

Bogotá D.C., miércoles 28 de septiembre de 2022

## **CIRCULAR**

**PARA:** **TODA LA COMUNIDAD TOMASINA SEDES Y SECCIONALES**

**DE:** **COLOMBIA**

**ASUNTO:** **REPOSITORIO DE DATOS PAPYRUS**

El Plan de Gestión de Datos es un documento que describe el ciclo de vida de los datos recopilados, generados y procesados durante el proyecto de investigación. Se especifica la recolección, organización, descripción, almacenamiento, difusión, reutilización y preservación de los datos.

Teniendo en cuenta la importancia de la gestión de Datos de Investigación (GDI), la Universidad Santo Tomás como miembro del Consorcio Colombia, suscribe el Repositorio de Gestión de Datos Papyrus, “repositorio de carácter multidisciplinario, que alberga los conjuntos de Datos científicos de proyectos de investigación de las Instituciones colombianas que pertenecen al Consorcio, con el fin de darles a los datos, visibilidad, garantizar su preservación y facilitar su acceso, recuperación y reutilización a través de la custodia y gestión centralizada, dando cumpliendo a estándares internacionales”.(Consortia, 2022)

Por lo anterior, queremos dar a conocer el plan de trabajo en Papyrus y los invitamos a participar de este proyecto.

Fraternamente;

**FRAY VICENTE BECERRA REYES, O.P.**  
Director General  
Centro de Recursos para el Aprendizaje y la Investigación  
CRAI SANTOTO - BOGOTÁ

## Plantilla de Gestión de Datos

Información administrativa



Los permisos se otorgarán a cada uno de los miembros del proyecto que hará uso del Repositorio de Datos. Se asignan permisos sobre todo el Dataverse del proyecto o sobre cada uno de los Datasets que lo conforman.

## Roles de los participantes de Papyrus





## Estructura PGD Universidad Santo Tomás

DATAVERSE



DATASET



DATAFILE

Programa o  
Departamento

Para cada programa  
o departamento se  
creará un DataVERSE

Grupos de  
Investigación

Para cada de Grupo  
de investigación se  
generará un Dataset

Archivos de cada  
Dataset

Cada archivo  
tendrá asignado  
un DOI

### 1. Directrices de acceso abierto

La Universidad Santo Tomás, a través del CRAI SANTOTO, gestiona actualmente diversas plataformas que facilitan el acceso y la visibilidad de la producción académica institucional.

En el marco del desarrollo de la ciencia abierta, y en concordancia con las directrices internacionales de la UNESCO para el acceso abierto, así como con los lineamientos de Colciencias para una política nacional de ciencia abierta en Colombia, la Universidad fortalece sus líneas de acción en infraestructura física, tecnológica y de comunicaciones, así como en vigilancia tecnológica, bibliometría e innovación. Dentro de esta oferta de servicios se incluye la Gestión de Datos de Investigación (GDI).

Este proceso impulsa a los autores tomasinos a elaborar un **Plan de Gestión de Datos**, en el cual se describa el ciclo de vida de la información recopilada en sus investigaciones y el tratamiento que debe recibir en cada etapa de desarrollo. De igual manera, el **CRAI SANTOTO** se encargará de garantizar la **custodia y el acceso** a dichos datos, de manera que puedan ser utilizados como insumo para nuevas producciones científicas, en el marco de la ciencia abierta, a través del Repositorio de Datos de Investigación (RDI) – Papyrus..

## 2. Principios que rigen la GDI en el CRAI SANTOTO

Para la custodia y el acceso a los conjuntos de datos de investigación, el CRAI SANTOTO establece una serie de principios que orientan la publicación, divulgación y preservación de la información a través del RDI-Papyrus.

Los aspectos clave para su desarrollo se organizan en tres grupos:

1. Administrativo: relacionado con la adquisición, almacenamiento y organización de los datos.
2. Preservación: referido al uso de formatos estándar, los medios de migración y las estrategias de resguardo de la información.
3. Descubrimiento: centrado en los mecanismos de búsqueda y recuperación, el acceso y el licenciamiento que permita poner los datos del repositorio a disposición de la comunidad.

Con base en lo anterior, se adoptaron los siguientes **principios estándar**:

**Principios FAIR** (Localización, Accesibilidad, Interoperabilidad y Reutilización de activos digitales): permiten implementar procesos que aseguren que los datos sean localizables mediante metadatos, identificadores y palabras clave; que sean accesibles a través de licencias y plataformas; que logren interoperabilidad mediante el uso de estándares para indexación y registro; y que fomenten la reutilización de los conjuntos de datos como insumo para nuevas producciones científicas.

**Principios DataCite**: tienen como finalidad garantizar la localización y citación de los datos de investigación, en articulación con los principios FAIR. Este proceso se lleva a cabo mediante la asignación de DOIs y metadatos específicos que facilitan el acceso a los conjuntos de datos.

**Principios OpenAIRE**: establecen pautas internacionales orientadas a asegurar la accesibilidad abierta de los datos, así como la adopción de políticas de investigación que promuevan la interoperabilidad de los sistemas de información.

### 3. Compromisos

#### Universidad Santo Tomás:

- Proveer el acceso y mantenimiento a las plataformas requeridas para la custodia de los datos de investigación.
- Garantizar el talento humano para el asesoramiento de almacenamiento y uso de la información contenida en las plataformas

#### Investigadores:

- Diligenciar claramente la plantilla de datos y proporcionar los conjuntos de datos necesarios debidamente descritos para su depósito.
- Acudir al personal del CRAI para el respectivo resguardo de la información y la permanente actualización de los datos.
- Mantener copias de seguridad de los datos que permitan recuperar la Información en caso de siniestros en los sistemas de información del CRAI SANTOTO.
- Garantizar los documentos soporte de uso libre de la información contenida en los conjuntos de datos.

