**ANEXO: GESTIÓN DE DATOS**

**ANTECEDENTES**

ISABIAL promueve la Ciencia Abierta y mediante este documento pretende alinear e incentivar la gestión de datos de investigación en su Convocatoria de Ayudas para el Apoyo y Fomento de la Investigación con los principios *FAIR* (Localizable, Accesible, Interoperable, Reutilizable) en relación a la gestión de datos.

Por ello, a continuación, se presentan una serie de planteamientos útiles a la hora de planificar la gestión de los datos involucrados en la presente propuesta y, así, facilitar su posterior revisión por parte del personal evaluador y del Comité de Ética de la Investigación.

El tratamiento de datos personales se regula en el Reglamento (UE) 2016/679 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de éstos y en la Ley Orgánica 3/2018, de Protección de Datos Personales y garantía de los derechos digitales. De acuerdo con este marco, ISABIAL como encargado de su tratamiento para fines de investigación ofrece garantías de cumplimiento de la normativa y de las condiciones para el tratamiento de los datos por parte de su comunidad investigadora, siendo el/la investigador/a responsable del mismo.

Para el depósito, difusión y preservación de los sets de datos utilizados una vez finalizado el proyecto, la infraestructura europea creada para promover el acceso abierto *OpenAIRE* recomienda emplear, por orden de prioridad:

1. Repositorio temático consolidado para la disciplina en cuestión. En <https://www.re3data.org/> se puede encontrar un registro de repositorios de datos para elegir el más adecuado.
2. Repositorio institucional de datos. ISABIAL se encuentra a la espera de poder emplear REPISALUD, repositorio del Instituto de Salud Carlos III, con este fin.
3. Repositorio multidisciplinar de datos. Se ha habilitado una [comunidad ISABIAL](https://zenodo.org/communities/isabial/), administrada por Gestión Científica del Instituto, en [ZENODO](https://zenodo.org/), repositorio financiado por la Comisión Europea. El depósito en ZENODO permite seleccionar el tipo de acceso para cada archivo (i.e. abierto; embargado; restringido; cerrado) y la licencia de uso (p.e. Creative Commons Attribution 4.0 International, que requiere citar la base de datos y el estudio original, así como mencionar los cambios realizados; la cual se sugiere a nivel institucional para archivos identificados con acceso en abierto). Es importante señalar que el uso de ZENODO es temporal, hasta que sea viable emplear el repositorio REPISALUD por parte de la comunidad investigadora de ISABIAL.

**INSTRUCCIONES**

La cumplimentación de este anexo es obligatoria. Será necesario adaptar la información resaltada en gris.

Para ello, por favor responda a la pregunta siguiente y, en función de la respuesta elegida, complete los detalles requeridos en la plantilla que se muestre:

* ¿Los datos involucrados en su propuesta se van a recopilar y/o generar para este proyecto o son ya existentes?
	+ [Se van a generar y/o recopilar y se va a emplear REDCap.](#REDCap)
	+ [Se van a generar y/o recopilar y no se va emplear REDCap.](#BBDD)
	+ [Se van a reutilizar datos ya existentes.](#Reutilizar)

**NOTAS**

Por favor, siga las [recomendaciones](https://www.aepd.es/es/documento/tecnicas-seudonimizacion-sector-sanitario-enisa.pdf) sobre adopción de técnicas de seudonimización para el sector sanitario de la Agencia Europea para la Ciberseguridad.

De la misma forma, valore la posibilidad de emplear el *software* [REDCap](https://isabial.com/redcap/) (*Research Electronic Data Capture*), una plataforma online para proyectos de investigación que permite la recogida de datos y el diseño y creación de bases de datos, en un entorno colaborativo, de acceso restringido, ubicado en servidor privado e interoperable con los programas estadísticos más habituales, entre otras ventajas. Además, la comunidad investigadora de ISABIAL cuenta con el apoyo del personal de Bioinformática para facilitar el proceso de desarrollo sobre REDCap y comprobar la correcta seudonimización de los datos de investigación recogidos.

Respecto al depósito, se deben incluir, al menos, dos ficheros diferenciados: i) El correspondiente al set de datos; ii) Uno o más archivos con información complementaria (p.e. resultados de expresión diferencial, fichero de picos de ChIP-seq, formato BigWig, datos brutos, etc.) y/o metadatos del estudio (p.e. proyecto, autores, publicación relacionada, palabras clave, etc.).

