

## Documento de apoyo para la preparación de un Plan de Gestión de Datos para postulantes al Fondo Clemente Estable

Como parte de la postulación al Fondo Clemente Estable (FCE) convocatoria 2024, la ANII sumó un incentivo para promover la apertura de datos de investigación: los equipos de investigación que presenten un plan de gestión de datos obtendrán fondos extras para el desarrollo de sus proyectos. Este documento es una introducción para postulantes a la gestión de datos de investigación, los planes de gestión de datos, su importancia y beneficios, y una guía práctica de cómo generar un plan para la postulación al FCE.

### Gestión de datos de investigación

El manejo de datos de investigación permite una gestión efectiva y responsable de la investigación, e incluye la recolección, documentación, almacenamiento, intercambio y preservación a largo plazo de los datos de investigación dentro y más allá del curso de un determinado proyecto.

Los datos de investigación son la evidencia comúnmente aceptada por la comunidad académica específica como necesaria para apoyar científica y técnicamente la investigación, para sustentar los resultados de investigación y validar las conclusiones. Se generan como parte de actividades de investigación como experimentos, mediciones, encuestas, entrevistas, observaciones, etc. y son tan variados como las disciplinas académicas y los métodos con que se generan. Por ejemplo, pueden incluir notas de campo, bitácoras, textos, ilustraciones, fotografías, sonidos, bases de datos, programas informáticos, entre otros.

La publicación de los datos recogidos o generados durante la investigación contribuye a que el proceso de investigación sea más transparente y reproducible, y permite que otros/as puedan entender mejor los métodos, verificar o contrastar los resultados y reutilizar los datos para otros fines.

### Principios de gestión de datos

“Lo más abierto posible, tan cerrado como sea necesario”

Los datos de investigación colectados o generados con financiación pública deben ser gestionados de forma responsable y segura y pueden hacerse disponibles para la reutilización por parte de otras personas, siempre que sea ética, legal y comercialmente posible. Para esto se recomienda que los datos se gestionen siguiendo los [Principios FAIR](#) (acrónimo de *Findable, Accessible, Interoperable* y *Reusable*).

Los Principios FAIR proporcionan una guía para hacer que los resultados de la investigación sean encontrables, accesibles, interoperables y reutilizables. Sólo si los datos de la investigación se almacenan de forma estructurada y con metadatos (datos que describen los datos) podrán ser recuperados y utilizados por otros.

**Encontrable (findable).** El primer paso para (re)utilizar los datos es encontrarlos:

- Los datos están descriptos con metadatos enriquecidos y tienen asignado un identificador único persistente. Los datos y metadatos están indexados.

**Accesible (accessible).** Una vez que la persona encuentra los datos, necesita saber cómo puede acceder a ellos:

- Los metadatos y datos son recuperables por su identificador y pueden ser entendidos por humanos y máquinas.

**Interoperable (interoperable).** Por lo general, los datos deben poder integrarse con otros datos e interoperar con aplicaciones o flujos de trabajo para su análisis, almacenamiento y procesamiento.

- Los datos y metadatos usan un lenguaje formal, accesible, compartido y de amplia aplicación para la representación del conocimiento.

**Reutilizable (reusable).** Para optimizar la reutilización de los datos, los metadatos y los datos deben estar bien descritos para que puedan reproducirse y/o combinarse en distintos entornos.

- Los datos y metadatos se describen detalladamente con atributos precisos y relevantes. Los datos se publican con licencias de uso claras y proporcionan información precisa sobre su procedencia.

Al realizar investigaciones en colaboración con poblaciones específicas, los/as investigadores deben consultar a las personas y grupos involucrados para elaborar conjuntamente los principios de gestión de datos y ser especialmente conscientes de los datos sensibles y/o de alto riesgo.

## **Plan de gestión de datos (PGD)**

Un plan de gestión de datos es un documento formal que detalla las estrategias y herramientas que serán usadas para la gestión eficiente de los datos generados o recolectados durante el proyecto de investigación y luego de su cierre. El objetivo de un PGD es servir de herramienta de apoyo en cada etapa del ciclo de vida de los datos: recolección, generación, tratamiento y publicación.

