos (i)

- Numerar cada conjunto de datos de forma continua.
- Utilizar nombres únicos, no asignar nombres iguales en carpetas diferentes
- Elegir nombres cortos y representativos, utilizando solo letras, números, guiones bajos y guiones. Decidir cómo se van a usar mayúsculas, guiones, etc.
- 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.

Descripción y organización de los datos: Denominación de archivos (ii)

- En 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)

Descripción y organización de los datos:

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.

Elección de formatos de archivos (abiertos)

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.

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

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, .tiff) 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

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



Federal Records Management

[Home](#) > [Federal Records Management](#) > [Records Management Regulations, Policy, and Guidance](#) > [Appendix A: Tables of File Formats](#)

Records Management Resources

Email Management

[Records Management FAQs](#)
[Memorandums to Agency Records Officers](#)[Federal Records Centers \(FRC\) Guidance and Policy for Accessioning](#)[Records Management Policy and Guidance](#)[Records Management Self Assessment \(RMSA\)](#)[Records Management Training Electronic Records Archives \(ERA\)](#)[Basic Laws & Authorities](#)
[Meetings & Activities](#)

Operating Status & Schedules

[View the National Archives Operating Status](#)

Appendix A: Tables of File Formats

Quick Links

Computer Aided Design	Digital Audio
Digital Cinema	Digital Video
Digital Photographs	Scanned Text
Geospatial Formats	Presentation Formats
Structured Data Formats	Email
Calendars	Navigational Charts

Symbol Key

Preferred Formats	
Acceptable Formats	
Acceptable for Imminent Transfer Formats	

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)

U.S. National Archives and Records Administra tion

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

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 sujetos 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](https://doi.org/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>

Choose a License

Answer the questions or use the search to find the license you want

↻ Start again



Are all the elements of your dataset licensed under a public license or in the Public Domain?

Yes

No

Search for a license...

Public Domain Dedication (CC Zero)

CC Zero enables scientists, educators, artists and other creators and owners of copyright- or database-protected content to waive those interests in their works and thereby place them as completely as possible in the public domain, so that others may freely build upon, enhance and reuse the works for any purposes without restriction under copyright or database law.

Publicly Available



OPEN DATA

Creative Commons Attribution (CC-BY)

This is the standard creative commons license that gives others maximum freedom to do what they want with your work.

Publicly Available



OPEN DATA

Creative Commons Attribution-ShareAlike (CC-BY-SA)

This creative commons license is very similar to the regular Attribution license, but requires you to release all derivative works under this same license.

Publicly Available



OPEN DATA

EUDAT licence selector.

<https://ufal.github.io/public-license-selector/>

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



How to License Research Data

Alex Ball (DCC)



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

¡¡Buscar apoyo legal especializado!!

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.

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

Aspectos éticos y de protección de datos

Reglamento General de Protección de Datos (RGPD) y recomendaciones para la formulación del consentimiento para la reutilización de datos personales

Examples of sensitive data are:

- **Personal data:** identifiers such as names or identification numbers, physical, physiological, genetic, mental, economic, cultural or social characteristics, it also includes location data from GPS or mobile phones
- **Confidential data:** trade secrets, investigations, data protected by intellectual property rights Security: passwords, financial information, national safety, military information...
- **Combination of different datasets** that can be combined into sensitive or personal data
- **Biological data:** endangered (plant or animal) species, where their survival is dependent on the protection of their location data (biodiversity community) 
- **Personal and sensitive metadata** 

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.

Exención en el RGPD sobre el uso de datos con fines de investigación, si se trata de "aspectos de interés público, investigación científica o histórica o fines estadísticos" (Art. 5.1 2016/679 / UE):

Data Management Expert Guide

1. Plan
2. Organise & Document
3. Process
4. Store
5. Protect
 - Ethics and data protection
 - Ethical review process
 - Processing personal data
 - Diversity in data protection
 - Informed consent**
 - Anonymisation
- Copyright
 - Diversity in copyright
- Adapt your DMP: part 5
- Sources and further reading
6. Archive & Publish
7. Discover
8. Contributors

Informed consent

The following statement has been adapted from an actual consent form: **"Any information I give will be used for research only and will not be used for any other purpose"**.

Consider the implications for data sharing for any data generated using this consent statement. Do you have any suggestions for alternative wording or other changes?