**CONSULTAS**

Gestión Científica (965 913991)

Bioinformática (965 913981)

**PLANTILLA 1: DATOS QUE SE VAN A GENERAR Y USO DE REDCAP**

(Ejemplo en azul)

Los datos van a recopilarse y/o generarse específicamente para este proyecto y se va a emplear REDCap.

1. TITULARIDAD, UBICACIÓN Y ACCESO
* Project ID (PID): 41
* ¿Están los datos seudonimizados? Sí; No. Sí

Por favor, consulte las [recomendaciones](https://www.aepd.es/es/documento/tecnicas-seudonimizacion-sector-sanitario-enisa.pdf) sobre adopción de técnicas de seudonimización para el sector sanitario de la Agencia Europea para la Ciberseguridad.

* ¿Encriptación que se va a utilizar? Contador; Número aleatorio; Función resumen (HASH), Código de autenticación de mensaje resumido (HMAC); Cifrado. Contador
* ¿Quién es el responsable de la base de datos?
	+ - Nombre y apellidos: Óscar Moreno Pérez
		- Cargo y servicio/departamento; Grupo de Investigación: Servicio de Endocrinología. Responsable del Grupo de Investigación en Diabetes y Enfermedades Metabólicas Asociadas; Subdirector científico para Innovación, Calidad, Formación y Simulación (ISABIAL).
		- Vinculación: Departamento de Salud (DS), ISABIAL, UA, UMH) DS.
		- ORCID: 0000-0002-8670-6404

El acceso a los datos en fase de ejecución del proyecto estará restringido al equipo investigador relacionado en la propuesta, mediante acceso de doble factor en la plataforma REDCap.

1. TIPOLOGÍA Y FORMATO DE DATOS
* Carácter de los datos:observacional; experimental; operativo; de terceros; de simulación; de seguimiento; de otro tipo (identificar). Estudio observacional y de seguimiento.
* Formato del set de datos: Microsoft Excel (.xlsx); imágenes sin comprimir (.tiff); PDF (.pdf); genomas de referencia (.fa); datos brutos de secuenciación (.fastq); datos procesados de secuenciación (.sam, .bam, .bed, .bw, .R, .csv, .txt); proyectos derivados de SPSS (.sav); base de datos Access (.accdb); REDCap (.csv, .sav (SPSS), .R (R), .sas (SAS), .do (STATA), .xml (XML)); de otro tipo (identificar). Proyecto derivado de SPSS (.sav) y Microsoft Excel (.csv o .xlsx).
1. DEPÓSITO

Los sets de datos utilizados se van a depositar en:

[ ] Repositorio temático consolidado para la disciplina en cuestión: Link

(En [https://www.re3data.org/](https://www.re3data.org/%20) se puede encontrar un registro de repositorios de datos para elegir el más adecuado.)

[ ] Comunidad ISABIAL en ZENODO. [x] Comunidad ISABIAL en ZENODO.

**PLANTILLA 2: DATOS QUE SE VAN A GENERAR Y USO DE BASE DE DATOS**

(Ejemplo en azul)

Los datos van a recopilarse y/o generarse específicamente para este proyecto y no se va a emplear REDCap.

1. TITULARIDAD, UBICACIÓN Y ACCESO
* Nombre/Identificador de la base de datos: WATERLAND
* ¿Están los datos seudonimizados? Sí; No. Sí

Por favor, consulte las [recomendaciones](https://www.aepd.es/es/documento/tecnicas-seudonimizacion-sector-sanitario-enisa.pdf) sobre adopción de técnicas de seudonimización para el sector sanitario de la Agencia Europea para la Ciberseguridad.

* ¿Encriptación que se va a utilizar? Contador; Número aleatorio; Función resumen (HASH), Código de autenticación de mensaje resumido (HMAC); Cifrado. Contador
* ¿Quién es el responsable de la base de datos?
	+ - Nombre y apellidos: Enrique de Madaria Pascual
		- Cargo y servicio/departamento; Grupo de Investigación: Servicio de Aparato Digestivo. Responsable del Grupo de Investigación Clínica en Pancreatología; Subdirector científico para Proyectos (ISABIAL).
		- Vinculación: Departamento de Salud (DS), ISABIAL, UA, UMH) DS.
		- ORCID: 0000-0002-2412-9541.
* ¿Existen requisitos de acceso? Contraseña

El acceso a los datos en fase de ejecución del proyecto estará restringido al equipo investigador relacionado en la propuesta.

