

# Plan de Gestión de Datos

Plantilla y guía para proyectos financiados en el marco del Plan Estatal de Investigación Científica y Técnica y de Innovación

## Control de versiones

Versión	Fecha	Autor/revisor	Comentarios
<b>1.0</b>	17/01/2024	Jordi Moretón Lorena López	Primera versión

Este documento pretende servir de apoyo a los investigadores del Institut de Recerca Sant Joan de Déu (IRSJD) para crear sus Planes de Gestión de Datos (PGD, o *Data Management Plan, DMP*); concretamente, para los proyectos financiados en el marco del Plan Estatal de Investigación Científica y Técnica y de Innovación.

A continuación, se muestra:

- Con un número, los diferentes apartados de la plantilla.
- Con una letra mayúscula, los elementos que se deberían tener en cuenta en el momento de rellenar el apartado.
- Con una letra minúscula, las descripciones y una muestra de ejemplos reales.
- Con tablas resaltadas de color rojo se resaltan ejemplos institucionales, información institucional relevante o pequeños párrafos de guía institucional.

### Información institucional

Se proporcionan respuestas institucionales y/o información clave del IRSJD para varios campos de este documento. Estos se proporcionan para guiar y ayudar a responder cada ítem de acuerdo con las políticas, protocolos y sistemas institucionales establecidos, pero pueden necesitar modificaciones para adaptarse mejor a las particularidades de cada proyecto. Por favor contacta con [freerca.support-dmp@sjd.es](mailto:freerca.support-dmp@sjd.es) si necesitas apoyo en la elaboración de tu DMP.

Este documento ha sido adaptado por el IRSJD a partir de la guía de PGD versión 1 de enero de 2020 elaborada por el Grupo de Trabajo de Apoyo a la Investigación del CSUC que está formado por representantes de las siguientes universidades: Universitat de Barcelona, Universitat Autònoma de Barcelona, Universitat Politècnica de Catalunya, Universitat Pompeu Fabra, Universitat de Girona, Universitat de Lleida, Universitat Rovira i Virgili, Universitat Oberta de Catalunya, Universitat de Vic-Universitat Central de Catalunya, Universitat Ramon Llull, Universitat de les Illes Balears y Universitat Jaume I. Especial reconocimiento al Servicio de Apoyo del Consorcio de Servicios Universitarios de Catalunya (CSUC), por su visión del Plan de Gestión de Datos en castellano.

Este documento se encuentra bajo Licencia de [Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/)



## Información preliminar

Este documento está destinado a ayudar a los investigadores y gestores del IRSJD en la creación y actualización de un Plan de Gestión de Datos de Investigación (PGD).

El entregable también suele incluir otra información preliminar como: el logotipo del proyecto, el nivel de diseminación, las revisiones históricas, la tabla de contenido y una lista de acrónimos utilizados.

## Plantilla / Tabla de contenido

<b>1. Resumen de los datos .....</b>	<b>5</b>
1.A Describir los objetivos de los datos generados y/o recolectados y su relación con los objetivos del proyecto.....	5
1.B Especificar el tipo y el formato de los datos que se generarán/recolectarán.....	6
1.C Especificar si se reutilizan los datos existentes y como .....	9
1.D Especificar el origen de los datos .....	10
1.E Indicar el tamaño esperado de los datos.....	10
1.F Describir la utilidad de los datos: para quienes pueden ser útiles.....	11
<b>2. Datos FAIR .....</b>	<b>13</b>
2.1 Datos encontrables (incluyendo los metadatos) .....	13
2.2 Datos accesibles .....	16
2.3 Datos interoperables .....	24
2.4 Datos reutilizables.....	25
<b>3. Asignación de recursos.....</b>	<b>30</b>
3.A Estimar el coste para hacer que los datos sean FAIR en el proyecto .....	30
3.B Especificar como se prevé cubrir el coste .....	30
3.C Especificar quién es el responsable de la gestión en cada proyecto .....	31
3.D Describir los recursos para la preservación a largo plazo .....	32
<b>4. Seguridad de los datos .....</b>	<b>33</b>
4.A Tratamiento de recuperación de datos, así como el almacenaje y la transferencia de datos confidenciales	33
<b>5. Aspectos éticos.....</b>	<b>37</b>
5.A Problemas éticos o legales que pueden afectar la recogida e intercambio de datos.....	37
<b>6. Otros temas .....</b>	<b>42</b>
6.A Referenciar otros procedimientos nacionales/sectoriales/institucionales/de financiación usados en la gestión de datos de investigación .....	42
<b>7. Otro soporte en el desarrollo del plan.....</b>	<b>44</b>
7.A Explicar los recursos utilizados para el desarrollo del plan.....	44

## 1. Resumen de los datos

---

1.A Describir los objetivos de los datos generados y/o recolectados y su relación con los objetivos del proyecto

### **1.A a) Descripción**

Explicar de manera breve el objetivo de los datos generados y/o recolectados y su relación con los objetivos del proyecto.

### **1.A b) Ejemplo real**

**Ej. 1** Los datos se originarán a partir de mediciones, calibraciones, comparaciones y validaciones. Se utilizará para cumplir los objetivos del proyecto y en conferencias y publicaciones revisadas por pares.

El consorcio recopilará datos experimentales para cumplir los objetivos 1 a 4. Los datos de medición y calibración resultarán de los objetivos 1 y 3 y los datos de comparación y validación de los objetivos 2 y 4. Los datos de cuestionarios y estudios de mercado se utilizarán para respaldar aceptación por parte del usuario final (objetivo 5).

**Ej. 2** Recopilar y poner a disposición los datos del análisis de materiales superconductores para respaldar la credibilidad y elevar la calidad de las publicaciones científicas basadas en esos datos. Facilitar el intercambio de datos dentro del Consorcio y promover la caracterización distribuida de muestras con diferentes métodos. Permitir proyectos de seguimiento y que futuras generaciones de estudiantes continúen el trabajo para aprovechar los conjuntos de datos existentes, validar los resultados y documentar la mejora de los materiales y las técnicas de producción de manera verificable. Este enfoque garantizará un impacto duradero de este proyecto financiado por la CE más allá del período del proyecto.

El objetivo del proyecto es mejorar el rendimiento de los cables superconductores y, en una etapa posterior, las películas delgadas, obteniendo una mejor comprensión del comportamiento del material, la influencia de las técnicas de producción en el rendimiento y dilucidar las limitaciones de rendimiento (por ejemplo, el factor de calidad para los cables superconductores). películas delgadas sobre sustrato, límites de corriente en cables en condiciones de campo magnético alto). La recopilación y publicación gestionada de los datos ayudará a establecer una biblioteca duradera de resultados que pueda ayudar a documentar la evolución del desempeño a lo largo de varios años y permitir que otros investigadores validen los resultados de forma independiente.

## 1.B Especificar el tipo y el formato de los datos que se generarán/recolectarán

### **1.B a) Descripción**

Describir el contenido y alcance de los datos. Los datos de investigación se generan por diferentes motivos a través de diferentes procesos y pueden ser de los siguientes tipos:

- Observacionales: datos capturados en tiempo real (neuroimágenes, datos de muestras, datos de sensores, datos de encuestas, etc.).
- Experimentales: datos capturados en equipos de laboratorio (secuencias de genes, cromatogramas, datos de campos magnéticos, etc.).
- Simulación: datos generados a partir de modelos de prueba (climatológicas, matemáticas o modelos económicos, etc.).
- Derivadas o compiladas: datos reproducibles, pero de difícil reproducción (texto y minería de datos, modelos 3D, base de datos compilada, etc.).
- De referencia: conglomerado o conjunto de datos (bancos de datos de secuencias de genes, estructuras químicas, portales de datos espaciales, etc.).
- Otros.

El formato de los datos (texto, numérico, imagen, etc.) también debería indicarse.

### **Información Institucional**

Puede encontrar información de formatos (aceptados y recomendados para conservación a largo plazo) en el siguiente enlace:

<https://confluence.csuc.cat/display/RDM/Recomanacions+de+formats>

Tipo de Archivo	Formatos Recomendados	Formatos NO recomendados, pero comúnmente aceptados
Documento de texto	<a href="#">PDF/A (.pdf)</a> <a href="#">ODT (.odt)</a>	<a href="#">Microsoft Word (.doc)</a> <a href="#">Office Open XML (.docx)</a> <a href="#">Rich Text File (.rtf)</a> <a href="#">PDF different to a PDF/A (.pdf)</a>
Texto plano	<a href="#">Unicode text (.txt)</a>	<a href="#">Non-Unicode text (.txt)</a>
Lenguajes de marcado	<a href="#">XML (.xml)</a> <a href="#">HTML (.html)</a> <a href="#">YAML (.yaml)</a> <a href="#">JSON (.json)</a> <a href="#">ReStructuredText (.rst)</a> Related files: <a href="#">.css</a> , <a href="#">.xslt</a> , <a href="#">.js</a> , <a href="#">.es</a>	<a href="#">SGML (.sgml)</a> <a href="#">Markdown (.md)</a>
Lenguajes de programación	<a href="#">NetCDF</a> <a href="#">TextFabric</a> <a href="#">R (.r)</a> Octave (.mat)	<a href="#">MATLAB (.mat)</a>
Hojas de cálculo	<a href="#">ODS (.ods)</a>	<a href="#">Microsoft Excel (.xls)</a>