Hay muchas formas de presentar un PGD, sin embargo, el contenido suele cubrir los siguientes aspectos: descripción de los datos a recolectar o generar y cómo se van a documentar (metadatos), qué aspectos éticos y legales se tendrán en cuenta, cómo se hará el almacenamiento, respaldo y seguridad de los datos, y cómo se van a compartir y reutilizar.

Los PGDs son documentos ‘vivos’ y sujetos a cambios, que deben mantenerse y actualizarse a lo largo del curso de la investigación. El plan que se presenta en la postulación al fondo es una primera versión. En el transcurso del proyecto de investigación, los PGD pueden ser expandidos o modificados para acomodar cambios.

## **Importancia y beneficios de los PGD**

Los planes de gestión de datos se reconocen cada vez más como requisitos para la excelencia en la investigación científica. La preparación de un PGD contribuye a mejorar la productividad y calidad de la investigación, a garantizar una gestión responsable y efectiva de los datos, a mejorar la reutilización de los datos y amplificar el impacto del conocimiento generado.

Algunas de las razones por las que un PGD es importante

- Estimula a los/as investigadores/as a considerar mejores prácticas para la recolección, el almacenamiento, el análisis y la preservación de los datos, que puede ayudar a asegurar su integridad y calidad a lo largo del proyecto y después de finalizado.
- Colabora con que los datos se almacenen de manera organizada y se documenten adecuadamente, lo que maximiza su valor y utilidad para la comunidad científica.
- Ayuda asegurar el cumplimiento de las regulaciones legales y éticas aplicables a datos de investigación
- Detalla medidas de seguridad y copias de respaldo que reducen el riesgo de pérdida o daño de los datos.
- El proceso de preparación puede ayudar a los investigadores a planificar y coordinar mejor el trabajo del proyecto, lo que puede mejorar la eficiencia, minimizar los problemas y reducir los costos.
- Mejora la transparencia y la reproducibilidad de la investigación al asegurar que los datos estén disponibles para otros investigadores para su verificación y validación

## **Guía práctica para generar el PGD**

Al preparar su plan de gestión de datos, sea tan específico como fuera posible/corresponda. A continuación, presentamos una guía sobre cómo completar las diferentes secciones del documento para su postulación al FCE 2024.

## 1. Datos a recolectar o generar

<p><b>¿Qué datos se generarán o recolectarán en el marco del proyecto?</b></p> <p><b>Incluya si aplica tipo, volumen, unidad y formatos de datos que se espera obtener.</b></p>	<p>Detalle por separado qué tipos de datos serán generados y cuáles serán recolectados (reutilizados) en el marco de su proyecto. Ejemplos de generación pueden ser ‘un paquete de código en lenguaje R’ y ejemplos de reutilización pueden ser ‘conjuntos de datos meteorológicos existentes’.</p> <p>Ejemplos de tipo de datos son fotos, sonidos, videos, documentos, formularios, mediciones, muestras geológicas, modelos 3D, encuestas online, código, programas (software), análisis químicos, secuencias genéticas, mapas interactivos, reconstrucciones digitales.</p> <p>El volumen (incluyendo su unidad) puede ser conocido o puede estimarse. Indique, en lo posible, el tamaño para un archivo y mencione cuántos se crearán, o indique la cantidad de muestras/análisis/secuencias, número de personas encuestadas, etc., y el tamaño total esperado de los datos (menos de 1GB, entre 1GB y 1TB, más de 1 TB).</p> <p>El formato de los datos puede ser digital (.png, .tiff, .wav, .csv, etc.), material o en papel. Es aconsejable preferir formatos abiertos y bien establecidos en la comunidad académica correspondiente.</p>
<p><b>¿Cómo se obtendrán los datos?</b></p>	<p>Esta información es necesaria para poder reconstruir el proceso por el que se generarán los datos y, si corresponde, de dónde serán obtenidos para su reutilización. Detalle aquí, por ejemplo, qué herramientas, programas informáticos, tecnologías o procesos serán utilizados. Ejemplos de origen de los datos pueden ser trabajo de campo, de laboratorio, archivos, publicaciones, equipos de medición, simulaciones, compilaciones, etc.</p>
<p><b>¿Cómo se organizarán y estructurarán?</b></p>	<p>Detalle con qué estructura se guardarán los datos del proyecto (los que se generarán y los que se reutilizarán). Por ejemplo, si existen guías internas del proyecto (o buenas prácticas de la comunidad) para la organización de los datos, convenciones sobre cómo nombrarlos y cómo versionarlos.</p>

