

## Política de Datos para el Programa de Apoyo a la Formación Doctoral ZEF-IDEA (PAFD)

(Antonio Rogmann, Universidad de Bonn)

**Nota:** Este documento puede estar sujeto a modificaciones de acuerdo con las disposiciones del órgano jurídico y legal de la Universidad Nacional de Colombia – Sede Bogotá

### Contenido

Abreviaciones y acrónimos.....	2
Definición de términos técnicos.....	2
1 Introducción .....	3
1.1 Objetivos.....	3
1.2 Principios en el SGD.....	3
1.3 Alcance .....	4
2 Gestión de Derechos de Datos .....	4
2.1 Reglamento de Intercambio de Datos .....	4
2.1.1 Marco Legal .....	4
2.1.1.1 Propiedad Intelectual .....	5
2.1.1.2 Privacidad de Datos .....	5
2.1.2 Reutilización de Datos .....	5
2.1.3 Licencia .....	5
2.1.4 Usuarios de Datos.....	6
2.1.5 Proveedores de Datos.....	6
2.1.5.1 Estudiantes de Becas .....	6
2.1.5.2 Otros Proveedores de Datos.....	7
3 Cita y Agradecimientos.....	8
4 Marco Organizacional.....	8
4.1 Tareas básicas .....	8
4.2 Responsabilidades .....	9

## Abreviaciones y acrónimos

- SGD (Sistema de Gestión de Datos científicos): El conjunto de todos los componentes técnicos, organizativos y legales de la gestión de datos del PAFD;
- PDM (Portal de Datos y Mapas): Portal web cual permite a los usuarios acceder a las bases de datos científicos del PAFD utilizando un catálogo de metadatos;
- UGD (Unidad de Gestión de Datos): El Grupo encargado de implementar y mantener el SGD

## Definición de términos técnicos

- Datos: Un dato es una representación simbólica (numérica, alfabética, algorítmica, espacial, etc.) de un atributo o variable cuantitativa o cualitativa. Los datos describen hechos empíricos, sucesos y entidades. Es un valor o referente que recibe el computador por diferentes medios, los datos representan la información que el programador manipula en la construcción de una solución o en el desarrollo de un algoritmo<sup>1</sup>;
- Metadatos: Los datos que proporcionan información sobre los datos disponibles y la disponibilidad, lo que permite al usuario evaluar su aptitud para su propio uso y propósitos;
- Datos primarios: Datos recopilados / creados por el usuario de datos;
- Datos secundarios: Datos recopilados y proporcionados por alguien que no sea el usuario;
- Datos brutos: Datos sin procesar. Proviene de los registradores de datos u otros instrumentos de campo o encuestas sin ninguna revisión adicional en cuanto a la calidad de los datos o cualquier otro procesamiento;
- Usuario de datos: Un individuo autorizado para acceder y utilizar datos.
- Proveedor de datos: Un individuo o una institución que distribuye datos a usuarios de datos.
- Estándar de datos: Norma o requisito para un artículo (por ejemplo, archivo de datos, servicio, proceso, técnica) basado en un acuerdo común.
- Gestión de datos: "La Gestión de Recursos de Datos es el desarrollo y ejecución de arquitecturas, políticas, prácticas y procedimientos que gestionan apropiadamente las necesidades del ciclo de vida completo de los datos de una empresa"<sup>2</sup>.
- Plan de Gestión de Datos: Documento formal creado al inicio de un proyecto de investigación que describe:
  - flujos de trabajo de gestión de datos (control de calidad, formateo, codificación, creación de metadatos, etc.) en el ciclo de vida de los datos desde la adquisición, el procesamiento, el intercambio, el archivado hasta la eliminación de datos.
  - estándares en uso.
  - instalaciones técnicas.
  - marco organizativo de gestión de datos (responsabilidades en los flujos de trabajo de gestión de datos).

---

<sup>1</sup> <https://es.wikipedia.org/wiki/Dato>

<sup>2</sup> Data Management Association (DAMA) [https://es.wikipedia.org/wiki/Gestion\\_de\\_datos](https://es.wikipedia.org/wiki/Gestion_de_datos)

- compartimiento de datos y reglas de intercambio (política de intercambio de datos, licencias).