⊕ Some thoughts on this statement



Informed consent is the process by which a researcher discloses appropriate information about the research so that a participant may make a voluntary, informed choice to accept or refuse to cooperate.

Normally informed consent is given before the start of the research. Gaining informed consent is crucial to meeting your legal and ethical obligations towards participants whilst simultaneously enhancing the value of your research data.

To obtain informed consent, researchers should:

- Inform participants about the purpose of the research;
- Discuss what will happen to their contribution (including the future archiving and sharing of their data);
- Indicate the steps that will be taken to safeguard their anonymity and confidentiality;
- Outline their right to withdraw from the research.

Request for participation in research project

[This template serves as an example of an information letter.

Clear text and insert your own.

NB - all information should be concise and easy to understand]

"[Insert project title]"

Background and Purpose

[Describe the purpose of the project and briefly sketch the whether the project is a master's or Ph.D. project at the inst as commissioned research, in cooperation with other inst]

[Describe how the sample has been selected and/or why participate.]

What does participation in the project imply?

[Describe the main features of the project: data collection (surveys, interviews, observation, tests, etc., preferably any collection of data about the participant from other files, other informants, etc.). Describe the types of data concern...), and the manner(s) in which data will be collected.]

If parents are to give consent on behalf of their children the questionnaire/interview guide etc. If multiple samples project, it must be explicitly indicated what participants a separate information letter must be made for each group.

What will happen to the information about you?

All personal data will be treated confidentially. [Describe data (e.g. only the project group, student and supervisor personal data/recordings will be stored to ensure confidentiality stored separately from other data).]

[Describe whether participants will be recognizable]

The project is scheduled for completion by [insert personal data and any recordings at that point. If project completion: state the purpose of further study will have access, as well as the final date for analysis]

Version 3 (MAY 2016)

INFORMATION SHEET - VOLUNTEERS

PROJECT TITLE: project title here
Ethics Code: ethics code here

You are being invited to take part in a research study. Before you decide it is important for you to understand why the research is being done and what it will involve. Please take time to read the following information carefully and discuss it with others if you wish. Ask us if there is anything that is not clear or if you would like more information. Take time to decide whether or not you wish to take part.

What is the purpose of this study?

Please type here the purpose of the study. Please type here the purpose of the study. Please type here the purpose of the study. Please type here the purpose of the study. Please type here the purpose of the study.

Why have I been chosen?

You have been chosen because you are a member of the CBSU Volunteer Panel, or because you answered an advert asking for volunteers to participate in research studies using functional magnetic resonance imaging (fMRI).

Who is organizing this study?

The study is organized by a researcher working at the Medical Research Council Cognition and Brain Sciences Unit (MRC CBSU), Cambridge. Your Principle Researcher is:

Principle Researcher Name + (CBSU direct dial number)

What does the procedure involve?

Before your scan a member of staff will ask you some questions to ensure that you have no metal within you before you enter the strong magnetic field. You will then be asked to lie in the scanner and the scanning will start. The scanning can be noisy and so we shall give you ear plugs as well as headphones to reduce this noise. It may not be appropriate for you to be scanned if you are very claustrophobic. During some of the scans we will ask you to perform simple tasks. ~~Replace this phrase and explain with details of the tasks specific to your project. For example: These tasks will involve remembering items, such as digits or objects, or simple task rules. You will also have to make simple decisions by pressing a button on a button box in your right hand, or by telling us what those decisions are.~~

You will have ample time before scanning to practice the tasks to ensure you are comfortable with them. The tasks we will be using have been used at the MRC CBSU and usually present volunteers with no significant problems. The scanning session will take about one and a half hours, although you will not actually be scanned for more than 45 minutes of this time.

What is the device involved?

We can learn a great deal about how the brain works by looking at the blood flow to different parts of the brain whilst the brain performs different tasks. We need to obtain this information in both health and disease. We measure brain function using images taken with a magnetic resonance imaging scanner. This scanner uses a strong magnetic field to create detailed images of brain structure and function. By taking a series of images whilst you perform a task we can build up a picture of the brain areas activated by this type of function. The scan does not involve any injections or X-rays.

What are the possible risks/side effects of taking part?

The scanner can be loud when it takes images, and you will be given earplugs and ear defenders to block out some of the sound. Also, the MR environment is quite confined, and people who are uncomfortable in small or confined spaces may not be able to participate. If this should be you, remember that you may withdraw from the study at any time without explaining why. Otherwise fMRI is generally thought to be a safe, non-invasive imaging technique. There are no known risks or side effects.