	<a href="#">CSV</a> (.csv)	<a href="#">Office Open XML Workbook</a> (.xlsx) <a href="#">PDF/A</a> (.pdf)
Bases de datos	<a href="#">SQL</a> (.sql) <a href="#">SIARD</a> (.siard) <a href="#">CSV</a> (.csv) <a href="#">FITS</a> (.fits, .fit, .fts) <a href="#">(Apache) Parquet</a> (.parquet)	<a href="#">Microsoft Access</a> (.mdb, .accdb) <a href="#">dBase</a> (.dbf) <a href="#">HDF5</a> (.hdf5, .he5, .h5)
Datos estadísticos	<a href="#">SPSS</a> (.dat/.sps) <a href="#">STATA</a> (.dat/.DO) <a href="#">R</a> (.rdat/.rdara)	<a href="#">SPSS Portable</a> (.por) <a href="#">SPSS</a> (.sav) <a href="#">STATA</a> (.dta) <a href="#">SAS</a> (.7dat; .sd2; .tpt)
Imágenes (mapa de bits)	<a href="#">JPEG</a> (.jpg, .jpeg) <a href="#">TIFF</a> (.tif, .tiff) <a href="#">PNG</a> (.png) <a href="#">JPEG 2000</a> (.jp2) <a href="#">DICOM</a> (.dcm)	
Imágenes vectoriales	<a href="#">SVG</a> (.svg)	<a href="#">Adobe Illustrator</a> (.ai) <a href="#">EPS</a> (.eps) <a href="#">WMF/EMF</a> (.wmf, .emf) <a href="#">CDR</a> (.cdr)
Audio (Contenedor y Códec)	<a href="#">BWF</a> (.bwf)	
Audio (Contenedor)	<a href="#">MXF</a> (.mxf) <a href="#">Matroska</a> (.mka)	Audio (container)
Audio (Códec)	<a href="#">FLAC</a> (.flac) <a href="#">OPUS</a>	<a href="#">WAVE</a> (.wav) <a href="#">MP3</a> (.mp3) <a href="#">AAC</a> (.aac, .m4a) <a href="#">AIFF</a> (.aif, .aiff) <a href="#">OGG</a> (.ogg)
Video (Contenedor)	<a href="#">MXF</a> (.mxf) <a href="#">Matroska</a> (.mkv)	<a href="#">MPEG-4</a> (.mp4, .m4a, .m4v) <a href="#">MPEG-2</a> (.mpg, .mpeg, .m2v, mpg2) <a href="#">AVI</a> (.avi) <a href="#">QuickTime</a> (.mov, .qt)
Video (Códec)	<a href="#">Theora</a> (.ogv)	
Diseño asistido por computación (CAD)	<a href="#">AutoCAD DXF version R12 (ASCII)</a> (.dxf) <a href="#">SVG</a> (.svg)	<a href="#">AutoCAD DXF diferent a R12 (ASCII)</a> (.dwg, .dxf) <a href="#">DWG</a> (.dwg) <a href="#">DGN</a> (.dgn)
Sistemas de información geográfica (GIS)	<a href="#">GML</a> (.gml) <a href="#">MIF/MID</a> (.mif/.mid)	<a href="#">Esri Shapefiles</a> (.shp & related files) <a href="#">MapInfo</a> (.tab & related files) <a href="#">KML</a> (.kml) <a href="#">Esri Geodatabase</a> (.gdb) <a href="#">Project files/Workspaces</a> (.mxd, .wor, .qgs)
Imágenes georeferenciadas	<a href="#">GeoTIFF</a> (.tif, .tiff)	<a href="#">TIFF World File</a> (.tfw & .tif, possibly with additional files) <a href="#">JPEG World File</a> (.jgw & .jpg, possibly with additional files) <a href="#">ERDAS IMAGINE File Format</a> (.img)
GIS raster	<a href="#">ASCII GRID</a> (.asc, .txt)	<a href="#">Esri GRID</a> (.grd & related files)

		<a href="#">Surfer Grid (.grd; .srf)</a> <a href="#">ERDAS IMAGINE File Format (.img)</a>
3D	<a href="#">WaveFront Object (.obj)</a> <a href="#">Polygon file format (.ply)</a> <a href="#">X3D (.x3d)</a> <a href="#">COLLADA (.dae)</a>	<a href="#">Autodesk FBX (.fbx)</a> <a href="#">Blender (.blend)</a> <a href="#">3D PDF (.pdf)</a>
RDF	<a href="#">RDF/XML (.rdf)</a> <a href="#">Trig (.trig)</a> <a href="#">Turtle (.ttl)</a> <a href="#">N-Triples (.nt)</a> <a href="#">JSON-LD</a> <a href="#">ATLAS.TI Copy bundle</a> <a href="#">NVivo Project file</a>	
Computer Assisted Qualitative Data Analysis (CAQDAS)	<a href="#">REFI-QDA (Qualitative Analysis)</a>	Data

### **1.B b) Ejemplo real**

**Ej. 1** La mayoría de los datos estarán en archivos de datos ASCII (Código estándar americano para el intercambio de información), por ejemplo, en formato de variables separadas por comas (CSV), que se pueden importar a archivos de texto enriquecido para procesamiento de textos o a hojas de cálculo. Si se utiliza software especializado, se proporcionará información sobre lectores gratuitos. Los datos se generarán en los siguientes formatos:

- Gráficos: jpeg, odg, pdf, png, pptx
- Tablas: odsu, opj, xlsx
- Texto: docx, pdf, txt
- Otros: nb, cpp

**Ej. 2** Los datos de acceso abierto serán los conjuntos completos de datos de resultados de muestras caracterizadas que se utilizan para crear figuras y gráficos en publicaciones científicas, de modo que otros investigadores puedan comparar sus resultados más fácilmente y que se puedan producir más resultados, incluidos datos históricos, más rápidamente. Los datos son tablas de valores en formato de hoja de cálculo de documento abierto (.ODS) para cantidades limitadas de datos con columnas escritas. Para cantidades mayores de datos numéricos, se utilizará codificación UTF-8, valores separados por comas en archivos de formato textual (.CSV) con valor de columna y descripción del formato de datos (FORMAT.TXT). Además, las imágenes y los archivos de datos de medición sin procesar proporcionados por los instrumentos de medición se almacenarán en una



plataforma de almacenamiento de datos interna del proyecto. Los archivos de datos y las imágenes se incluirán en los conjuntos de datos abiertos. Los datos brutos de propiedad proporcionados por los instrumentos de medición no se publicarán. Para todos los archivos publicados, se incluirá un registro de documento y un seguimiento de cambios (información de contacto del autor, estado, versión, motivo y fecha del cambio, descripción del contenido, título, origen de los datos, incluida una breve descripción de la medición y/o configuración del experimento) en un archivo de metadatos separado para cada acción de caracterización llamado METADATA.ODS.

## 1.C Especificar si se reutilizan los datos existentes y como

### **1.C a) Descripción**

Si se reutiliza un conjunto de datos (dataset), especificar la fuente de donde se ha extraído, por ejemplo, de un repositorio relevante. Si se compran o reutilizan fuentes de datos existentes, explica cómo se han tratado temas como los derechos de autor y la ley de propiedad intelectual.

Cuando se crean nuevas fuentes de datos, explicar por qué no se reutilizan las fuentes ya existentes.

### **1.C b) Ejemplo real**

**Ej. 1** Algunas de las tareas del proyecto utilizarán datos existentes en formatos hdf, txt y xlsx. Estos datos se utilizarán en la validación de los resultados del proyecto.

**Ej. 2** Los datos existentes de proyectos de I+D en curso en el ámbito del estudio <AcronymProject> sobre cables superconductores y películas delgadas servirán como base para los archivos de datos.

**Ej. 3** Las imágenes y datos seleccionados y existentes de las bases de datos de los museos asociados (<Partner1>, <Partner2>, <Partner3>...) se utilizarán en pruebas específicas, como las pruebas de almacenamiento en WP6. El último tipo de datos que se creará es la información contenida en los entregables del proyecto, que debe conservarse, hacerse accesible y transmitirse a las personas posteriores que trabajen en <AcronymProject>.

## 1.D Especificar el origen de los datos

### **1.D a) Descripción**

Indicar el origen de los datos, si éstos se generan dentro del proyecto.

Indicar la fuente de donde se han obtenido, si éstos son recolectados.

### **1.D b) Ejemplo real**

**Ej. 1** Los datos existentes procederán de varias fuentes, que incluirán: datos preexistentes del socio, datos de la literatura científica, datos de mediciones del mundo real y datos de experimentos de simulación. Los datos recopilados de propiedades nacionales permanecerán confidenciales y no se incluirán en el repositorio.

**Ej. 2** Los datos provienen de experimentos y campañas de medición realizadas por los ESR y sus colegas en los institutos beneficiarios: 1. Fase A: Cables y cintas superconductoras: <Partner1>, <Partner2>, <Partner3>... 2. Fase B: Superconductora películas delgadas: <Partner1>, <Partner2>, <Partner3>....

**Ej. 3** Estos datos han sido digitalizados en diversos proyectos anteriores.

## 1.E Indicar el tamaño esperado de los datos

### **1.E a) Descripción**

Indicar, de manera aproximada, el volumen del conjunto de datos (datasets). Hay que considerar el volumen de los datos en términos de almacenamiento, copias de seguridad, costes y acceso. Estimar el volumen de los datos en MB / GB / TB y el crecimiento gradual para asegurar que el almacenamiento y el soporte técnico adicional puedan ser proporcionados.

#### **Información Institucional**

Haga una estimación fundamentada del orden de magnitud (1 MB, 10 MB, 100 MB, 1 GB, 10 GB, 100 GB, 1 TB).

En cuanto al repositorio institucional de datos de IRSJD FAIR, el CORA – Research Data Repository (RdR) (<https://dataverse.csuc.cat>), éste permite un conjunto de datos de hasta 100 GB, con archivos individuales de hasta 10 GB.

### **1.E b) Ejemplo real**

**Ej. 1** Actualmente no se conoce el tamaño esperado de los datos, pero es probable que sea <10 GB y que los archivos individuales sean  $\leq 1$  MB.

**Ej. 2** Actualmente se desconoce el tamaño de los datos. La experiencia inicial con el almacenamiento de resultados de diferentes tipos de mediciones permitirá revisar este plan inicial de gestión de datos. Los principales tamaños de datos relevantes provendrán de imágenes como las características de muestras microscópicas que se almacenan en formato de mapa de bits de alta resolución. Sin embargo, se espera que el tamaño total del conjunto de datos para la caracterización de una sola muestra sea del orden de decenas de MB solamente.

**Ej. 3** El tamaño de los datos manejados por <AcronymProject> es bastante pequeño, como menos de 10 GB, excepto en las pruebas de la infraestructura de datos en WP6, donde el proyecto necesita experiencia en la gestión de grandes volúmenes de datos, como se explicó anteriormente.

## 1.F Describir la utilidad de los datos: para quienes pueden ser útiles

### **1.F a) Descripción**

Indicar para qué colectivo pueden ser de interés los datos.

### **1.F b) Ejemplo real**

**Ej. 1** Los datos serán aptos para su uso por otros grupos de investigación que trabajen en los siguientes temas: biogás, biometano y gases energéticos. También será útil para los comités de normas, incluido el Grupo de Trabajo de Biometano ISO/TC193/SC1/WG25, el Análisis de Gases ISO/TC 158 y los reguladores.

**Ej. 2** Dentro del Consorcio:

Los conjuntos de datos se compartirán dentro del consorcio como base de trabajo para producir publicaciones científicas, para verificar y validar los resultados mediante experimentos repetidos en diferentes lugares y como base para una documentación integral de la evaluación del rendimiento del material superconductor en el alcance del programa mundial de I+D de tecnología Future Circular Collider.

Más allá del Consorcio:

Los datos pueden ser utilizados por investigadores independientes para comprender mejor los contenidos y conclusiones de las publicaciones científicas, que basan sus hallazgos en los datos. Además, los investigadores independientes pueden utilizar los archivos para producir cifras y publicaciones, mostrando comparaciones de sus propios resultados y los resultados de <AcronymProject>. Los científicos también pueden utilizar los archivos de

datos para repetir los experimentos y mediciones para verificar y validar la investigación de <AcronymProject>. Por último, los escritores científicos y la prensa también pueden utilizar los conjuntos de datos para producir infografías de alta calidad, que demuestren el potencial de impacto de la tecnología.