## 1 Introducción

### 1.1 Objetivos

El objetivo del Sistema de Gestión de Datos científicos (SGD) en el “*Programa de Apoyo a la Formación Doctoral*” (PAFD) es apoyar la investigación y la enseñanza dentro como fuera del PAFD con un fácil acceso a datos científicos generados o recopilados por los investigadores del IDEA, los estudiantes del PAFD, o de sus socios. El apoyo se materializa por:

- el ahorro de tiempo y costos en la búsqueda, y el libre acceso a los datos científicos;
- evitar la duplicación de esfuerzos en la recopilación de datos;
- el aumento de la calidad de los resultados científicos por permitir su reproducción con los datos originales
- el respaldo en el desarrollo de nuevos proyectos con base en los datos generados por proyectos anteriores
- la distribución de datos científicos bien documentados a usuarios externos a la academia, por ejemplo a miembros de movimientos ambientales que buscan hacer sus propias interpretaciones y tomar decisiones mejor informadas
- la consideración de preocupaciones éticas y aspectos legales tales como la protección de la privacidad o la protección de la propiedad intelectual.

Los datos científicos se guardan como archivos en un banco de datos. Este banco de datos es instalado en servidores en la Universidad Nacional de Colombia (sede Bogotá), de propiedad y administración del IDEA, y accesible para cualquiera a través de un Portal de Datos y Mapas (PDM) el cual ofrece diferentes servicios web para acceder los datos. En cierto caso, por ejemplo, si se necesita asignar un DOI<sup>3</sup> a un conjunto de datos, los datos se pueden archivar y publicar en otros repositorios. El PAFD usa para eso un catálogo en *Dataverse*.<sup>4</sup> Bajo el término "datos" en este documento se resumen artefactos digitales de varios tipos como, por ejemplo, datos brutos en tablas, productos de datos como informes, diagramas, mapas, videos, códigos, aplicaciones derivadas de la explotación, procesamiento y combinación de los datos brutos, las bases de datos y tablas generadas del procesamiento de estos datos en bruto, y en diferentes formatos digitales y de contenido variado que son producidos por actividades investigativas en el PAFD, u otros. El conjunto de la base de datos, el portal web, la política de datos establecido en este documento, los flujos de trabajo y sus responsabilidades necesarios para gestionar, archivar y publicar los datos se llama “**Sistema de Gestión de Datos Científicos en el Programa de Apoyo a la Formación Doctoral IDEA-ZEF**” (SGD).

### 1.2 Principios en el SGD

Para lograr los objetivos del SGD los principios organizativos y legales claramente definidos se establecen y describen en esta política de la siguiente manera:

---

<sup>3</sup> Identificador de objeto digital (<https://www.doi.org/>)

<sup>4</sup> <https://dataverse.org/>

- El SGD adhiera a los principios F.A.I.R<sup>5</sup>, cuales exigen que datos sean localizable (Findable), Accesible, Interoperable y Reusable;
- Los datos deben ser de alta calidad, de acuerdo con las mejores prácticas de las respectivas disciplinas científicas. La gestión de la calidad y su documentación son responsabilidad del científico que proporciona los datos
- Los datos deben estar bien documentados, de modo que los usuarios puedan evaluar la idoneidad de los datos para sus fines
- Los datos se deben registrar de una manera que se puedan buscar y acceder fácilmente mientras sean protegidos contra uso indebido
- Los derechos de los datos que sirven para proteger la propiedad intelectual y la confidencialidad de la información personal, así como otros requisitos para restringir el acceso a los datos, deben considerarse de manera coherente, tanto durante la preparación y distribución de los datos
- Los autores de datos deben citarse cuando los datos se reutilizan y los contribuyentes a la generación o el procesamiento de los datos deben ser reconocidos
- El uso de datos obtenido por el PDM sin la debida cita está estrictamente prohibido.