* ¿Cuál es la ubicación de la base de datos? Carpeta del Servicio, en el servidor del DS (HGENAREA).
1. TIPOLOGÍA Y FORMATO DE DATOS
* Carácter de los datos:observacional; experimental; operativo; de terceros; de simulación; de seguimiento; de otro tipo (identificar). Ensayo clínico.
* Formato del set de datos: Microsoft Excel (.xlsx); imágenes sin comprimir (.tiff); PDF (.pdf); genomas de referencia (.fa); datos brutos de secuenciación (.fastq); datos procesados de secuenciación (.sam, .bam, .bed, .bw, .R, .csv, .txt); proyectos derivados de SPSS (.sav); base de datos Access (.accdb); REDCap (.csv, .sav (SPSS), .R (R), .sas (SAS), .do (STATA), .xml (XML)); de otro tipo (identificar). Microsoft Excel (.xlsx).
1. DEPÓSITO

Los sets de datos utilizados se van a depositar en:

[ ] Repositorio temático consolidado para la disciplina en cuestión: Link

(En [https://www.re3data.org/](https://www.re3data.org/%20) se puede encontrar un registro de repositorios de datos para elegir el más adecuado.)

[ ] Comunidad ISABIAL en ZENODO. [x] Comunidad ISABIAL en ZENODO.

**PLANTILLA 3: DATOS QUE SE VAN A REUTILIZAR**

Los datos involucrados en esta propuesta son ya existentes.

1. TITULARIDAD, UBICACIÓN Y ACCESO

Nombre/Identificador de la base de datos:

¿Están los datos seudonimizados? Sí; No.

Por favor, consulte las [recomendaciones](https://www.aepd.es/es/documento/tecnicas-seudonimizacion-sector-sanitario-enisa.pdf) sobre adopción de técnicas de seudonimización para el sector sanitario de la Agencia Europea para la Ciberseguridad.

¿Encriptación que se va a utilizar? Contador; Número aleatorio; Función resumen (HASH), Código de autenticación de mensaje resumido (HMAC); Cifrado.

¿Quién es el responsable de la base de datos?

 Nombre y apellidos:

 Cargo y servicio/departamento; Grupo de Investigación:

 Vinculación: Departamento de Salud (DS), ISABIAL, UA, UMH)

 ORCID:

¿Existen requisitos de acceso? Por ejemplo: usuario y contraseña; únicamente contraseña; sin requisito de acceso (en este caso, indicar cómo se protegen los datos y controla el acceso.

El acceso a los datos en fase de ejecución del proyecto estará restringido al equipo investigador relacionado en la propuesta.

¿Cuál es la ubicación de la base de datos? Indicar la ruta si está ubicada en el servidor del DS (HGENAREA); si no lo está, indicar cómo se puede acceder a ella.

Si los sets de datos utilizados han sido ya depositados en algún repositorio, por favor indique el link a su ubicación y/o DOI (Identificador de Objeto Digital): Link; DOI

1. TIPOLOGÍA Y FORMATO DE DATOS

Carácter de los datos:observacional; experimental; operativo; de terceros; de simulación; de seguimiento; de otro tipo (identificar).

Formato del set de datos: Microsoft Excel (.xlsx); imágenes sin comprimir (.tiff); PDF (.pdf); genomas de referencia (.fa); datos brutos de secuenciación (.fastq); datos procesados de secuenciación (.sam, .bam, .bed, .bw, .R, .csv, .txt); proyectos derivados de SPSS (.sav); base de datos Access (.accdb); REDCap (.csv, .sav (SPSS), .R (R), .sas (SAS), .do (STATA), .xml (XML)); de otro tipo (identificar).

1. DEPÓSITO

Si los sets de datos utilizados han sido ya depositados en algún repositorio, por favor indique el link a su ubicación y/o DOI (Identificador de Objeto Digital): Link; DOI

Si no es así, por favor elija:

[ ] Repositorio temático consolidado para la disciplina en cuestión: Link

(En [https://www.re3data.org/](https://www.re3data.org/%20) se puede encontrar un registro de repositorios de datos para elegir el más adecuado.)

[ ] Comunidad ISABIAL en ZENODO