**Ej. 3** Los datos de estos pilotos limitados serán útiles para usuarios e instituciones que puedan estar considerando tecnologías similares en su trabajo de digitalización y gestión de datos. Esto se aplica en particular a los experimentos realizados por el WP6, pero también a los demás. En particular, los datos digitalizados de los experimentos del WP3 harán evidente la calidad de los resultados de digitalización logrados con las nuevas tecnologías. Los datos de los experimentos del WP5 serán útiles para los museos.

## 2. Datos FAIR

---

### 2.1 Datos encontrables (incluyendo los metadatos)

#### **2.1.A Describir la descubierta y la identificación de los datos producidos/usados**

##### **2.1.A a) Descripción**

Descubierta de los datos (incluyendo los metadatos): Los metadatos deben documentar como se generaron los datos, bajo qué licencia y como se reutilizarán. Además, los metadatos ayudan a descubrir nuevos datos y proporcionan el contexto para otras interpretaciones adecuadas. Ver también el 2.1.5.

Identificar los datos y consultar el mecanismo de identificación estándar: Explicar cómo los datos y los metadatos asignan un identificador globalmente eterno y permanente (DOI, Handle...).

#### **Información Institucional**

El repositorio institucional de datos de IRSJD FAIR, el CORA – Research Data Repository (RdR) (<https://dataverse.csuc.cat>), asigna DOI para la identificación persistente y la citabilidad del conjunto de datos.

##### **2.1.A b) Ejemplo real**

**Ej. 1** El repositorio asigna identificadores Handle/DOIs para tener una identificación persistente y citabilidad del conjunto de datos o dataset.

#### **2.1.B Describir cómo se nombrarán y se estructurarán los archivos**

##### **2.1.B a) Descripción**

Describir la organización de los datos: estructura y nombre de los ficheros.

##### **2.1.B b) Ejemplo real**

**Ej. 1** La identificación del conjunto de datos del Proyecto sigue el nombre: Data\_<WPno>\_<SerialNumberDataset>\_<DatasetTitle>. Ejemplo: Datos\_WP2\_1\_01.

**Ej. 2** Los archivos se estructurarán en términos de proyecto, socio principal, identificación de publicación, figura y nombres de archivo.

## **2.1.C Proporcionar palabras claves que optimicen las posibilidades de reutilización**

### **2.1.C a) Descripción**

Indicar como se crearán las palabras clave para optimizar la recuperación y la reutilización de los datos.

Los metadatos se tienen que proporcionar estructurados, mediante l'Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) de manera que se permita el intercambio con otros repositorios. Al mismo tiempo, los metadatos proporcionados tienen que ser lo más detallados posibles, para permitir que se indexen y que los datos sean localizables y recuperables.

#### **Información Institucional**

El CORA – Research Data Repository (RdR) (<https://dataverse.csuc.cat>) proporciona campos para palabras clave de metadatos específicos para optimizar las posibilidades de recuperación y reutilización. Es posible especificar si se utiliza algún vocabulario controlado.

### **2.1.C b) Ejemplo real**

**Ej. 1** Los datos deben poder encontrarse de forma fácil, rápida e idéntica. Por lo tanto, se deben utilizar medidas exactas y estándar para identificar los conjuntos de datos. Esto puede incluir la definición y el uso de convenciones de nomenclatura, palabras clave de búsqueda, números de versión, estándares de metadatos e identificadores de datos estándar.

**Ej. 2** Todos los resultados de proyectos abiertos depositados en un repositorio proporcionarán palabras clave de búsqueda junto con sus metadatos. Las palabras clave para datos abiertos se seleccionarán de vocabularios controlados que sean adecuados para el tipo específico de datos.

## **2.1.D Describir el planteamiento a seguir para un buen control de versiones**

### **2.1.D a) Descripción**

Describir cómo se organizará el control de versiones de los ficheros.

### **2.1.D b) Ejemplo real**

**Ej. 1** Los mecanismos de control de versiones deben establecerse y documentarse antes de recopilar o generar datos.

**Ej. 2** El software de código abierto seguirá el esquema de control de versiones semántico (<ExampleURL>). Lo mismo también se puede aplicar a los conjuntos de datos. Además, todos los datos abiertos, publicaciones y software de código abierto depositados en el repositorio de Zenodo utilizarán versiones DOI. El control de versiones de DOI permite actualizar un conjunto de datos después de su publicación y citar una versión específica de un conjunto de datos o todas las versiones de un conjunto de datos.

## **2.1.E Especificar que metadatos se crearan**

### **2.1.E a) Descripción**

Describir los estándares de metadatos y como se utilizarán. Se recomienda la utilización de [estándares específicos para cada disciplina](#). En el caso que no se utilice ningún estándar, describir los metadatos que se generaran (manual o automáticamente) y cómo.

#### **Información Institucional**

El repositorio institucional de datos IRSJD FAIR, el CORA – Research Data Repository (RdR) (<https://dataverse.csuc.cat>) utiliza el estándar de la Data Documentation Initiative (DDI) para describir el conjunto de datos y cumple con los Dublin Core Metadata Terms y DataCite. Más información sobre los metadatos: <https://guides.dataverse.org/en/5.5/user/appendix.html>

### **2.1.E b) Ejemplo real**

**Ej. 1** El estándar de metadatos utilizado para describir el conjunto de datos será el Dublin Core Schema, ya que es un estándar flexible y de uso común y también es el adoptado por el repositorio europeo OpenAIRE.

**Ej. 2** Los depositantes crean manualmente los metadatos en el formulario de depósito en el repositorio.

**Ej. 3** (1) Se espera que los datos se proporcionen en formato ANSI SQL, XML o texto (ASCII). Para este conjunto de datos, se considerarán las prácticas de citación de datos y metadatos derivadas de la comunidad.

(2) No existen normas para estos registros. Una posible solución son los servidores de proyectos, como los servidores AAA. En este caso, los registros incluirían los atributos definidos por "proyecto".

**Ej. 4** Cada archivo asociado con datos irá acompañado de metadatos específicos únicos para permitir un fácil acceso y reutilización. A continuación, se presenta la forma a seguir.

**Ej. 5** Se tendrán en cuenta estándares como Dublin Core y el Registro de Metadatos (MDR) ISO/IEC 11179, que aborda cuestiones en el espacio de modelado de datos y metadatos.

**Ej. 6** Existen muchos estándares de metadatos diferentes para muchos tipos diferentes de datos y puede que no sea posible encontrar uno que se ajuste a todos los propósitos. Por lo tanto, un enfoque pragmático y factible es acordar un esquema de metadatos de catálogo mínimo y común para aquellos conjuntos de datos que se publican en catálogos y repositorios de datos públicos y utilizar extensiones de esquema específicas para tipos de datos, si es necesario.

En general, el modelo de dominio de metadatos de deposición de Zenodo, que se basa en los términos mínimos y recomendados del esquema de metadatos de DataCite, se utilizará para los datos abiertos generados por el proyecto y depositados en un repositorio apropiado.

## 2.2 Datos accesibles

### 2.2.A Especificar qué datos producidos o utilizados en el proyecto estarán disponibles de manera abierta por defecto

#### 2.2.A a) Descripción

Describir cómo se compartirán los datos, incluyendo el procedimiento para acceder a los mismos, los periodos de embargo (si los hay) y definir si el acceso será totalmente abierto o restringido a un grupo específico. Si algunos datos no pueden ser abiertos totalmente debe justificarse el porqué.

#### Información Institucional

Los principios FAIR permiten datos abiertos, restringidos o incluso -justificadamente- cerrados, y el CORA – Research Data Repository (RdR) (<https://dataverse.csuc.cat>) permite establecer periodos de acceso restringido y/o embargo para los conjuntos de datos. El grado de “apertura” dependerá de las características específicas de los datos que se publican (es decir, el grado de anonimización y los riesgos de reidentificación (GDRP), acuerdos de confidencialidad, solicitudes de patentes, etc.).

Contacte con [freerca.support-dmp@sjd.es](mailto:freerca.support-dmp@sjd.es) para contestar este ítem.

En el caso de muchos conjuntos de datos, puede crear una tabla:

Dataset	Abierto/Restringido/Cerrado	Justificación de las restricciones
...	...	...



### **2.2.A b) Ejemplo real**

**Ej. 1** Todos los datos producidos por los experimentos de WP3, WP4, WP5 y WP6, que se han descrito anteriormente, estarán disponibles públicamente. Se trata de cualquier imagen y resultado de la interpretación humana automática o asistida por computadora de los datos, que se puede ver en las imágenes. Esto no significa que también los detalles del equipo utilizado y los algoritmos utilizados en la interpretación estarán disponibles abiertamente, ya que pueden contener información privada. En Zenodo, existe la opción de proporcionar acceso abierto, acceso embargado y acceso cerrado.

**Ej. 2** Todos los datos asociados con publicaciones científicas estarán disponibles abiertamente de forma predeterminada, a menos que exista una razón específica para no publicar los datos. Conjuntos de datos que no se pueden compartir: restricciones voluntarias. Es posible que se pongan a disposición otros datos caso por caso si son relevantes para terceros.

Los siguientes datos no se pondrán a disposición del público:

- Datos obtenidos con el permiso de terceros, pero los terceros no han aceptado ponerlos a disposición del público.
- Datos que revelen la identidad de un fabricante.
- Datos que comprometan la protección de la propiedad intelectual de un socio. También se considerará el nivel de datos disponibles; por ejemplo, no se proporcionarán datos preprocesados a menos que exista una razón clara para hacerlo.

Conjuntos de datos que no se pueden compartir: razones legales y contractuales.

Todos los datos del proyecto estarán disponibles, con excepción de los datos de encuestas de mercado o de clientes, que son comercialmente sensibles y no se pueden compartir.

### **2.2.B Especificar cómo se accederá a los datos (Ej. depositando en un repositorio)**

#### **2.2.B a) Descripción**

Describir cómo se compartirán los datos, es decir, quién tendrá acceso al conjunto de datos. Se puede crear un procedimiento para que temporalmente los datos se hagan accesibles al resto de miembros del grupo, los socios del proyecto, y al público en general. Hay que

indicar si los datos se pondrán en acceso abierto y en qué plazo razonable. Una posibilidad es ofrecer juntamente con las publicaciones. Si hay que establecer un periodo de embargo, es aquí donde hay que especificarlo.

### **Información Institucional**

Los datos finales se pueden depositar en acceso abierto/restringido/cerrado en el repositorio Fair Data del IRSJD, el CORA – Research Data Repository (RdR) (<https://dataverse.csuc.cat>).

Estos datos se conservarán durante la vida útil del repositorio.

Si es necesario, se puede embargar el acceso a los conjuntos de datos durante un período de tiempo.