### 1.3 Alcance

La política de datos en el SGD define los derechos de uso y accesibilidad con respecto a los datos de investigación que son almacenado en las bases de datos del SGD. El SGD se dirige a todos los usuarios del PDM, sean usuarios o proveedores de datos. Los coordinadores del programa determinan quién está autorizado de usar el PDM para archivar y publicar datos.

Proyectos, que tienen sus propias políticas y regulaciones de intercambio de datos y son asociados con el IDEA/UNAL, pueden ser invitados de usar el SGD para la publicación de sus datos científicos. En este caso, se necesitará una colaboración con la Unidad de Gestión de Datos en el IDEA/UNAL (UGD) para organizar la integración de datos y para gestionar los derechos de acceso a los datos en el SGD mediante el PDM. Se recomienda crear un Plan de Gestión de Datos adecuado para el proyecto de antemano.

## 2 Gestión de Derechos de Datos

La gestión de derechos de datos se refiere al marco legal utilizado en la distribución y reutilización de los datos de investigación e incluye:

- Protección de la propiedad intelectual y otros derechos relacionados con los datos;
- Protección de información sensible; y
- Reconocimiento de aportes intelectuales y prácticos durante la generación de los datos.

A continuación, el alcance de los derechos de propiedad intelectual y otros derechos se conoce como "derechos de datos".

### 2.1 Reglamento de Intercambio de Datos

#### 2.1.1 Marco Legal

---

<sup>5</sup> <https://www.force11.org/group/fairgroup/fairprinciples>

La gestión de la protección de la propiedad intelectual y de la privacidad de datos cae debajo de la ley colombiana. Con respeto a la propiedad de datos valen las prescripciones legales de la institución por cual el proveedor de datos está contratado.

#### **2.1.1.1 Propiedad Intelectual**

Valen las regulaciones de la Universidad Nacional del Colombia<sup>6</sup>.

#### **2.1.1.2 Privacidad de Datos**

Los datos personales implican cualquier información relativa a circunstancias personales o materiales de una persona física identificada o identificable. Los datos asociados con un individuo (por ejemplo, los datos del hogar con información privada / personal que pueden relacionarse con el individuo) se pueden usar y procesar en investigaciones científicas. Sin embargo, dichos datos deben anonimizarse lo antes posible sin comprometer el propósito de la investigación. Si esto no es posible, los datos deben ser pseudoanonimizados, es decir, se deben dividir los registros de datos personales en partes y almacenarse por separado para impedir la identificación de las personas físicas. Con una función de asignación, los datos se pueden volver a unir si es necesario. Esta función puede no ser conocida por las personas involucradas en el procesamiento de datos, excepto por las personas autorizadas para custodiar los datos confidenciales.

Los datos personales pueden usarse en publicaciones sólo si las personas interesadas han declarado por escrito su acuerdo o si el interés público en la presentación de los resultados de la investigación supera sustancialmente los intereses legítimos de la persona o personas afectadas.

El autor de los datos con restricciones de privacidad o divulgación y requisitos de confidencialidad debe protegerlos contra la pérdida, alteración o cualquier otro tipo de mal uso. Los datos personales confidenciales almacenados en el SGD deben cumplir con los requisitos legales de la UNAL<sup>7</sup>.

#### **2.1.2 Reutilización de Datos**

El uso posterior de datos por parte de terceros, por ejemplo, al procesar nuevos datos en un marco de investigación profesional y proporcionar datos a terceros, requiere el consentimiento previo del titular de los derechos. Este consentimiento se otorga generalmente en la forma de una licencia. El derecho a "reutilizar" los datos incluye el derecho a reproducirlos, distribuirlos y hacerlos accesibles (por ejemplo, a través de Internet), en relación con las actividades de investigación o los procesos de toma de decisiones. La "reutilización" solo está permitida dentro de los límites legales y / o límites de la licencia. Las posibles restricciones en los datos pueden limitar los derechos de uso, por ejemplo, al prohibir una mayor distribución. El derecho de "reutilización" no libera al usuario de la obligación de indicar las fuentes y / o sus obligaciones hacia el autor y, cuando corresponda, la agencia financiadora que permitió la realización de los estudios por medio de los cuales los datos fueron creados.