## 2. Descripción de los datos y metadatos

<p><b>¿Cómo describirá los datos para que sean fácilmente encontrables,</b></p>	<p>Los metadatos describen a los datos de una manera estructurada, aportan información sobre el contenido del recurso y las relaciones que existen entre las partes de un conjunto de datos (tanto para las</p>
---	---

<p><b>interpretables y reutilizables?</b></p>	<p>personas como para las máquinas). Muchos repositorios de datos facilitan la entrada de metadatos mediante formularios online.</p> <p>Algunos de los estándares de metadatos más reconocidos son Dublin Core Metadata Standard para información bibliográfica, Data Documentation Initiative (DDI) para encuestas y datos de observación y las Text Encoding Initiative (TEI) para datos textuales. Vea aquí más <a href="#">ejemplos de metadatos en diferentes áreas del conocimiento</a> y el <a href="#">buscador de estándares de FAIRsharing.org</a>.</p> <p>Especifique el/los estándar/es que usará para la creación de metadatos. Detalle por ejemplo:</p> <ul style="list-style-type: none"> <li>● Se utilizarán estándares, ontologías, clasificaciones, etc. específicas de la disciplina (especifique)</li> <li>● Se utilizará un sistema de descripción personalizado (describalo brevemente y, si es necesario, indique dónde está documentado con más detalle)</li> <li>● No se utilizará ningún sistema fijo para la descripción.</li> </ul> <p>Detalle si usará identificadores persistentes en sus datos y metadatos (ejemplo: DOI, handle, ARK, URN, ORCID, ROR, etc.).</p> <p>Detalle si sus datos permitirán el intercambio y la reutilización de datos entre investigadores, instituciones, organizaciones, países, etc. Por ejemplo, el conjunto de datos adhiere a formatos estandarizados, o puede combinarse fácilmente con otros conjuntos de datos de distintos orígenes, o el conjunto de datos se puede utilizar con programas informáticos disponibles (abiertos) o aplicaciones informáticas establecidas y ampliamente utilizadas en la comunidad correspondiente.</p>
<p><b>¿Qué documentación se generará para describir los métodos utilizados para la generación/recolección y tratamiento?</b></p>	<p>Indique cómo detallará la información necesaria para que otras partes puedan comprender cómo realizó la recolección o creación de sus datos, y el análisis y resultados de investigación obtenidos a partir de ellos. Por ejemplo, sobre el proceso de creación, metodología, fuentes, tiempos, herramientas/tecnología, contenido, identificadores, ubicación.</p>

### 3. Aspectos éticos y legales

<p><b>¿Quién tendrá la propiedad de los datos?</b></p>	<p>La propiedad de los datos en un proyecto depende de varios factores, como la fuente de los datos, los acuerdos establecidos entre los miembros del proyecto y las políticas institucionales. En caso de dudas</p>
--	--