### **2.2.B b) Ejemplo real**

**Ej. 1** Los datos se depositarán en los sistemas de almacenamiento que serán probados por WP6, según corresponda (<Repositorio1>, <Repositorio2>, <Repositorio3>). Se proporcionarán enlaces desde el sitio web <AcronymProject> a estos sistemas de almacenamiento. Según su definición de servicio, los datos almacenados en <Repositorio1> permanecen disponibles permanentemente. No está previsto el acceso permanente a los datos de las pruebas nacionales <Repository2> y <Repository3>. Los datos de los pilotos de digitalización pueden permanecer disponibles permanentemente si se publican en el <Repositorio4>. Estos arreglos se revisarán una vez que se hayan creado los datos de los pilotos.

**Ej. 2** Una vez que se completen el procesamiento, el control de calidad, la organización, el análisis y la publicación, los datos serán accesibles mediante su depósito en repositorios de acceso abierto (por ejemplo, Zenodo).

### **2.2.C Especificar qué métodos o softwares son necesarios para acceder a los datos**

#### **2.2.C a) Descripción**

Incluir cualquier requerimiento técnico para acceder a los datos y reutilizarlos.

#### **2.2.C b) Ejemplo real**

**Ej. 1** Navegador web y/o interfaces de programación de aplicaciones (API) que ofrecen estos sistemas de almacenamiento, complementados con herramientas personalizadas desarrolladas por usuarios en dominios específicos. Zenodo proporciona servicios básicos,

robustos y rápidos. Se prevé que todo lo que esté encima de él esté en capas, y no necesariamente forme parte de la infraestructura de Zenodo. Por ejemplo, la visualización y búsqueda de múltiples imágenes debe manejarse fuera de Zenodo, por ejemplo, usando <ExampleURL> que está siendo desarrollado actualmente por <Partner1> para el Repositorio de Literatura sobre Biodiversidad específico del dominio.

**Ej. 2** Se podrá acceder a los datos utilizando el siguiente software: MS Office, Matlab, Mathematica, Origin, Open Office, Adobe Reader, Image Viewer.

## **2.2.D Especificar en la documentación qué software es necesario para acceder a los datos**

### **2.2.D a) Descripción**

Incluir la documentación del software que se necesita para acceder a los datos.

### **2.2.D b) Ejemplo real**

**Ej. 1** Si se accede a través de la API, se necesitará documentación.

**Ej. 2** Siempre que sea posible, se utilizará software estándar disponible públicamente, pero si se desarrollan herramientas de software especializadas, es decir, creadas dentro de Matlab, se proporcionará un archivo de texto corto (por ejemplo, ASCII) con el archivo de datos para explicar el software requerido.

## **2.2.E Especificar, si es posible, el software relevante**

### **2.2.E a) Descripción**

En caso de que sea un software específico, incluirlo si es posible. Por ejemplo, el código.

#### **Información Institucional**

El repositorio Fair Data del IRSJD, CORA – Research Data Repository (RdR) (<https://dataverse.csuc.cat>) se puede utilizar para depositar el código de software junto con el conjunto de datos y la documentación relevante. Además, el repositorio puede vincular el conjunto de datos al código depositado en Github.

### **2.2.E b) Ejemplo real**

**Ej. 1** Los proveedores de estos sistemas de almacenamiento ya han lanzado dicho software.

**Ej. 2** La mayoría de los programas de software están disponibles como productos comerciales o como software gratuito. Para el software desarrollado en el proyecto, el código fuente se depositará en el repositorio (por ejemplo, Zenodo).

## **2.2.F Especificar donde se depositan los datos, metadatos, documentación y el código**

### **2.2.F a) Descripción**

Indicar el repositorio en el que los datos, los metadatos, la documentación y el código se almacenarán. Puede ser en el mismo repositorio o en diferentes según el tipo de contenido. Por ejemplo, el código debería depositarse en un repositorio específico. Se puede consultar el documento de [recomendaciones para seleccionar un repositorio de datos del CSUC](#).

Es importante usar repositorios que proporcionen a los datos enlaces permanentes (DOI, Handle) para facilitar el descubrimiento y la citación.

### **Información Institucional**

Los datos generados en el marco de este proyecto y la información que respalda la preservación y reutilización de la información se depositarán en CORA – Research Data Repository (RdR) (<https://dataverse.csuc.cat/>). Es un repositorio de datos multidisciplinar de confianza que permite a las universidades catalanas publicar conjuntos de datos de investigación en modo FAIR y siguiendo las directrices de la EOSC. Este repositorio está diseñado para recopilar, difundir y proporcionar acceso abierto persistente, confiable y a largo plazo a los datos. El CORA – Repositorio de datos de investigación (RdR) asigna DOI para la identificación persistente y la identificabilidad del conjunto de datos.

### **2.2.F b) Ejemplo real**

**Ej. 1** Los datos se depositarán en los sistemas de almacenamiento que serán probados por el WP6, según corresponda (OSC nacional, EUDAT, Zenodo). Se proporcionarán enlaces desde el sitio web <AcronymProject> a estos sistemas de almacenamiento.

**Ej. 2** Los datos y los metadatos, la documentación y el código asociados se depositarán en el repositorio de acceso abierto llamado Zenodo o en el Repositorio de acceso abierto (<ExampleURL>).

## 2.2.G Especificar si se han examinado los requisitos del repositorio

### 2.2.G a) Descripción

Indicar si se han explorado adecuadamente cuáles son los requisitos del repositorio identificado.

#### Información Institucional

El IRSJD ha firmado un acuerdo con el Consorci de Serveis Universitaris de Catalunya (CSUC), proveedor de CORA – Research Data Repository (RdR) (<https://dataverse.csuc.cat/>), para utilizarlo como repositorio institucional.

Como ya se ha mencionado anteriormente, es un repositorio de datos multidisciplinar de confianza que permite a las universidades catalanas publicar conjuntos de datos de investigación en modo FAIR y siguiendo las directrices de la EOSC.

### 2.2.G b) Ejemplo real

**Ej. 1** Ya hemos explorado los acuerdos apropiados con los servicios nacionales de nube en Finlandia (CSC), EUDAT a través del trabajo del piloto <AcronymProject> y Zenodo a través del trabajo de la comunidad del Repositorio de Literatura <Disciplinary>.

**Ej. 2** Sí, el Repositorio de Acceso Abierto es funcional y etiqueta correctamente conjuntos de datos con un esquema de metadatos compatible con DataCite).

## 2.2.H Especificar cómo se da acceso si hay restricciones de uso

### 2.2.H a) Descripción

En caso de que el acceso a los datos sea restringido a todo el público por un motivo justificado, especificar qué datos serán accesibles a socios individuales, a todos los socios o bajo solicitud. Especificar los procedimientos para solicitar el acceso a los datos restringidos y en qué condiciones se concederán. Además, especificar también si habrá más restricciones en un período de tiempo.

#### Información Institucional

Si, según la respuesta dada en 2.2B, los datos finales deben depositarse en acceso restringido o cerrado en el repositorio Fair Data del IRSJD, el CORA – Research Data Repository (RdR) (<https://dataverse.csuc.cat/>), o va a ser embargado por un periodo de tiempo, explique los motivos.

Este ítem también puede responderse de acuerdo con 2.2J y 2.4B, donde se indica la licencia para la reutilización. Si la licencia es CC-BY-NC o equivalente, no se permite su reutilización con fines comerciales.

## **2.2.H b) Ejemplo real**

**Ej. 1** No existen restricciones de uso, excepto cuando se haya elegido la licencia CC BY-NC. <AcronymProject> debe abordar cuestiones de datos confidenciales (por ejemplo, ubicación de plantas protegidas), pero <AcronymProject> evitará trabajar con datos confidenciales. Si se reciben datos personales en cuestionarios que recibirá <AcronymProject>, dichos datos se anonimizarán antes de estar disponibles fuera del proyecto.

**Ej. 2** No existen restricciones sobre el uso de los datos publicados, pero los usuarios deberán reconocer al consorcio y la fuente de los datos en cualquier publicación resultante.

## **2.2.I Especificar si es necesario un comité de acceso a los datos**

### **2.2.I a) Descripción**

Especificar por qué es o no necesario un comité de acceso a los datos.

### **2.2.I b) Ejemplo real**

**Ej. 1** Debido a la pequeña escala de estos experimentos, no es necesario crear un comité de acceso a los datos.

**Ej. 2** Este consorcio contará con un comité de acceso a datos. Su competencia será seleccionar, caso por caso, los datos que serán accesibles abiertamente. Se considerarán los aspectos éticos y de seguridad de los datos, incluidos los requisitos de propiedad intelectual. Si es necesario, se retendrán algunos o todos los datos de una posible publicación. Esto se decidirá en consulta con los socios pertinentes.

## **2.2.J Describir correctamente las condiciones de acceso**

### **2.2.J a) Descripción**

Describir cuáles son las condiciones de acceso definidas por el repositorio de vuestra elección (indicando el URL de donde proviene la información). Por ejemplo: con una licencia legible por máquina.

### **Información Institucional**

El CORA – Research Data Repository (RdR) (<https://dataverse.csuc.cat>) asigna una licencia CC0 por defecto, pero se permiten otras licencias personalizadas.

El IRSJD sugiere una licencia CC-BY o una licencia CC-BY-NC para los datos siempre que la unidad de innovación necesite limitar la reutilización debido a problemas de transferibilidad. Consulte con

la unidad de Innovación sobre posibles problemas antes de asignar una licencia al conjunto de datos.

Los metadatos siempre tendrán una licencia CC0 asignada.

### **2.2.J b) Ejemplo real**

**Ej. 1** Se utilizarán las Licencias Creative Commons soportadas por GBIF. Estos incluyen CC0, CC-BY y CC BY-NC (consulte <ExampleURL>). Zenodo admite una gran variedad de licencias legibles por máquina, ampliamente utilizadas y específicas de dominio. El propietario de los datos determinará cuál de estas licencias se utilizará cuando los datos se publiquen en los repositorios <AcronymProject>. Sin embargo, la recomendación del proyecto es elegir CC0 para datos y CC-BY para medios y evitar CC-BY-NC, que tiene problemas en algunas jurisdicciones nacionales.

**Ej. 2** Sí, Zenodo proporciona condiciones de acceso bien descritas (consulte <http://about.zenodo.org/policies/>).

## **2.2.K Determinar la identidad de las personas que acceden a los datos**

### **2.2.K a) Descripción**

Describir el procedimiento establecido por el repositorio para determinar la identidad de la persona que accede a los datos.

Los usuarios deben registrarse para utilizar el repositorio institucional para publicaciones. En el tratamiento de datos personales recuerda que deberás respetar la legislación específica europea y española y la normativa del Instituto. Por favor, consulte la información específica proporcionada por el IRSJD en la página web Investigación y datos personales.

Para solicitar una revisión ética del proyecto, consultar al Comité de Ética en Investigación (CEIm-IRSJD).

### **Información Institucional**

El CORA – Repositorio de datos de investigación (RdR) (<https://dataverse.csuc.cat>) tiene una funcionalidad de libro de visitas opcional. Esto nos permite saber quién accedió al conjunto de datos. Además, el acceso al conjunto de datos puede restringirse mediante un formulario de contacto previo.