#### **2.1.3 Licencia**

Una licencia es un acuerdo que establece el alcance de los derechos de uso. Normalmente, la licencia debe ser concluida con una firma de ambas partes. En el PDM donde se pueden descargar datos con derechos de autor se incluirá una licencia en los metadatos. La licencia será integrada en el proceso de descarga de los datos del PDM, de modo que durante dicho proceso y en el ingreso de la información de contacto, el usuario acepta las condiciones de la licencia.

---

<sup>6</sup> <http://investigacion.unal.edu.co/servicios/propiedad-intelectual-un>

<sup>7</sup> [http://unal.edu.co/fileadmin/user\\_upload/docs/POLITICA\\_DE\\_TRATAMIENTO\\_DATOS.pdf](http://unal.edu.co/fileadmin/user_upload/docs/POLITICA_DE_TRATAMIENTO_DATOS.pdf)

#### 2.1.4 Usuarios de Datos

Un usuario de datos dentro del alcance de SGD se define como cualquier persona que usa los datos producidos, elaborados y dispuestos por los científicos que trabajan en estudios o proyectos realizados o coordinados por el PAFD (IDEA). Los usuarios que deseen acceder a datos restringidos deben tener una cuenta de usuario en el PDM. Las cuentas de usuario están reguladas por la UGD en colaboración con los titulares de los derechos sobre los datos solicitados. La información sobre cómo obtener permiso para descargar datos restringidos se incluye en los metadatos de los datos solicitados.

#### 2.1.5 Proveedores de Datos

Los compromisos legales con respecto a la provisión de datos al SGD dependen del tipo de empleo del autor de los datos. Los estudiantes de doctorado y maestría en el PAFD (en lo sucesivo denominado "becario") tienen contratos de beca estudiantil con el Centro de Investigación para el Desarrollo (ZEF) de la Universidad de Bonn (Alemania), por lo tanto, están bajo la legislación laboral alemana. Todos los demás proveedores de datos (en lo sucesivo denominado "otro proveedor de datos"), ya sea de la UNAL u otras organizaciones que deseen utilizar el PDM como depósito de datos, se encuentran bajo sus derechos como autores de datos de acuerdo con la legislación colombiana, de acuerdo con las disposiciones de la UNAL.

##### 2.1.5.1 Estudiantes de Becas

Para el suministro de datos, las siguientes pautas y opciones son aplicables a los becarios en el PAFD:

- a) Los becarios mantienen todos sus derechos de autor de los datos que generan (= datos primarios) siempre que se cumplan los requisitos de la ley nacional de derechos de autor. En consecuencia, un becario es libre en la decisión de compartir sus datos con terceros;
- b) Los datos en bruto, así como los datos derivados generados y utilizados en un estudio del PAFD deben entregarse al supervisor científico, a más tardar, simultáneamente con el estudio, de modo que el supervisor pueda verificar los resultados del estudio si es necesario;
- c) Los datos deben ir acompañados por una documentación clara y separada de los mismos (metadatos), para que el supervisor o cualquier otra persona autorizada pueda identificar el conjunto de datos, comprender su contenido y los métodos utilizados durante el procesamiento con el mínimo esfuerzo;
- d) Los becarios voluntariamente pueden otorgar una licencia abierta de uso de datos primarios a IDEA/UNAL. Si los derechos de uso de datos fueron transferidos por el becario a IDEA/UNAL a través de una licencia<sup>8</sup>, IDEA/UNAL pone estos datos a disposición para el público científico como no científico con base en dicha licencia;
- e) El becario puede agregar licencias personalizadas, por ejemplo, publicando los detalles de contacto en los metadatos de un conjunto de datos, de modo que los usuarios potenciales puedan contactar al autor directamente y solicitarle el acceso o los derechos de uso. Las licencias a medida deben diseñarse en colaboración con el administrador de datos y el departamento legal de UNAL;
- f) El almacenamiento de datos y metadatos se realiza dentro del PDM en consideración de los estándares de datos definidos por la UGD en el Plan de Gestión de Datos para el PAFD<sup>9</sup>. Los