What are the possible benefits of taking part?

We will reimburse you for your time and contribute towards the cost of your travel, and you will have the pleasure of knowing that you have made a contribution to our understanding of the relationship between brain and behaviour. If you have made a contribution to our understanding of the relationship between brain and behaviour, you can take away a picture of your brain on the day of your scan. There will be no other direct benefits from being involved in the study.

MRC Cognition & Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 7EF
Tel: 01223 355294; web: <http://www.cbsu.mrc.ac.uk/>

Version 100-01-1008

<https://www.cessda.eu/Training/Training-Resources/Library/Data-Management-Expert-Guide/5.-Protect/Informed-consent>

Compartir datos: los Repositorios.

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
- Se recomienda que los datos se hagan públicos siguiendo los *Science Commons [Protocol for Implementing Open Access Data](#) and the [Open Knowledge/Data Definition](#)*
- Elegir un repositorio adecuado, **que cumpla con los criterios FAIR**

README.TXT



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

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

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.consorciomadrano.es/readme.xhtml>

Plantilla readme.txt

Readme.txt ([acceso](#) [descarga](#)) es una plantilla para que rellenen los investigadores que quieran depositar un dataset en e-cienciaDatos.

Todos los conjuntos de datos han de estar documentados para ser comprensibles y reutilizables: autoría, título, descripción, metodología, proyectos financiadores, cobertura temporal y geográfica, derechos de uso y privacidad, etc.

Los gestores del repositorio e-cienciaDatos utilizarán la información proporcionada en la plantilla readme.txt para:

- cumplimentar los metadatos estandarizados que describen el dataset y que son recolectados por diversos servidores científicos.
- elaborar un fichero readme.txt que se incluirá junto al resto de ficheros del dataset.

Copyright © 2020 Consorcio
Madroño



re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES
<http://doi.org/e-cienciaDatos>



Powered by The
Dataverse
Project

```
readme-en.txt - Editor
Datei Bearbeiten Format Ansicht Hilfe

-----
GENERAL INFORMATION
-----
1. Title of dataset

2. Contact

    Principal Investigator Contact Information
    Name:
    Institution:
    Email:
    ORCID:

3. Description of the project

4. Description of the dataset

5. Notes

6. Deposit date

7. Date

8. Language

-----
AUTHOR INFORMATION
-----
1. Author

    Name:
    Last name :
    Institution:
    Email:
    ORCID:
```

```
readme-en.txt - Editor
Datei Bearbeiten Format Ansicht Hilfe

-----
METHODOLOGY
-----
1. Methodology

2. Software

-----
KEYWORDS
-----
1. Keywords

-----
SPONSORSHIP INFORMATION AND GRANT IDs
-----
1. Grant Information

-----
RELATED PUBLICATIONS
-----
1. Related publication

2. Related dataset

-----
GEOGRAPHIC INFORMATION
-----
1. Spatial coverage

-----
TEMPORAL INFORMATION
-----
1. Time period coverage
```

```
readme-en.txt - E...
Datei Bearbeiten Format Ansicht Hilfe

-----
FILES
-----
1. Files

-----
LICENSES AND PRIVACY
-----
1. Licenses

2. Privacy

-----
OTHERS
-----
1. Data dictionary
```



Make a README

Because no one can read your mind (yet)



Need an interactive developer hub, complete with documentation, reference guides, tutorials, realtime API logs and community? Try [ReadMe.com!](#)

Make a README is proudly sponsored by ReadMe.com

Make a README

<https://www.makeareadme.com/>

README 101

What is it?

A [README](#) is a text file that introduces and explains a project. It contains information that is commonly required to understand what the project is about.

Why should I make it?

It's an easy way to answer questions that your audience will likely have regarding how to install and use your project and also how to collaborate with you.

Who should make it?

Anyone who is working on a programming project, especially if you want others to use it or contribute.

When should I make it?

Definitely before you show a project to other people or make it public. You might want to get into the habit of making it the [first file you create](#) in a new project.

Where should I put it?

In the top level directory of the project. This is where someone who is new to your project will start out. Code hosting services such as [GitHub](#), [Bitbucket](#), and [GitLab](#) will also look for your README and display it along with the list of files and directories in your project.

How should I make it?