## **2.2.K b) Ejemplo real**

**Ej. 1** No se comprobará directamente la identidad de la persona que accede a los datos. Sin embargo, esperamos que los usuarios sigan las normas estándar de citas científicas y el uso de los datos en este contexto será rastreado a través de citas científicas.

## 2.3 Datos interoperables

### **2.3.A Evaluar la interoperabilidad de los datos. Especificar qué datos y qué vocabularios de metadatos, estándares o metodologías se seguirán para facilitar la interoperabilidad**

#### **2.3.A a) Descripción**

Explicar qué datos y qué vocabularios de metadatos, estándares o metodologías se siguen para facilitar la interoperabilidad. La interoperabilidad de los datos del proyecto permite el intercambio i la reutilización de datos entre investigadores, instituciones, organizaciones, países, etc. Adherirse a los estándares de formatos que sean, en la medida de lo posible, compatibles con programas i aplicaciones abiertas.

#### **Información Institucional**

Además de explicar los formatos de datos y los estándares utilizados, explique la interoperabilidad del conjunto de datos de la siguiente manera:

El CORA – Research Data Repository (RdR) (<https://dataverse.csuc.cat>) sigue el modelo Open Archives Initiative, que permite la interoperabilidad con el protocolo de transmisión de metadatos OAI-PMH (Open Archive Initiative - Protocol for Metadata Harvesting). Este protocolo permite la visibilidad de los documentos de diferentes plataformas y recopiladores: Google Scholar, BASE, CORE, etc. Este repositorio de datos es compatible con OpenAIRE y cumple con todos los requisitos de metadatos exigidos por la Comisión Europea.

Se proporcionan ID persistentes para cada documento (DOI) y los identificadores de autor (ORCID) se incluyen en los metadatos.

El estándar de metadatos utilizado para describir el conjunto de datos es el esquema de metadatos de DDI compatible con Dublin Core, un estándar flexible y de uso común que también es adoptado por el repositorio europeo OpenAIRE.

#### **2.3.A b) Ejemplo real**

**Ej. 1** Los datos producidos en el proyecto serán interoperables ya que los conjuntos de datos se adherirán a formatos estandarizados: ASCII, txt, csv, xml, tiff. Si no se puede utilizar MS Office, el visor de PDF o el visor de imágenes, se proporcionará un archivo de



texto (ASCII) con el conjunto de datos que explica dónde se puede obtener un lector gratuito.

### **2.3.B Especificar si se usarán vocabularios estandarizados para todos los tipos de datos del conjunto de datos (dataset), para facilitar la interoperabilidad entre disciplinas**

#### **2.3.B a) Descripción**

Especificar si se usan vocabularios estandarizados para los distintos tipos de datos. Si no es el caso, indicar la correspondencia entre las clasificaciones más usadas en tu especialidad.

#### **Información Institucional**

Además de la información proporcionada en el apartado 2.3.A, indicar en este apartado si se usaran vocabulario NO estandarizados (no comunes, o generados *ad-hoc* para el proyecto), y en dicho caso si se proveerán tablas de correspondencia entre estos y otros vocabularios comúnmente aceptados.

#### **2.3.B b) Ejemplo real**

**Ej. 1** Otro tipo de datos han sido registrados siguiendo las codificaciones internas, claramente especificadas en el fichero.

## 2.4 Datos reutilizables

### **2.4.A Especificar cómo se asignan las licencias para permitir la máxima la reutilización de los datos**

#### **2.4.A a) Descripción**

Si los datos se ponen a disposición de otros investigadores y del público en general, especificar el grado de reutilización permitido. Se establecerá el grado de reutilización a través de licencias. La CE propone el uso de licencias Creative Commons CC-BY o CC0, pero existen otras.

#### **Información Institucional**

El CORA – Research Data Repository (RdR) (<https://dataverse.csuc.cat>) asigna una licencia CC0 por defecto, pero se permiten otras licencias personalizadas.

El IRSJD sugiere una licencia CC-BY o una licencia CC-BY-NC para los datos siempre que la unidad de innovación necesite limitar la reutilización debido a problemas de transferibilidad. Consulte con

la unidad de Innovación sobre posibles problemas antes de asignar una licencia al conjunto de datos.

Los metadatos siempre tendrán una licencia CC0 asignada.

#### **2.4.A b) Ejemplo real**

**Ej. 1** Los entregables asociados al conjunto de datos tienen licencia a través de una licencia Todos los derechos reservados, ya que son documentos de trabajo que no están destinados a ser reutilizados. Sin embargo, la base de datos debe compartirse como un posible conjunto de datos reutilizable. Por este motivo, al depositarse en el repositorio se solicitará una licencia de Atribución-No Comercial (by-nc). Los datos están actualmente disponibles para su reutilización en el sitio web del proyecto y también se podrán encontrar y reutilizar a través del repositorio de depósito final (el institucional o Zenodo) y desde OpenAire, el último al final del proyecto.

**Ej. 2** Siempre que sea posible, los datos se compartirán inmediatamente después de la producción siguiendo la Licencia Internacional Creative Commons 4.0 con Atribución (CC4BY). En algunos casos, los datos de prueba de datos experimentales solo estarán disponibles después de la finalización del proyecto o la publicación de los resultados, lo que ocurra primero, y se compartirán utilizando la misma licencia CC4BY.

### **2.4.B Especificar cuando los datos estarán disponibles para su reutilización. Si procede, especificar porqué y durante qué período será necesario embargar los datos**

#### **2.4.B a) Descripción**

Describir cuando los datos estarán disponibles para su reutilización. Se podrá elaborar un procedimiento para que los datos puedan ser accesibles a otros miembros del grupo, a los socios de proyecto y, posteriormente al público en general. Se debe indicar si los datos serán de acceso abierto y cuando. Una posibilidad es hacerlos públicos junto con las publicaciones. Si se requieren periodos de embargo, se debe indicar en este apartado.

#### **2.4.B b) Ejemplo real**

**Ej. 1** Los datos seguirán siendo reutilizables una vez finalizado el proyecto por cualquier persona interesada en ellos, sin restricciones de acceso ni de tiempo.

**Ej. 2** Los datos recopilados en este proyecto estarán disponibles para su reutilización una vez finalizado el experimento. Como se ha señalado anteriormente, por motivos de ventajas

competitivas podrá aplicarse un embargo de datos, incluida la realización de una tesis doctoral, en cuyo caso se mantendrá un embargo de tres años.

**2.4.C Especificar si los datos producidos y/o utilizados en el proyecto podrán ser utilizados por terceros, en particular, una vez finalizado el proyecto. Si se restringe la reutilización de algunos datos se debe concretar los motivos**

**2.4.C a) Descripción**

En principio, los datos deben estar a disposición de otros investigadores y del público en general con las menos restricciones posibles. Sin embargo, existen motivos para no compartir los datos: razones éticas, protección de datos personales, implicación de derechos de propiedad intelectual y/o industrial, intereses comerciales, etc. Si procede, se debe especificar los motivos por los cuales no se compartirán algunos conjuntos de datos.

**2.4.C b) Ejemplo real**

**Ej. 1** Derechos de propiedad intelectual y cuestiones de privacidad. Las actividades de acceso e intercambio de datos se implementarán rigurosamente de conformidad con las normas y reglamentos de privacidad y recopilación de datos, tal como se aplican a nivel nacional y en la UE, así como con las normas H2020. Los datos brutos recopilados a través de entrevistas con fuentes externas del consorcio pueden estar disponibles para todo el consorcio o socios específicos previa autorización de los propietarios. Este tipo de datos no estarán disponibles para el público. Los resultados del proyecto estarán disponibles públicamente sobre la base de los DPI, como se describe en el Acuerdo de Consorcio.

**Ej. 2** El conjunto de datos completo será confidencial y sólo los miembros del consorcio tendrán acceso a él. Además, si se decide hacer que partes específicas de ellos (por ejemplo, metadatos, estadísticas, etc.) sean de amplio acceso, se creará un portal de gestión de datos que deberá proporcionar una descripción del conjunto de datos y un enlace a una sección de descarga. Por supuesto, estos datos serán anonimizados para no tener ninguna posible correlación e identificación de cuestiones éticas con su publicación y difusión.

**Ej. 3** Cada conjunto de datos archivados tendrá su propio ID de repositorio permanente y será fácilmente accesible. Esperamos que la mayoría de los datos generados estén disponibles sin restricciones y solo se restringirán los conjuntos de datos sujetos a derechos de propiedad intelectual y cuestiones de confidencialidad. Cuando este sea el caso, los

acuerdos se realizarán basándose en los conjuntos de datos individuales. Las solicitudes de uso de los datos por parte de terceros serán aprobadas por el consorcio del proyecto.

## **2.4.D Describir el proceso que garantiza la calidad de los datos**

### **2.4.D a) Descripción**

Describir el proceso que garantiza la calidad de los datos. Especificar cómo i cuando se llevarán a cabo las evaluaciones internas.

La calidad de los datos se puede asegurar a través de diferentes medidas. Estas incluyen la validación de la muestra, las réplicas y la comparación con resultados de estudios similares y el control de la distorsión sistemática.

### **2.4.D b) Ejemplo real**

**Ej. 1** La calidad del conjunto de datos está garantizada por el funcionamiento de la plataforma.

**Ej. 2** La calidad de los datos está garantizada por diferentes medidas. Estos incluyen la validación de la muestra, la replicación y comparación con resultados de estudios similares y el control de la distorsión sistemática.

**Ej. 3** La garantía y el control de la calidad de los datos son fundamentales y la razón de ser de este proyecto. Alrededor del 80% de los esfuerzos realizados en nuestros Centros Temáticos se dirigen al aseguramiento de la calidad de los datos.

**Ej. 4** Para nuestra recopilación de datos de investigación, el control de calidad de los datos puede ocurrir en varias etapas durante el proceso de garantía de calidad. Se necesita un control de calidad inicial a nivel local y en las primeras etapas del proceso de recolección. Se llevarán a cabo controles adicionales en una etapa posterior del ciclo de vida de los datos. El control de calidad final de los metadatos se lleva a cabo durante su introducción en el IMIS. El control de calidad inicial de los datos, durante la recopilación de datos, es responsabilidad principal del creador/propietario de los datos del proyecto, quien debe garantizar que los datos registrados reflejen los hechos, respuestas, observaciones y eventos reales. La calidad de los métodos de recopilación de datos utilizados influye fuertemente en la calidad de los datos, y documentar en detalle cómo se recopilan proporciona evidencia de dicha calidad. También pueden ocurrir errores durante la entrada de datos. Los datos se digitalizan, se transcriben, se introducen en una base de datos o en una hoja de cálculo, o se codifican. Aquí, la calidad está garantizada mediante

procedimientos estandarizados y consistentes para la entrada de datos con instrucciones claras.