	<p>se recomienda contactarse con referentes de Propiedad Intelectual en su institución para obtener más información sobre este punto.</p> <p>Algunas referencias a políticas institucionales sobre propiedad intelectual y repositorios: UdelaR (<a href="#">Ordenanza de los Derechos de la Propiedad Intelectual</a>, <a href="#">Ordenanza del Repositorio Abierto</a>), LATU (<a href="#">Código de Ética (Capítulo 11- Propiedad Intelectual)</a>), UTEC (<a href="#">Ordenanza de Propiedad Intelectual</a>), ORT (<a href="#">Reglamento estudiantil de ORT (art. 9)</a>), INIA (<a href="#">Código de Ética y Estatuto del Personal</a>).</p> <p>En muchos casos, si los datos son recolectados o generados por investigadores en el marco de su trabajo académico, es probable que la propiedad de los datos sea de la institución en la que se llevó a cabo el proyecto. En este caso, la universidad/instituto podría compartir los derechos con el autor o tener derechos exclusivos para utilizar, publicar y compartir los datos, incluso si el/la investigador/a que los recolectó o generó ya no forma parte de la institución.</p> <p>La propiedad de los datos puede afectar su disponibilidad y reutilización en el futuro. Por lo tanto, es recomendable que los proyectos definan claramente la propiedad de los datos desde el inicio y establezcan acuerdos para su uso y acceso en el futuro.</p>
<p><b>¿Existen restricciones para la publicación de datos bajo propiedad de terceros?</b></p>	<p>Especifique si existen restricciones legales (derechos de autor, propiedad intelectual) o acuerdos de confidencialidad que limiten su uso o publicación. Ver más sobre la <a href="#">Ley de Derechos de Autor en Uruguay</a>.</p> <p>Si hubiera restricciones, se deben obtener los permisos necesarios antes de publicar la información bajo propiedad de terceros.</p>
<p><b>En caso de tratarse de datos de personas, ¿cómo se asegurará el cumplimiento de la ley 18.331 de protección de datos personales y el decreto 158/019 relativo a la investigación en seres humanos?</b></p>	<p>Ejemplos de datos personales son: nombre, dirección, teléfono, cédula de identidad, número de RUT, huella digital, número de socio, número de estudiante, una fotografía o incluso hasta el ADN.</p> <p>La Ley establece que dentro del universo de datos personales algunos son considerados como sensibles y por tanto tienen un régimen especial para su tratamiento. Estos son datos que contienen información sobre origen racial y étnico, opiniones políticas, creencias religiosas o morales, afiliación sindical, salud o vida sexual.</p> <p>La recolección, tratamiento, y almacenamiento de datos personales sólo es admisible cuando se ha obtenido el "consentimiento informado" de la persona en cuestión. Detalle si se obtendrá o no y por qué.</p>

	<p>Tenga en cuenta que si trabaja con datos sensibles es necesario disociarlos mientras el proyecto está en curso. Disociarlos implica el tratamiento de los datos de manera que la información obtenida no pueda vincularse a persona. A la hora de publicar datos personales sensibles deberá anonimizarlos, es decir, tratarlos de manera de evitar de forma irreversible su identificación. Vea más información sobre cómo hacer esto en la <a href="#">Guía Criterios de Disociación de Datos Personales</a> de la Unidad Reguladora y de Control de Datos Personales.</p> <p>Detalle si se anonimizarán los datos, cómo y cuándo. Detalle también cuándo se eliminarán de forma segura los datos originales (sin anonimizar).</p> <p>Dentro de la actualización del marco normativo nacional fue creada la figura del Delegado de Protección de Datos Personales que tiene como funciones asesorar sobre las de protección de datos personales y supervisar su correcto cumplimiento en la institución. Recomendamos consultar en su institución para obtener apoyo en el tema.</p>
<p><b>¿Existen otros aspectos éticos o legales a tener en cuenta en la recolección, generación, tratamiento y publicación de los datos?</b></p>	<p>En caso que no lo haya mencionado en los puntos anteriores, detalle aquí, por ejemplo, si en la recolección de datos además del consentimiento informado y la privacidad, también tendrá en cuenta evitar la discriminación de personas por su raza, género, orientación sexual, religión, discapacidad, entre otros. Tenga en cuenta que otro tipo de datos sensibles son los datos que contienen por ejemplo la localización geográfica de especies amenazadas.</p> <p>En cuanto a la generación de datos, puede mencionar aspectos de la calidad de los datos y los procedimientos para que sean precisos, confiables y objetivos.</p> <p>Respecto al tratamiento de los datos, detalle si prevé tomar medidas para ser transparentes en el tratamiento de los datos y/o para evitar el acceso no autorizado, la alteración o la pérdida de los datos.</p> <p>Especifique si planea publicar sus datos en formatos accesibles/abiertos.</p>