---

<sup>8</sup> Por ejemplo a base del marco de licencias Creative Commons, Open Data Commons (para datos que no sean debajo de propiedad intelectual) o General Public License (GNU) para código o scripts de software

<sup>9</sup> Aquí URL de descarga para el Plan de Manejo de Datos

metadatos en general son públicos. Si el becario no desea permitir el acceso a los datos según una licencia abierta, la visibilidad de los metadatos también puede ser restringida, por ejemplo, para los supervisores;

- g) Si el becario desea almacenar datos secundarios de fuentes externas en el PDM y compartirlos por este medio, los derechos de datos especificados por la fuente externa deben respetarse y formar parte de los metadatos;
- h) El becario recibe capacitación en gestión de datos, entre otros, para la creación de metadatos en el PDM, la estandarización de los datos para su intercambio, la concesión de licencias y la carga a la base de datos usando el PDM;
- i) El becario recibe una cuenta de usuario en el PDM y, por lo tanto, un control total sobre sus datos y metadatos;
- j) En consulta con la UGD, el becario es responsable de la publicación oportuna de los datos y la protección de los mismos en caso de que estén sujetos a restricciones, por ejemplo, porque todavía son parte de un proceso de publicación para un artículo científico o por razones de confidencialidad (mira sección 2.2.1.2).

#### 2.1.5.2 Otros Proveedores de Datos

Las siguientes reglas y opciones se aplican a la integración de datos generados por cualquier otro proveedor de datos, ya sea de IDEA/UNAL o de organizaciones asociadas con el IDEA/UNAL en un proyecto, a menos que las regulaciones específicas del proyecto, en forma de un Plan de Gestión de Datos (sección 1.3), lo definan de otra manera.

- a) El otro proveedor puede usar el PDM como plataforma de intercambio de datos, según lo autorice IDEA/UNAL;
- b) El otro proveedor posee todos los derechos de autor de los datos que produce, siempre que se cumplan los requisitos de la ley nacional de derechos de autor o la regulación de la Universidad Nacional del Colombia o la regulación de la política de datos de un cierto proyecto en cual trabaja (si existe);
- c) El derecho a reutilizar los datos protegidos por derechos de autor se otorga a IDEA/UNAL o a cualquier tercero en forma de licencia abierta<sup>10</sup>;
- d) Los datos deben ser publicados en el PDM en consideración de los estándares de datos definidos por la UGD o por un Plan de Gestión de Datos definido para un cierto proyecto (si existe);
- e) El otro proveedor recibe una cuenta de editor en el PDM para que pueda documentar y cargar sus datos a la base de datos;
- f) El contenido de los datos cuales se van a publicar debe ser controlados y aceptados por IDEA/UNAL (por parte de la UGD) antes de ser publicado. El otro proveedor tiene la responsabilidad de manejar el acceso a sus datos de acuerdo con las regulaciones de esta política o la política de un proyecto aceptado por IDEA/UNAL para usar el SGD;
- g) Si el otro proveedor almacena datos secundarios de fuentes externas en el PDM y compartirlos por medio del PDM, los derechos de datos especificados por la fuente externa deben respetarse y formar parte de los metadatos.

---

<sup>10</sup> Por ejemplo a base del marco de licencias Creative Commons, Open Data Commons (para datos que no sean debajo de propiedad intelectual) o General Public License (GNU) para código o scripts de software

### 3 Cita y Agradecimientos

- a) Si los datos son proporcionados por una institución (generalmente el caso con datos secundarios), la institución debe ser citada y, si está disponible, la URL<sup>11</sup> desde donde se descargaron los datos, debe ser indicado dentro de los metadatos del PDM;
- b) Cuando los datos almacenados en el SGD y obtenidos a través del PDM se utilizan en una publicación, la fuente de los datos debe ser citada utilizando los detalles de la cita incluidos en los metadatos de los respectivos datos. La cita se puede incluir como una nota en el pie de página, en la bibliografía o en un apéndice con referencias similares;
- c) Si dos o más personas crearon un único conjunto de datos que se publica mediante del PDM, la publicación conjunta, es decir, la denominación de todos los autores de datos como autores/coautores es necesario;
- d) Cuando la recopilación o el preprocesamiento de datos primarios fue sustancialmente apoyado, por ejemplo, por asistentes académicos, técnicos de campo, etc.), y los cuales no son coautores de los datos, deben ser reconocidos adecuadamente en los metadatos.