## **2.4.E Especificar los plazos en los que los datos podrán ser reutilizados**

### **2.4.E a) Descripción**

Indicar el tiempo de espera para que los datos puedan ser reutilizados.

### **2.4.E b) Ejemplo real**

**Ej. 1** Para su reutilización, los datos se almacenarán en la página web o en un sistema de repositorio cuando se implemente durante al menos diez años.

**Ej. 2** Los datos seguirán siendo reutilizables hasta que Zenodo suspenda el(los) conjunto(s) de datos (es decir, con garantía por un mínimo de 20 años).

**Ej. 3** La infraestructura tiene un horizonte temporal de al menos <Número> años, los datos seguirán siendo útiles y utilizables más allá de ese período. Por ejemplo, ahora las series temporales generadas desde <Año> de concentraciones de CO2 en <Lugar> todavía se utilizan.

## 3. Asignación de recursos

---

### 3.A Estimar el coste para hacer que los datos sean FAIR en el proyecto

#### **3.A a) Descripción**

Indicar el coste aproximado de hacer que los datos sean FAIR y como pensar en taparlos: con costos directos e indirectos relacionados con el almacenamiento, archivo, reutilización, seguridad, etc.

#### **Información Institucional**

No hay costes para el presupuesto de su proyecto por utilizar un depósito en el repositorio institucional, ya que es administrado por la Unidad de Gestión del Conocimiento del Instituto. Esto incluye todos los costes relacionados con el almacenamiento y archivo a largo plazo, la reutilización, la seguridad, la conservación, etc.

\*Si no utiliza el repositorio institucional, puede haber costes en el Presupuesto del Proyecto.

\*Si tiene un presupuesto dedicado para hacer que los datos sean FAIR durante el Proyecto (es decir, *un data steward* dedicado, sistemas de almacenaje, etc), explíquelo en esta sección.

#### **3.A b) Ejemplo real**

**Ej. 1** No hay costos asociados a los mecanismos descritos para hacer que la base de datos sea JUSTA y preservada a largo plazo.

**Ej. 2** Los costos para depositar el conjunto de datos en el proyecto y los recursos posteriores necesarios para que el conjunto de datos esté disponible públicamente se han incluido dentro de paquetes de trabajo específicos dentro del proyecto.

### 3.B Especificar como se prevé cubrir el coste

#### **3.B a) Descripción**

Indicar como se quiere cubrir el coste para hacer que los datos sean FAIR, incluyendo los costes adicionales de archivo y conservación.

#### **3.B b) Ejemplo real**

**Ej. 1** <AcronymProject> está gestionado y respaldado por un equipo de expertos y es gratuito.

**Ej. 2** El coste de conservación de la base de datos será asumido por el <Partner1>.

**Ej. 3** (1) Probablemente se asignará una unidad de disco duro dedicada al conjunto de datos. Actualmente no se prevén costes relacionados con su conservación.

(2) El coste será cubierto por el instituto anfitrión local en el marco del proyecto.

(3) El costo será cubierto por el instituto de alojamiento local como parte del mantenimiento estándar del sistema de red.

### 3.C Especificar quién es el responsable de la gestión en cada proyecto

#### **3.C a) Descripción**

Explicar las responsabilidades de la gestión de los datos en cada proyecto.

#### **Información Institucional**

[Si se utiliza CORA RDR institucional] La Unidad de Gestión del Conocimiento del Instituto gestiona el repositorio de datos del FAIR, el CORA – Research Data Repository (RdR) (<https://dataverse.csuc.cat>).

Explicar todos los demás actores en la gestión de datos durante el Proyecto.

#### **3.C b) Ejemplo real**

**Ej. 1** El coordinador del proyecto tiene la responsabilidad final de la gestión de datos en el proyecto y, por tanto, de la gestión de la plataforma Marketplace.

**Ej. 2** Cada socio tiene que respetar las políticas establecidas en este PGD. Los conjuntos de datos deben crearse, gestionarse y almacenarse de forma adecuada y de conformidad con la legislación aplicable.

- El Coordinador del Proyecto tiene la responsabilidad particular de garantizar que los datos compartidos a través del sitio web estén fácilmente disponibles, pero también que se realicen copias de seguridad y que los datos de propiedad estén protegidos.
- Líder del WP1, garantizará la integridad y compatibilidad del conjunto de datos para su uso durante la vida del proyecto por parte de diferentes socios.
- La validación y registro de conjuntos de datos y metadatos es responsabilidad del socio que genera los datos en el WP.
- Hacer copias de seguridad de los datos para compartirlos a través de repositorios de acceso abierto es responsabilidad del socio que posee los datos. Quality control of

these data is the responsibility of the relevant WP leader, supported by the Project Coordinator.

### 3.D Describir los recursos para la preservación a largo plazo

#### **3.D a) Descripción**

Indicar como se prevé preservar a largo plazo y quién decide qué datos se conservarán y durante cuánto tiempo.

#### **Información Institucional**

Los elementos se conservarán durante la vida útil del repositorio de datos institucional de FAIR, el CORA – Research Data Repository (RdR) (<https://dataverse.csuc.cat>).

#### **3.D b) Ejemplo real**

**Ej. 1** En cuanto a la cuestión de la conservación de datos a largo plazo, el consorcio aún no ha llegado a ningún acuerdo específico. Sin embargo, con un alto grado de confianza, se puede confirmar que es el coordinador del proyecto, con la ayuda de los recursos locales de <AcronymProject>, quien desempeñará el papel principal en esta tarea.



## 4. Seguridad de los datos

### 4.A Tratamiento de recuperación de datos, así como el almacenaje y la transferencia de datos confidenciales

#### 4.A a) Descripción

Describir las medidas adoptadas respecto la seguridad de los datos. Incluye el almacenaje y las copias de seguridad, la preservación a largo plazo y cómo se mantendrán los datos seguros y protegidos.

#### Servicios de Almacenamiento Institucional: Esplugues de Llobregat

Los datos recopilados se almacenarán en un <Servicio de almacenamiento> sujeto a copias de seguridad periódicas controladas por personal especializado.

El departamento de Sistemas realiza operaciones por tipo: de misión crítica (datos de usuario, máquinas virtuales, resultados científicos, etc.) y estáticas (conjuntos de datos científicos, archivos intermedios, etc.).

El contenido se revisará periódicamente para preservar su integridad, seguridad y durabilidad.

Servicios de almacenamiento:

- Redcap
  - Type of server: Virtual
  - Location: Hospital Sant Joan de Déu Data Center (Pg. de Sant Joan de Déu, 2, 08950 Esplugues de Llobregat, Barcelona, Spain)
  - Backups: diario, seminal, mensual y annual en el Hospital Sant Joan de Déu Data Center 2 (Pg. de Sant Joan de Déu, 2, 08950 Esplugues de Llobregat, Barcelona, Spain)
  - Recovery: cualquier fichero eliminado puede ser recuperado a partir de los backups.
  - Access: acceso nominal para cada usuario previa solicitud, si es externo deberá ser solicitado por alguien del hospital.
- Filesystem ("Recursos" folders)
  - Type of server: Virtual
  - Location: Hospital Sant Joan de Déu Data Center (Pg. de Sant Joan de Déu, 2, 08950 Esplugues de Llobregat, Barcelona, Spain)
  - Backups: diario, seminal, mensual y annual en el Hospital Sant Joan de Déu Data Center 2 (Pg. de Sant Joan de Déu, 2, 08950 Esplugues de Llobregat, Barcelona, Spain)
  - Recovery: cualquier fichero eliminado puede ser recuperado mediante el software DataProtector.
  - Access: acceso nominal
- Microsoft Office365/Onedrive
  - Type of server: Cloud
  - Location: Azure Microsoft (Europa)
  - Backups: diario, seminal, mensual y annual en el Hospital Sant Joan de Déu Data Center 2 (Pg. de Sant Joan de Déu, 2, 08950 Esplugues de Llobregat, Barcelona, Spain)
  - Recovery: disponemos de cabina Sinology Onpremise que realiza backups de todo el servidor.
  - Access: acceso nominal mediante MFA (autenticación multi factor).
- Genetics
  - Type of server: Physical

- Location: Hospital Sant Joan de Déu Data Center (Pg. de Sant Joan de Déu, 2, 08950 Esplugues de Llobregat, Barcelona, Spain)
  - Backups: bajo demanda. Unidades de cita en el Hospital Sant Joan de Déu Data Center 1 (Pg. de Sant Joan de Déu, 2, 08950 Esplugues de Llobregat, Barcelona, Spain)
  - Recovery: bajo demanda, restauración a partir de las cintas de backup.
  - Access: sólo tiene acceso el departamento de Bioinformática.
- uPad
    - Type of server: Cloud
    - Location: Amazon Web Services (AWS)
    - Backups: AWS Backup (pendiente implementación)
    - Recovery: AWS Backup (pendiente implementación)
    - Access: pendiente

Estos procedimientos están diseñados, establecidos y aplicados con el fin de dar pleno cumplimiento a los datos personales según lo dispuesto en la Directiva 95/46/CE (Reglamento General de Protección de Datos) y demás legislación nacional y normativa institucional vigente. Los miembros del equipo de investigación tendrán un nivel de acceso adecuado según su rol en el proyecto.

### **Servicios de Almacenamiento Institucional: Parc Sanitari Sant Boi de Llobregat**

Los datos recopilados por el grupo de investigación para el Proyecto se almacenarán en el <Servicio de Almacenamiento>, que está sujeto a una copia de seguridad periódica controlada por el personal de Sistemas.

El departamento de Sistemas realiza operaciones por tipo: de misión crítica (datos de usuario, máquinas virtuales, resultados científicos, etc.) y estáticas (conjuntos de datos científicos, archivos intermedios, etc.).

El contenido se revisará periódicamente para preservar su integridad, seguridad y durabilidad.