While READMEs can be written in any text file format, the most common one that is used nowadays is [Markdown](#). It allows you to add some lightweight formatting. You can learn more about it at the [CommonMark website](#), which also has a helpful [reference guide](#) and an [interactive tutorial](#).

Some other formats that you might see are [plain text](#), [reStructuredText](#) (common in [Python](#) projects), and [Textile](#).

You can use any text editor. There are plugins for many editors (e.g. [Atom](#), [Emacs](#), [Sublime Text](#), [Vim](#), and [Visual Studio Code](#)) that allow you to preview Markdown while you are editing it.

You can also use a dedicated Markdown editor like [Typora](#) or an online one like [StackEdit](#) or [Dillinger](#). You can even use the editable template below.

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

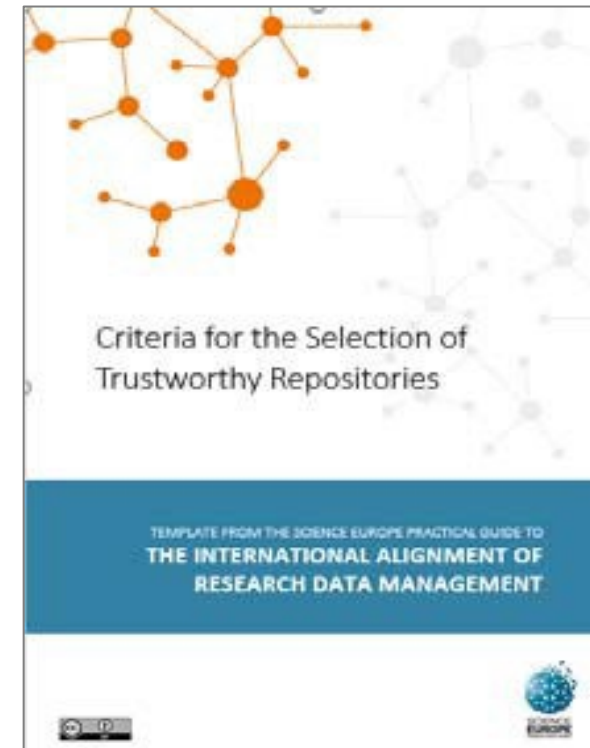
- a. Enable finding of data
- b. Enable referencing to related relevant information, such as other data and publications
- c. Provide information that is publicly available and maintained, even for non-published, protected, retracted, or deleted data
- d. Use metadata standards that are broadly accepted (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>

Ejemplo de dataset abierto

February 9, 2021 (1.0.0)

Dataset

Open Access

Catalan Textual Corpus

De Gibert Bonet, Ona; Armengol-Estapé, Jordi; Rodríguez-Penagos, Carlos; Melero, Maite; Villegas, Marta; Carrino, Casimiro Pio; Armentano-Oller, Carme; Gonzalez-Agirre, Aitor; Asensio, Alejandro

The Catalan Textual Corpus is a 1760-million-token web corpus of Catalan built from several sources: existing corpus such as DOGC, CaWac (non-dedup version), Oscar (unshuffled version), Open Subtitles, Catalan Wikipedia; and three brand new crawlings: the Catalan General Crawling, obtained by crawling the 500 most popular .cat and .ad domains; the Catalan Government Crawling, obtained by crawling the .gencat domain and subdomains, belonging to the Catalan Government; and the ACN corpus with 220k news items from March 2015 until October 2020, crawled from the Catalan News Agency.

Uploaded on February 9, 2021

<https://zenodo.org/record/4519349#.YQyQHbMKUk>

February 9, 2021

Dataset

Open Access

Catalan Textual Corpus

De Gibert Bonet, Ona; Armengol-Estapé, Jordi; Rodríguez-Penagos, Carlos; Melero, Maite; Villegas, Marta; Carrino, Casimiro Pio; Armentano-Oller, Carme; Gonzalez-Agirre, Aitor; Asensio, Alejandro

The Catalan Textual Corpus is a 1760-million-token web corpus of Catalan built from several sources: existing corpus such as DOGC, CaWac (non-dedup version), Oscar (unshuffled version), Open Subtitles, Catalan Wikipedia; and three brand new crawlings: the Catalan General Crawling, obtained by crawling the 500 most popular .cat and .ad domains; the Catalan Government Crawling, obtained by crawling the .gencat domain and subdomains, belonging to the Catalan Government; and the ACN corpus with 220k news items from March 2015 until October 2020, crawled from the Catalan News Agency.