#### 4. Almacenamiento, respaldo y seguridad

<p><b>¿Cómo se almacenarán los datos en el transcurso de la investigación?</b></p>	<p>Detalle dónde se guardarán los datos durante el desarrollo del proyecto, y, por ejemplo, si prevé costos de almacenamiento.</p>
--	--

<p><b>¿Qué respaldos y medidas de seguridad se prevén a lo largo del proyecto?</b></p>	<p>Las medidas de seguridad y los respaldos deben ser apropiados para los riesgos asociados con los datos y deben ajustarse a las políticas de su institución y las regulaciones legales nacionales (por ejemplo, sobre protección de datos personales).</p> <p>Especifique si prevé realizar copias de seguridad regulares de los datos recolectados y generados, y si las hará en un disco duro o en un servicio de alojamiento online o datacenter. Detalle si los datos tendrán acceso restringido, por ejemplo, estableciendo protección con contraseña y/o la implementación de permisos de usuario.</p>
--	--

## 5. Acceso a los datos generados o recolectados

<p><b>¿Para quiénes considera que pueden ser útiles los datos que se generarán o recolectarán?</b></p>	<p>Detalle para qué otras partes pueden ser útiles sus datos. En otras palabras, cuál es el potencial de reutilización de sus datos.</p> <p>Algunos destinatarios previstos pueden ser la comunidad científica, el público en general, el sector privado, para uso interno de su organización, etc.</p>
<p><b>¿Qué datos serán publicados en el repositorio provisto por ANII y cuándo?</b></p>	<p>Especifique qué datos de investigación publicará en el repositorio de ANII (generados o recolectados en su proyecto), cuáles no y por qué razón justificada.</p> <p>Detalle cuáles son los criterios/reglas para la selección de los datos que se archivarán (una vez finalizado el proyecto).</p>
<p><b>¿Prevé alguna restricción en relación a quiénes y cómo deberían poder acceder a ellos?</b></p>	<p>Detalle este punto en referencia a los aspectos éticos y legales presentados previamente. Especifique, además, si habrá un periodo de embargo antes de que los datos estén disponibles.</p> <p>Los datos que se publiquen en acceso abierto en el repositorio de ANII utilizarán licencias Creative Commons a elección del autor. Ejemplos: Sin obra derivada (ND), No comercial (NC), Atribución (BY), Compartir igual (SA), Dominio público (CC0). Vea aquí cómo <a href="#">elegir una licencia de Creative Commons</a>.</p> <p>En caso de que los datos generados no puedan ser compartidos bajo estas licencias indique si se utilizará alguna otra licencia estandarizada o especifique los términos y condiciones de uso de los mismos.</p>
<p><b>En caso de prever restricciones o embargos para la</b></p>	<p>Detalle la razón de un período de embargo. Si aplica, justifique la necesidad de imponer restricciones a compartir sus datos.</p>



<b>publicación de los datos justifique.</b>	
<b>¿Prevé compartir los datos en otros repositorios?</b>	Mencione si sus datos serán depositados en repositorios específicos de su disciplina. Por ejemplo: <a href="#">HEASARC</a> (datos astronómicos), <a href="#">GenBank</a> (secuencias de nucleótidos), <a href="#">GBIF</a> (registros de biodiversidad), <a href="#">RADAR4Culture</a> (patrimonio cultural), <a href="#">MorphoSource</a> (objetos 3D). Ver más repositorios específicos de disciplinas en <a href="http://re3data.org">re3data.org</a>