Servicios de Almacenamiento:

- Redcap
  - Type of server: Virtual
  - Location: Parc Sanitari Sant Joan de Déu - Data Center 1 (Carrer Doctor Antoni Pujades 42, Sant Boi de Llobregat, Barcelona, Spain)
  - Backups: diario, semenal, mensual y anual en el Parc Sanitari Sant Joan de Déu - Data Center 2 (Camí Vell de la Colònia 25, Sant Boi de Llobregat, Barcelona, Spain)
  - Recovery: se dispone de plataforma de backup de disco Dell Data Domain On-Premise, con el software Dell Avamar para realizar copias de seguridad y restauraciones.
  - Access: Acceso nominal para cada usuario previa solicitud, si es externo deberá ser solicitado por alguien del hospital.
- Filesystem
  - Type of server: cabina de almacenamiento físico Huawei con HA
  - Location: Parc Sanitari Sant Joan de Déu - Data Center 1 (Carrer Doctor Antoni Pujades 42, Sant Boi de Llobregat, Barcelona, Spain) and Data Center 2 (Camí Vell de la Colònia 25, Sant Boi de Llobregat, Barcelona, Spain)
  - Backups: diario, semenal, mensual y anual en el Parc Sanitari Sant Joan de Déu - Data Center 2 (Camí Vell de la Colònia 25, Sant Boi de Llobregat, Barcelona, Spain)
  - Recovery: Cualquier archivo eliminado se puede recuperar desde la plataforma de restauración local Dell Data Domain utilizando el software Veeam Backup.
  - Access: acceso nominal

- Microsoft Office365/OneDrive
  - Type of server: Cloud
  - Location: Azure Microsoft (Europa)
  - Backups: diario, semenal, mensual y annual en el Parc Sanitari Sant Joan de Déu - Data Center 1 (Carrer Doctor Antoni Pujades 42, Sant Boi de Llobregat, Barcelona, Spain)
  - Recovery: Contamos con una cabina de disco de almacenamiento local de Sinology con copias de seguridad/restauraciones para cuentas de O365, incluido OneDrive.
  - Access: Nominal access with MFA (autenticación multi factor)

Estos procedimientos están diseñados, establecidos y aplicados con el fin de dar pleno cumplimiento a los datos personales según lo dispuesto en la Directiva 95/46/CE (Reglamento General de Protección de Datos) y demás legislación nacional y normativa institucional vigente. Los miembros del equipo de investigación tendrán un nivel de acceso adecuado según su rol en el proyecto.

### **Preservación a largo plazo (repositorio institucional de datos FAIR)**

La siguiente lista describe la configuración de seguridad de todos los datos finales depositados en el repositorio institucional de datos FAIR CORA RDR:

- Versiones: los archivos de datos están versionados. Los registros no están versionados. Los datos cargados se archivan como un paquete de información de envío. Se generan derivados de archivos de datos, pero el contenido original nunca se modifica. Los registros se pueden retirar de la vista del público; sin embargo, los archivos y registros de datos se conservan.
- Réplicas: Todos los ficheros de datos se almacenan en el Centro del CSUC, principalmente en Barcelona, con réplicas en el Consorcio Madroño de Madrid. Los archivos de datos se guardan en múltiples réplicas en un sistema de archivos distribuido, del que se realiza una copia de seguridad en cinta todas las noches.
- Período de retención: los elementos se conservarán durante la vida útil del repositorio.
- Preservación funcional: El RDR no hace promesas de usabilidad y comprensibilidad de los objetos depositados a lo largo del tiempo.
- Preservación de archivos: los archivos de datos y metadatos se respaldan todas las noches y se replican en múltiples copias en el sistema en línea.
- Fijeza y autenticidad: todos los archivos de datos se almacenan junto con una suma de verificación MD5 del contenido del archivo y el archivo tabular se almacena con huella digital numérica universal ([JNE](#)).
- Los archivos se comparan periódicamente con sus sumas de verificación para garantizar que el contenido del archivo permanezca constante.
- Planes de sucesión: En caso de cierre del repositorio, RDR ha garantizado la migración de todo el contenido a repositorios institucionales y/o temáticos alternativos adecuados.

#### **4.A b) Ejemplo real**

**Ej. 1** Los datos recopilados por el grupo de investigación para el Proyecto se digitalizarán y almacenarán en el <StorageService> de la Universidad. El Acuerdo <StorageService> incluye términos y condiciones que cumplen con la Ley de Protección de Datos de la UE y las normas y regulaciones del Comité Nacional de Bioética. <StorageService> proporciona cláusulas modelo de la UE en acuerdos, posee certificaciones <ISONumber> y <ISONumber> y opera sus centros de datos dentro del Espacio Económico Europeo. Sólo los investigadores que trabajen para el Proyecto tendrán acceso a los datos, utilizando su nombre de usuario y contraseñas para acceder a los archivos. El grupo de investigación también mantendrá una copia de seguridad en un disco duro que se almacenará en el edificio de la escuela en <Room>. Sólo el investigador principal designado DPO para el proyecto tendrá acceso al disco duro. La seguridad está garantizada para el edificio y la habitación las 24 horas.

**Ej. 2** La confidencialidad e integridad de los datos se implementan en varios niveles:

- Los datos en reposo, almacenados en el repositorio de datos XXX, están protegidos contra el acceso no autorizado mediante el inicio de sesión estándar de la UE (antigua autenticación ECAS). Se otorgarán niveles de acceso adecuados mediante la creación de grupos.
- Los datos en tránsito están protegidos mediante mecanismos seguros de transferencia de datos, como TLS 1.2.2 (Transport Layer Security)
- El acceso a los datos se registra mediante un mecanismo de registro a prueba de manipulaciones integrado en <NameSoftware>, los archivos de registro se almacenan dentro de un sistema de archivos cifrado y se configuran solo para agregar.
- Los socios del consorcio impondrán una política estricta a todos los empleados, compañeros de trabajo, subcontratistas... que tengan acceso a los datos. Esta política incluirá, entre otras cosas, permitir copias en dispositivos locales solo durante el procesamiento de los datos con garantía
  - borrado después de ser procesado
  - ampliar las políticas de control de acceso a las copias locales
  - cláusulas contractuales
  - acuerdo con los términos y condiciones antes de otorgar el acceso
- Los datos serán seudonimizados hasta el nivel que no interfieran con la calidad de la investigación.

Por último, se mejorará la conciencia sobre la privacidad y seguridad de los datos (por ejemplo, asistiendo a un seminario web sobre este tema antes de que se le conceda acceso al repositorio; asistir a este seminario web será obligatorio al menos una vez al año durante el transcurso del proyecto).

## 5. Aspectos éticos

---

### 5.A Problemas éticos o legales que pueden afectar la recogida e intercambio de datos

#### **5.A a) Descripción**

Describir los posibles problemas éticos durante la recogida, el almacenaje, el procesamiento y el archivo de los datos, junto a los procedimientos de aprobación ética relacionados con el proyecto. Es importante destacar en este punto cualquier aspecto mencionado en el artículo 34 de los acuerdos de financiamiento "[Article 34 — Ethics and research integrity](#)".

Si las actividades de investigación involucran a niños, pacientes, población vulnerable, uso de células madre embrionarias, cuestiones de privacidad y protección de datos o investigación en animales y primates, se deben cumplir los principios éticos y la legislación nacional pertinente de la Unión Europea e internacional, y en este punto es donde debe tratarse.

#### **Información Institucional**

##### **Reglamento General de Protección de Datos (GDPR)**

El IRSJD tiene la obligación de preservar la confidencialidad de los datos personales y clínicos de los sujetos del estudio en base al GDPR (UE) 2016/679 del Parlamento Europeo y del Consejo de 27 de abril de 2016 sobre la protección de las personas físicas con en lo que respecta al tratamiento de datos personales y a la libre circulación de dichos datos, y por la que se deroga la Directiva 95/46/CE y cualquier otra disposición aplicable en materia de protección de datos. Cualquier tratamiento será debidamente analizado y revisado por cada Delegado de Protección de Datos correspondiente. Los datos generados se desvincularán de una persona identificable y se tomarán las precauciones necesarias para garantizar su seguridad y la seguridad de los datos asociados.

##### **Ley nacional española (LOPD-GDD)**

El IRSJD se encuentra bajo la legislación española en materia de protección de datos, y la ley aplicable es la Ley Orgánica 3/2018, de 5 de diciembre, de protección de datos de carácter personal (LOPD-GDD); La norma vigente en los preceptos no queda derogada por el Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de dichos datos, y por la que se deroga la Directiva 95/46/CE (GPDR).

##### **Delegado de Protección de Datos (DPO)**

El IRSJD cuenta con un Delegado de Protección de Datos que participará en el proyecto, en particular en lo que respecta a las cuestiones del RGPD. El Delegado de Protección de Datos contribuirá al seguimiento regular y sistemático de la recopilación, procesamiento, análisis y almacenamiento de datos. <Si corresponde> En los demás sitios participantes involucrados en la

recopilación y el procesamiento de datos, se nombrará un Delegado de Protección de Datos. Los contactos del Delegado de Protección de Datos en cada sitio participante estarán disponibles para todos los sujetos involucrados en el proyecto y se agregarán a este documento cuando estén disponibles.

### **Formulario de consentimiento**

El Comité de Ética de la Investigación con medicamentos (CEIm) es un Comité Ético de Investigación con Medicamentos afiliado que evalúa proyectos de investigación considerando, entre otros, los siguientes criterios genéricos establecidos por las normas de buena práctica clínica y por la legislación vigente española y comunitaria:

1. La relevancia del proyecto de investigación.
2. La idoneidad del protocolo, su eficacia científica y la posibilidad de extraer conclusiones válidas.
3. La justificación de los riesgos e inconvenientes previsibles en relación con los beneficios esperados.
4. La justificación del grupo de control.
5. El plan de seguimiento del estudio.
6. Los criterios de selección y retirada de los sujetos del estudio.
7. La idoneidad de los equipos de investigación.
8. La idoneidad de la información escrita sobre las características del estudio que se entregará a los potenciales sujetos de la investigación.
9. Verificación de la indemnización y previsión de tratamiento que se ofrecerá a los sujetos participantes en caso de lesión o muerte imputable al estudio, y del seguro o indemnización para cubrir las responsabilidades especificadas por la legislación correspondiente (cuando sea necesario).

En el caso de los participantes del estudio que no están en la posición legal para dar su consentimiento para participar en un estudio científico debido a su edad (menores de edad), se debe proporcionar un consentimiento informado a los padres/tutores. Además del consentimiento por parte de padres/tutores, los mayores de 12 años deben prestar su asentimiento informado para participar en estudios clínicos.

En el caso de los participantes del estudio con obstáculos para expresar su propia voluntad, se le ofrecerán las medidas de apoyo pertinentes para otorgar su consentimiento para participar en un estudio científico. En caso de imposibilidad, se debe proporcionar un consentimiento informado a los padres/tutores.

### **Métodos de datos personales no identificados**

Explique las técnicas de anonimización, pseudoanonimización, etc. que vaya a utilizar en el Proyecto en el caso de que existan datos personales.

#### Pseudoanonimización

Una de las formas de mitigar las preocupaciones legales que surgen del uso de datos personales es la pseudoanonimización. La pseudoanonimización sustituye la identidad del interesado de tal manera que los datos personales ya no pueden atribuirse a un interesado específico sin el uso de información adicional. Dicha información adicional debe conservarse por separado y está sujeta a medidas técnicas y organizativas para garantizar que los datos no se remontan a una persona física identificable. Pero como el uso de "información adicional" puede conducir a la identificación de las personas, los datos pseudoanónimos siguen siendo datos personales.

En un proceso de pseudoanonimización, los pacientes serán indicados y rastreados mediante un código unívoco en todos los documentos. Sólo este número de identificación de sujeto único vinculará los datos o las muestras con el paciente. El médico del estudio almacenará de forma segura una lista de correspondencia confidencial que contiene el código asociado con el nombre y la identidad de cada paciente. Solo el médico del estudio podrá conectar el número de identificación del sujeto con los datos personales del paciente. Él/ella no compartirá esta información excepto como se explica en el Formulario de Información del Paciente y Consentimiento Informado.