It consists of 1.758.388.896 tokens, 73.172.152 sentences and 12.556.365 documents. Documents are separated by single new lines. These boundaries have been preserved as long as the license allowed it.

We license the actual packaging of these data under a Attribution-ShareAlike 4.0 International License.

Copyright (c) 2021 Text Mining Unit at BSC

If you use this resource in your work, please cite our latest paper:

@misc{armengolestape2021multilingual, title={Are Multilingual Models the Best Choice for Moderately Under-resourced Languages? A Comprehensive Assessment for Catalan}, author={Jordi Armengol-Estapé and Casimiro Pio Carrino and Carlos Rodríguez-Penagos and Ona de Gibert Bonet and Carme Armentano-Oller and Aitor Gonzalez-Agirre and Maite Melero and Marta Villegas}, year={2021}, eprint={2107.07903}, archivePrefix={arXiv}, primaryClass={cs.CL}}

Funded by the Generalitat de Catalunya, Departament de Politiques Digitals i Administració Pública (AINA), MT4ALL and Plan de Impulso de las Tecnologías del Lenguaje (Plan TL).

Preview

catalan_textual_corpus.zip

LICENSE

348 Bytes

README.md

2.7 kB

corpus

catalan_textual_corpus.txt

10.9 GB

607

views

123

downloads

[See more details...](#)

Indexed in

OpenAIRE

Publication date:

February 9, 2021

DOI:

DOI 10.5281/zenodo.4519349

Keyword(s):

Catalan AI Corpus NLP

Communities:

[Catalan AI language resources](#)

License (for files):

[Creative Commons Attribution Share Alike 4.0 International](#)

Versions

Version 1.0.0

Feb 9, 2021

10.5281/zenodo.4519349

Cite all versions? You can cite all versions by using the DOI 10.5281/zenodo.4519348. This DOI represents all versions, and will always resolve to the latest one. [Read more.](#)

Ejemplo de dataset embargado

https://zenodo.org/record/7627346#.Y_OzCnbMKUk

February 10, 2023

Dataset Embargoed Access

Replication code and data for "Variation in Hunter-Gatherer Diets Compared to the Paleo Diet".

Lieberman, Daniel E.; Worthington, Steven; Schell, Laura; Parkent, Christine M.; Devinsky, Orrin; Carmody, Rachel N.

Replication code and data for the publication "Variation in Hunter-Gatherer Diets Compared to the Paleo Diet". Files include:

- 1) R replication code (3 files that should be run sequentially in numerical order).
- 2) Raw data: 1 file that is ingested by the R code.
- 3) GeoTiff file: 1 file of GeoTiff data on annual mean temperature from the WorldClim v. 2.1 dataset (<http://www.worldclim.com/version2>) that is ingested by the R code.
- 4) Table 1: a cleaned version of the data used in analyses that is output by the R code.
- 5) Supplementary data: 2 files containing seasonal and HG diet data that were not used in analyses, which are output by the R code.

Files

Embargoed Access
Files are currently under embargo but will be publicly accessible after April 30, 2023.

Citations 0

Show only:

☐ Literature (0) ☐ Dataset (0) ☐ Software (0) ☐ Unknown (0)

☐ Citations to this version

Search

No citations.

6 views

2 downloads

[See more details...](#)

Indexed in

OpenAIRE

Publication date:

February 10, 2023

DOI:

DOI [10.5281/zenodo.7627346](https://doi.org/10.5281/zenodo.7627346)

Keyword(s):

[diet](#) [hunter-gatherers](#) [Paleo Diet](#)

License (for files):

Creative Commons Attribution 4.0 International

Versions

Version 1

Feb 10, 2023

[10.5281/zenodo.7627346](#)

Cite all versions? You can cite all versions by using the DOI [10.5281/zenodo.7627345](https://doi.org/10.5281/zenodo.7627345). This DOI represents all versions, and will always resolve to the latest one. [Read more.](#)

Share

Cite as

Lieberman, Daniel E., Worthington, Steven, Schell, Laura, Parkent, Christine M., Devinsky, Orrin, & Carmody, Rachel N. (2023). Replication code and data for "Variation in Hunter-Gatherer Diets Compared to the Paleo Diet". [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.7627346>

February 10, 2023 (v1) Dataset Embargoed Access

View

Ejemplo de dataset restringido

March 22, 2021 (v1) Dataset Restricted Access

Dataset related to the article "Prospective use of ablation index for the ablation of right ventricle outflow tract premature ventricular contractions: a proof of concept study."