### 4 Marco Organizacional

En los flujos de trabajo de gestión de datos, las tareas específicas necesitan competencias específicas, ya sea de los científicos, de los administradores de datos o de ambos. Por este motivo, las tareas, que se enumeran en la sección 3.1, deben asignarse entre la UGD y los individuos o proyectos que participan en el SGD.

En esta política IDEA/UNAL ha delineado un esquema general. Sin embargo, se pueden negociar detalles con cada programa o proyecto que actúa debajo del programa PAFD o del IDEA/UNAL para cumplir con sus requisitos y disponibilidad de recursos. IDEA/UNAL recomienda establecer flujos de trabajo de gestión de datos junto con el personal de gestión de proyectos en un plan de gestión de datos específico del proyecto. Los planes de gestión de datos se pueden preparar en cooperación con la UGD.

#### 4.1 Tareas básicas

El marco regulatorio y organizativo asigna tareas específicas de gestión de datos entre individuos, proyectos y la UGD. Las tareas clave en la gestión de datos son:

- El desarrollo de planes de gestión de datos (o documentos comparables) que contengan políticas de gestión de datos específicas del proyecto y directivas sobre estándares de datos que se deben aplicar y responsabilidades que asignar;
- La preparación de datos siguiendo los estándares en términos de control de calidad, reformato, denominación de archivos, recopilación y control de calidad de los metadatos y asignación de información de fondo;
- Integración de datos y metadatos en el PDM y la gestión de los derechos de acceso por nivel de conjunto de datos;
- La integración de enlaces web para acceder a los datos cuando se citan en publicaciones;

---

<sup>11</sup> Uniform Resource Locator: Dirección única para un recurso en el internet

- El apoyo de los proveedores de datos en cumplir con las demandas del SGD a través de capacitación y soporte técnico.

## 4.2 Responsabilidades

El SGD en IDEA/UNAL tiene las siguientes responsabilidades:

- Administración técnica de la SGD asegurando:
  - Disponibilidad del servicio 24/7;
  - Protección de datos e información personal del usuario contra accesos no autorizados;
  - protección de las bases de datos contra la pérdida de datos;
- Gestión de características funcionales y contenido temático en el PDM con:
  - Actualización del diseño de la interfaz web del PDM si es necesario;
  - Actualización o ampliación de las características funcionales a pedido.
  - Integración de elementos temáticos utilizados para la búsqueda de datos del proyecto, tales como categorías de datos y palabras clave-tesoros;
  - Proporcionar creación de capacidad mediante materiales de aprendizaje en línea, instrucciones de proceso y capacitaciones.
- En colaboración con proyectos y / o socios, el desarrollo de planes o directrices de gestión de datos para uno o varios proyectos (según corresponda) que contengan:
  - Un Memorandum de Entendimiento con respecto a la Política de Datos del PAFD en la medida en que se requiera (sección 1.3);
  - Especificación de los estándares que se aplicarán a los datos antes de compartirlos;
  - Especificación de licencias que definen la reutilización de datos;
  - Información sobre los datos que se integrarán en el SGD, incluidos los períodos de retención;
  - Asignación de personal o responsabilidades relacionadas con el puesto dentro de los proyectos para:
    - Membresía en el comité de gestión de datos;
    - Implementación del control de estandarización de datos respectivamente;
    - Medidas de aseguramiento de la calidad de los datos.
    - Creación de plantillas de metadatos;
    - Cargar datos y metadatos en el PDM y revisarlos antes de publicarlos;
    - Implementación de reglas de intercambio de datos por parte de la administración de cuentas de usuario en el PDM.