#### Anonimización

La anonimización es el proceso de convertir los datos en una forma que no identifique a las personas y donde no es probable que se produzca la identificación. Una vez que los datos son verdaderamente anónimos y las personas ya no son identificables, los datos no entran dentro del alcance del GDPR, por lo que no son datos personales y permiten una gran facilidad para compartirlos y publicarlos abiertamente.

El proceso de anonimización utiliza métodos contra la identificación, la membresía y la divulgación de atributos para minimizar el riesgo de reidentificación.

#### **Evaluación de Impacto de la Protección de Datos (DPIA)**

[Si se requiere un DPIA] Según la naturaleza de los datos que el presente proyecto pretende recopilar, podría ser necesaria una Evaluación de Impacto de la Protección de Datos (DPIA) según el RGPD (Art. 35). La DPIA es obligatoria para proyectos que probablemente impliquen “un alto riesgo” para la información personal de otras personas. Más concretamente, se requiere un DPIA cuando:

- Utilizar nuevas tecnologías.
- Seguimiento de la ubicación o el comportamiento de las personas.
- Monitorear sistemáticamente un lugar de acceso público a gran escala.
- El tratamiento de datos personales relacionados con “origen racial o étnico, opiniones políticas, creencias religiosas o filosóficas o afiliación sindical, y el tratamiento de datos genéticos, datos biométricos con el fin de identificar de forma unívoca a una persona física, datos relativos a la salud o datos relativos a la vida sexual o la orientación sexual de una persona natural”
- El procesamiento se utiliza para tomar decisiones automatizadas sobre personas que podrían tener efectos legales (o igualmente significativos).
- Tratamiento de datos de niños.

Además, siguiendo las Directrices de la CE para DPIA, también se requiere cuando:

- Evaluación o puntuación, incluidos perfiles y predicciones.
- Toma de decisiones automatizada con efectos jurídicos o que tenga un efecto similar y significativo sobre la persona física
- Monitoreo sistemático para observar, monitorear o controlar a los interesados de un área o redes de acceso público
- Datos sensibles o de carácter altamente personal
- Procesamiento de datos a gran escala.
- Conjuntos de datos que se combinan o fusionan, obtenidos para diferentes propósitos y/o diferentes fuentes
- Datos relativos a interesados vulnerables
- Utilización innovadora o aplicación de nuevas soluciones tecnológicas u organizativas.



- Tratamiento que por sí mismo impida el ejercicio de un derecho o la utilización de un servicio o contrato

Aunque no todo lo anterior se aplica al presente proyecto, nuestra recopilación de datos consistirá en parte en <insertar requisitos>.

Además, dada la edad de los participantes, los controladores de datos deben realizar evaluaciones del impacto en la protección de datos y la privacidad, considerando el impacto específico en los derechos de los niños y garantizando que los resultados algorítmicos sean los mejores para sus intereses y no afecten indebidamente su desarrollo.

Como se describe en el artículo 35, el RGPD exige que los DPIA contengan los siguientes elementos:

- Una descripción sistemática de las operaciones de tratamiento previstas y de los fines del tratamiento, incluido, en su caso, el interés legítimo perseguido por el responsable del tratamiento.
- Una evaluación de la necesidad y proporcionalidad de las operaciones de tratamiento en relación con los fines
- Una evaluación de los riesgos para los derechos y libertades de los interesados
- Las medidas previstas para abordar los riesgos, incluidas salvaguardias, medidas de seguridad y mecanismos para garantizar la protección de los datos personales y demostrar el cumplimiento del RGPD, teniendo en cuenta los derechos e intereses legítimos de los interesados y otras personas interesadas.

### **5.A b) Ejemplo real**

**Ej. 1** Todas las actividades llevadas a cabo en el marco del proyecto <AcronymProject> cumplen con los principios éticos y la legislación nacional, de la UE e internacional relevante, por ejemplo, la Carta de los Derechos Fundamentales de la Unión Europea y el Convenio Europeo de Derechos Humanos. Las tareas de <AcronymProject> solo se refieren a actividades de investigación básica y el proyecto no involucra a humanos, animales o células. Debido a que el dominio principal de la actividad del proyecto <AcronymProject> está relacionado con la ciencia de materiales con especial atención a los materiales refractarios, el riesgo de tener problemas éticos durante el proyecto es extremadamente limitado. De cualquier manera, dentro de la Parte A de <AcronymProject> DoA, el paquete de trabajo 8 está dedicado a las cuestiones éticas que establecen los 'requisitos éticos' que debe cumplir el proyecto <AcronymProject>. Se proporcionará un entregable: D8.1 NEC -Requisito No. 1. En el marco de D8.1, todos los beneficiarios y organizaciones asociadas deben confirmar que los estándares y directrices éticos de Horizonte 2020 se aplicarán rigurosamente, independientemente del país en el que se encuentren. la investigación se lleva a cabo.



**Ej. 2.** La transferencia de datos sobre seres humanos al repositorio <AcronymProject> solo se considera cuando: los consentimientos informados, la aprobación ética y, cuando corresponda, la aprobación por parte de las autoridades locales de protección de datos cubren el propósito para el cual se prevé que los datos se utilicen dentro de <AcronymProject> y permitir la transferencia de datos individuales o agregados al repositorio <AcronymProject>. Todos los datos que se transfieran al repositorio <AcronymProject> deberán estar seudonimizados o completamente anonimizados. El propietario de datos/proveedor de datos es responsable del proceso de anonimización o pseudonimización y de garantizar que las variables identificables no se transfieran al repositorio <AcronymProject>. Las variables directamente identificables incluyen, entre otras, número de identificación nacional, nombre, número de teléfono, código postal, dirección de correo electrónico, dirección y coordenadas geográficas (en una resolución que ponga en riesgo la identificación). También hay que tener en cuenta que se puede utilizar una combinación de unas pocas variables de identificación indirectas (como datos de nacimiento, género y código postal) para identificar una gran parte de personas en cualquier conjunto de datos. En este contexto, el Propietario de datos/Proveedor de datos solo proporcionará dichas variables con la resolución más baja posible que sea necesaria para el análisis, p. distrito en lugar de código postal; año de nacimiento o edad en lugar de fecha de nacimiento.

## 6. Otros temas

---

6.A Referenciar otros procedimientos nacionales/sectoriales/institucionales/de financiación usados en la gestión de datos de investigación

### **6.A a) Descripción**

Explicar los procedimientos nacionales/sectoriales/institucionales/de financiaciones usados en la gestión de datos de investigación.

#### **Información Institucional**

Cada uno de los socios, incluido el IRSJD, seguirá sus propias políticas y procedimientos institucionales para la gestión de datos de investigación, además de las políticas nacionales y europeas referentes a la gestión de datos de investigación y la ciencia abierta.

- La Política de Ciencia Abierta
- La Política de Privacidad del instituto / Fundació San Juan de Déu ([enlace](#))
- La Política de Privacidad del Hospital Sant Joan de Déu ([enlace](#))
- La Política de Privacidad del Parc Sanitari Sant Joan de Déu ([enlace](#))

### **6.A b) Ejemplo real**

**Ej. 1** Como parte del compromiso de la <Universidad> de garantizar datos FAIR y abiertos, se espera que todo el personal activo de investigación (becarios postdoctorales, estudiantes de doctorado) prepare DMP para sus propios datos, según la Política de gestión de datos de investigación de la Universidad. La política de gestión de datos de la <Universidad> define los datos de investigación como "la evidencia que sustenta la respuesta a la pregunta de investigación y puede usarse para validar los hallazgos independientemente de su forma". Por lo tanto, los datos cubren declaraciones cuantitativas y cualitativas, datos brutos de mediciones y datos derivados, ya sea limpiados o extraídos del conjunto de datos primario de un investigador o derivados de una fuente existente.

**Ej. 2** Además de las políticas de la Comisión Europea sobre gestión de datos abiertos, los socios del proyecto también deben cumplir con sus propias políticas y procedimientos institucionales para la gestión de datos:

Imperial College London:

- [Recommended file storage options](#)
- [Encrypt sensitive information](#)

UCAL

- [Regolamento per la gestione dell'innovazione e della proprietà intellettuale e industriale. Rectoral Decree n.1597, 19/10/2015](#)
- [Codice di comportamento dell'Università della Calabria. Rectoral Decree n. 2653, 23/12/2014](#)

University of Strathclyde Glasgow

- [Information Security](#)
- [Research Code of Practice](#)
- [Research Data Policy](#)

CPI

- *IT policies for the company are set out in written policies which are subject to periodic review*

FDB

- *FDB has its own set of internal policies and procedures on data management.*

## 7. Otro soporte en el desarrollo del plan

### 7.A Explicar los recursos utilizados para el desarrollo del plan

#### **7.A a) Descripción**

Lista los procedimientos, las políticas o las referencias que se han usado en el desarrollo del plan.

#### **Información Institucional**

Para la elaboración del presente plan se han utilizado los siguientes recursos:

- "EinaDMP" (<https://dmp.csuc.cat>)
- La "plantilla" y guía del Plan de gestión de datos de CSUC Plan Estatal de Investigación Científica y Técnica y de Innovación con personalizaciones por parte del IRSJD.
- La plantilla DPIA

#### **7.A b) Ejemplo real**

**Ej. 1** Este DMP ha sido creado con la herramienta "EinaDMP" (<https://dmp.csuc.cat>).

**Ej. 2** Cada uno de los socios seguirá sus procedimientos nacionales e institucionales para la gestión de datos, además de este <AcronymProject> DMP.

**Ej. 3** Referencias:

- Data Management in the context of Horizon 2020: [http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management\\_en.htm](http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-data-management/data-management_en.htm)
- Guidelines on Data Management & FAIR data principles under H2020 OA policy: [http://ec.europa.eu/research/participants/data/ref/h2020/grants\\_manual/hi/oa\\_pilot/h2020-hi -oa-data-mgt\\_en.pdf](http://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi -oa-data-mgt_en.pdf)
- The Open Research Data (ORD) Pilot in H2020: <https://www.openaire.eu/what-is-the-open-research-data-pilot>
- OpenAIRE Guidelines for Literature Repositories, Data Archives, and CRIS Managers based on CERIF-XML: <https://guidelines.openaire.eu/en/latest>

**SJD** Sant Joan de Déu  
Institut de Recerca 

© Institut de Recerca Sant Joan de Déu  
2024

Edifici Docent Sant Joan de Déu  
C/ Santa Rosa, 39-57, 4a planta  
08950 Esplugues de Llobregat  
Tel. 93 600 97 51  
[info@fsjd.org](mailto:info@fsjd.org) · [www.irsjd.org](http://www.irsjd.org)