Gasperetti, Alessio; Sicuso, Rita; Dello Russo, Antonio; Zucchelli, G; Saguner, AM; Notarstefano, P; Soldati, E; Bongiorno, Maria Grazia; Della Rocca, Domenico; Mohanty, S; Carbucicchio, Corrado; Duru, F; Di Biase, Luigi; Natale, Andrea; Tondo, Claudio; Casella, Michela

This record contains raw data related to the article "Prospective use of ablation index for the ablation of right ventricle outflow tract premature ventricular contractions: a proof of concept study". Abstract Aims: Radiofrequency catheter ablation (RFCA) represent

Uploaded on March 22, 2021

View

March 22, 2021

Dataset Restricted Access

Dataset related to the article "Prospective use of ablation index for the ablation of right ventricle outflow tract premature ventricular contractions: a proof of concept study."

Gasperetti, Alessio; Sicuso, Rita; Dello Russo, Antonio; Zucchelli, G; Saguner, AM; Notarstefano, P; Soldati, E; Bongiorno, Maria Grazia; Della Rocca, Domenico; Mohanty, S; Carbucicchio, Corrado; Duru, F; Di Biase, Luigi; Natale, Andrea; Tondo, Claudio; Casella, Michela

This record contains raw data related to the article "Prospective use of ablation index for the ablation of right ventricle outflow tract premature ventricular contractions: a proof of concept study".

14

views

0

downloads

[See more details...](#)

Indexed in

OpenAIRE

Publication date:

March 22, 2021

DOI:

[DOI: 10.5281/zenodo.4627191](https://doi.org/10.5281/zenodo.4627191)

Related identifiers:

Supplement to
[10.1093/europace/eaab228](https://doi.org/10.1093/europace/eaab228) (Journal article)

Communities:

[Centro Cardiologico Monzino IROCCS](#)

Versions

Version 1

Mar 22, 2021

[10.5281/zenodo.4627191](https://doi.org/10.5281/zenodo.4627191)

Cite all versions? You can cite all versions by using the DOI [10.5281/zenodo.4627190](https://doi.org/10.5281/zenodo.4627190). This DOI represents all versions, and will always resolve to the latest one. [Read more.](#)

Share

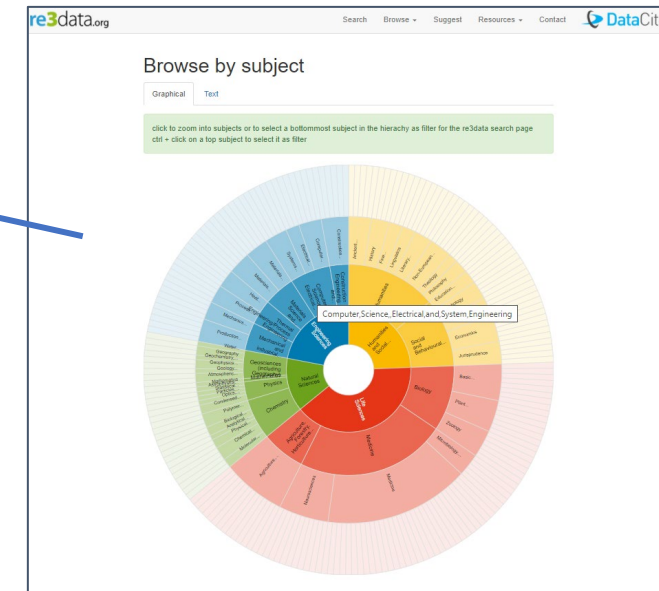
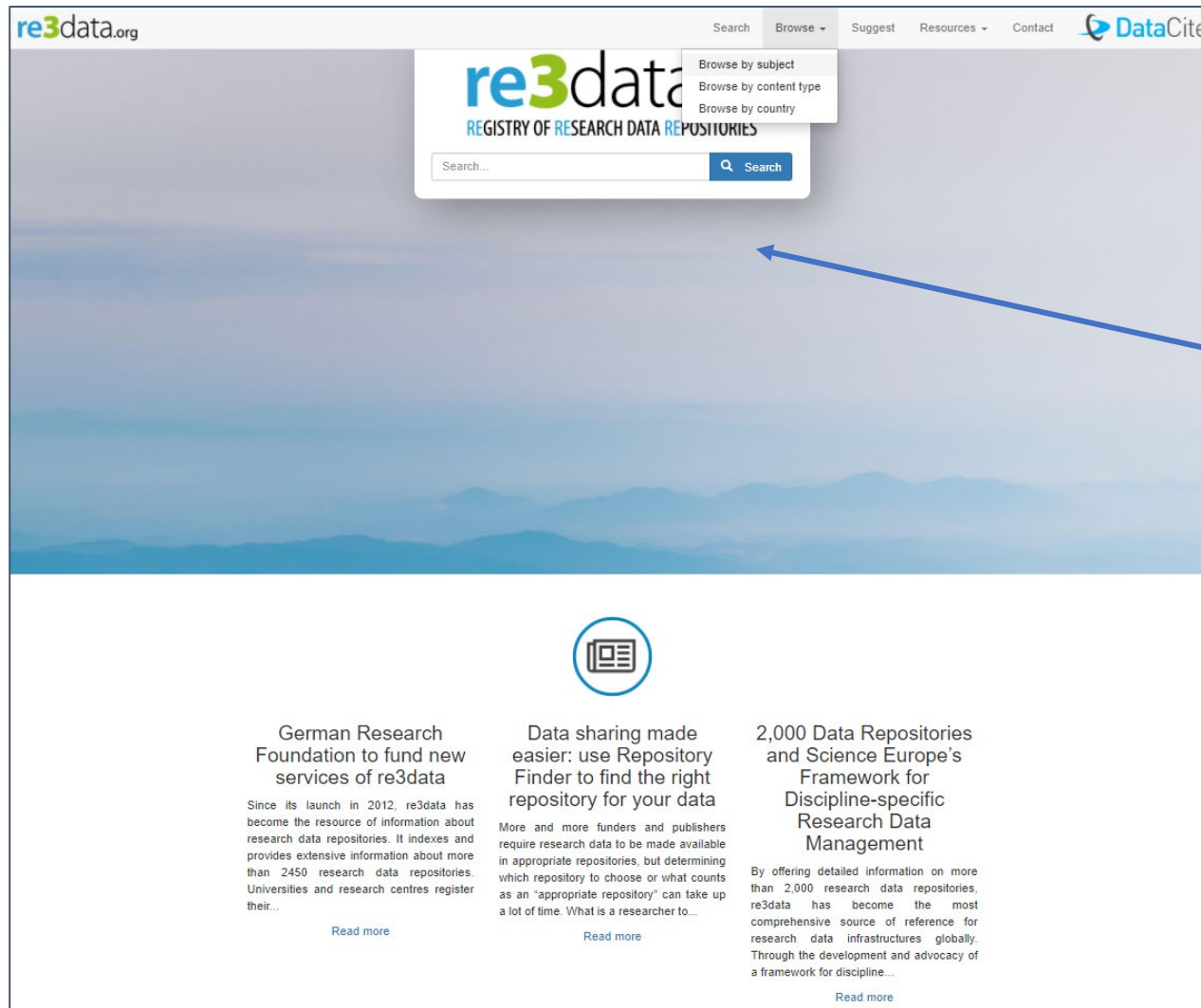


Cite as

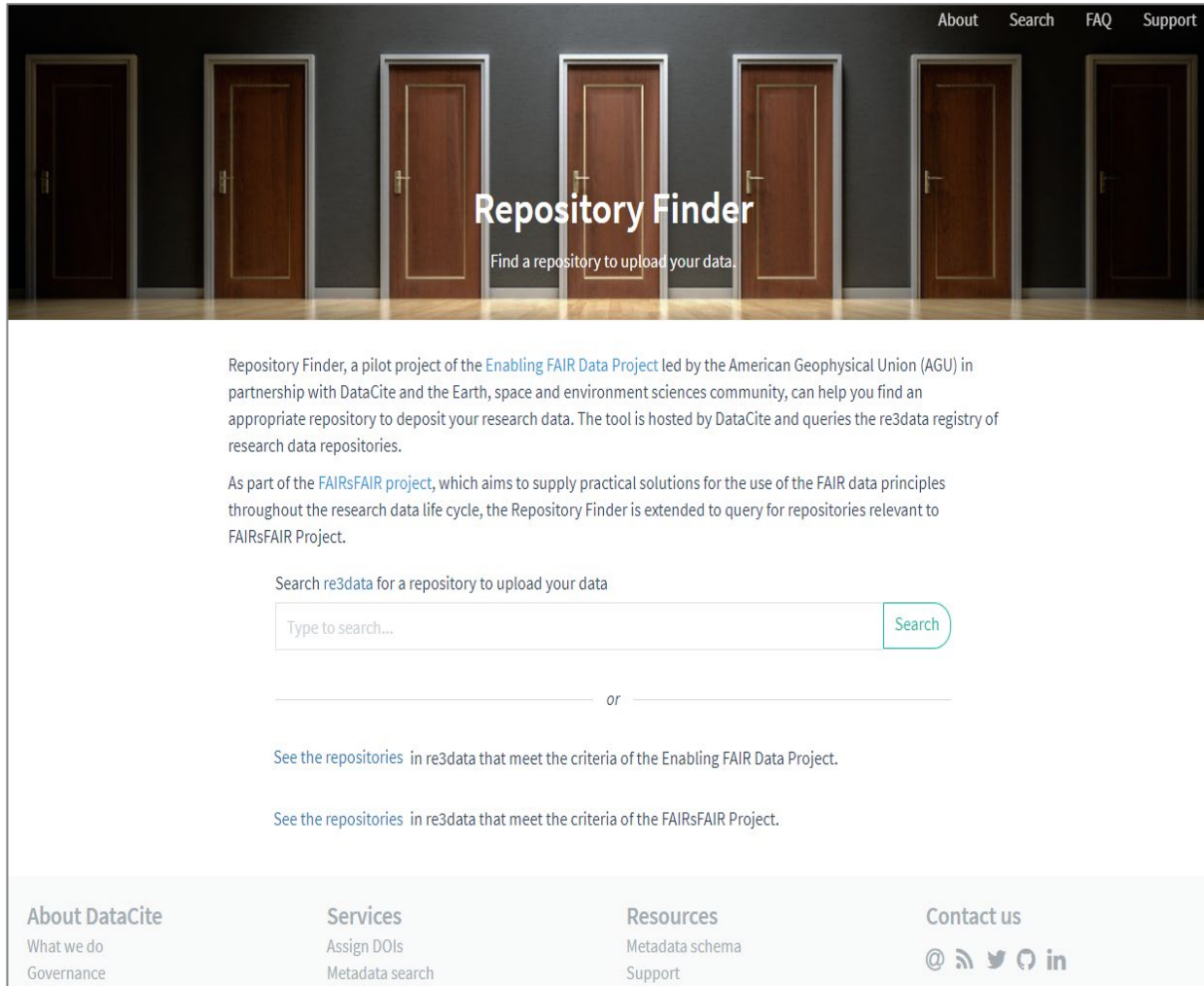
Gasperetti, Alessio, Sicuso, Rita, Dello Russo, Antonio, Zucchelli, G, Saguner, AM, Notarstefano, P, Soldati, E, Bongiorno, Maria Grazia, Della Rocca, Domenico, Mohanty, S, Carbucicchio, Corrado, Duru, F, Di Biase, Luigi, Natale, Andrea, Tondo, Claudio, & Casella, Michela. (2021). Dataset related to the article "Prospective use of ablation index for the ablation of right ventricle outflow tract premature ventricular contractions: a proof of concept study." [Data set]. Zenodo. <https://doi.org/10.5281/zenodo.4627191>

https://zenodo.org/record/4627191#.Y_OwAnbMKUk

No es
abierto pero
sí FAIR!



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](https://www.re3data.org/))



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**

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.

The screenshot shows the Zenodo 'New upload' interface. At the top, there's a blue header with the Zenodo logo, a search bar, and 'Upload' and 'Communities' tabs. Below the header, there are 'Delete', 'Save', and 'Publish' buttons. The main section is titled 'New upload' and includes instructions. It features a 'Files' section with a 'Choose files' button and a 'Communities' section with a 'recommended' dropdown. The 'Upload type' section shows various icons for different file types, with 'Dataset' selected. The 'Basic information' section contains fields for 'Digital Object Identifier' (with a 'Reserve DOI' button), 'Publication date' (set to 2020-12-12), 'Title', 'Authors' (with a family name and affiliation field), and an 'ORCID' field. A rich text editor for the 'Description' is at the bottom.

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

¡Muchas Gracias!

✉ alicia.gomez@ie.edu

🐦 <https://twitter.com/fagomsan>

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

id <https://orcid.org/0000-0003-4898-1680>