#### CRAI:

- Proporcionar a los diferentes grupos que conforman la línea de investigación (Investigadores, semilleros de investigación, coautores) un espacio de resguardo de información contenida en los datos de investigación para asegurar su preservación.
- Brindar el asesoramiento (Acompañamiento, formación) necesario a los investigadores en la subida, selección, y resguardo de sus conjuntos de datos
- Motivar la promoción y uso de los conjuntos de datos para la producción de Nuevo conocimiento en la comunidad de la CRAI SANTOTO
- Generar la correlación respectiva entre los conjuntos de datos contenidas en el Repositorio Papyrus y la investigación resguardada en el repositorio Institucional
- Preservar los conjuntos de datos en el repositorio de acuerdo con el tiempo establecido en la plantilla de datos garantizando su permanencia, acceso y seguridad.
- Establecer estrategias de promoción del Repositorio Papyrus en concordancia

con la política de ciencia abierta motivando de esta forma la visibilidad nacional e internacional de la producción tomasina.

- Velar por la correcta acreditación de propiedad de los conjuntos de datos.  
(Citación)

#### 4. Contenidos y Propiedad intelectual

El CRAI SANTOTO conservara en el RDI- Papyrus contenidos producidos por la comunidad CRAI SANTOTO que cumplan con los siguientes requerimientos:

##### 4.1 *Tipo de datos y formato*

- Conjuntos de datos que sean parte de una investigación producida por la comunidad de la CRAI SANTOTO y que respondan a un programa académico.
- Solo se conservarán datos que sean producidos por autores de la comunidad CRAI SANTOTO

**Tabla 1.**

**Tipos de archivos de datos**

<b>Tipo de dato</b>	<b>Descripción</b>
<b>Contenedores</b>	tar, gzip, zip, SQL, SPARQL
<b>Bases de datos</b>	xml, csv, json
<b>Específicos de programas</b>	Software
<b>Código fuente</b>	Java, C, R, Python
<b>Video</b>	mpeg, avi, mxf, mkv
<b>Sonido</b>	wave, aiff, mp3, mxf flac
<b>Estadísticas:</b>	dta, por, sas, sav
<b>Imágenes:</b>	tiff, jpeg 2000, pdf, dng, gif, bmp, svg
<b>Datos tabulares:</b>	CSV, txt, xls / xlsx, mbd / *.accdb, dbf
<b>Texto:</b>	xml, pdf / a, html, json, txt
<b>Cualitativos textuales:</b>	xml, rtf, RDF, OWL, RIS, BIBTex
<b>Archivo web:</b>	Warc, html

*Nota: Se recomienda utilizar formatos con estándares abiertos y documentados, no encriptados, sin compresión y usados entre investigadores de su área.*

De esa misma forma, los tipos de datos que no estén contemplados en este documento y sean necesarios de depósito serán estudiados previamente de acuerdo con la solicitud.

#### *4.2 Tipos de datos excluidos del RDI- Papyrus*

En el RDI-Papyrus en manejo de la universidad Santo Tomás y el CRAI SANTOTO se abstiene de publicar datos o conjuntos de datos con:

- Información confidencial
- Información sin autorización para publicación o divulgación
- Información que infrinja derechos de terceros/autor
- Que, contenga elementos de carácter publicitario o pornográfico, ya sea explícito o implícito
- Que contenga información que constituya acto ilícito alguno
- Datos que no tengan toda la información clara en la plantilla de datos
- Datos que no sean de producción de autores tomasinos.
- Documentos diferentes a conjuntos de datos (Artículos- Revistas- libros);  
Estos últimos se resguardarán en el repositorio institucional CRAI  
SANTOTO.

*Nota: Se podrá incluir un pre-print del artículo cuando este aun no este publicado y se mantendrá oculto del Set de datos.*

### **5. Aspectos legales de los conjuntos de datos**

Una vez la información sea ingresada en el Repositorio de Datos:

- El autor (es) conservarán los **Derechos patrimoniales y morales sobre su obra**.
- Entregar a la Universidad y el CRAI SANTOTO los derechos de reproducción y comunicación pública con fines de promoción y preservación de los datos; esto no implica explotación comercial ya que esta es expresa del autor(es)
- El CRAI SANTOTO no se hace responsable por la autenticidad y veracidad de los datos contenidos en el RDI-Papyrus; esta responsabilidad recae sobre el autor(es)
- Los autores están en la obligación de expresar cualquier cambio sobre los derechos/obligaciones de uso del conjunto de datos establecidos previamente.  
Estos deben ser expresados por escrito.

## **Licencias**

### ***Licencia de depósito.***

Para la publicación del set de conjunto de datos; el autor debe diligenciar y entregar al CRAI SANTOTO:

- Formulario GDI completamente diligenciado. En caso de duda con el diligenciamiento, el personal del CRAI estará disponible para el asesoramiento necesario.
- Formato de derechos de autor con la firma de todos los participantes de la investigación
- Los documentos que acrediten el depósito del resultado de investigación de esa investigación en un repositorio institucional o revista electrónica para realizar las correlaciones pertinentes.

### ***Licencias de uso.***

El licenciamiento de los conjuntos de datos se hará expresamente bajo las licencias Creative Commons 4.0. El autor deberá especificar en la plantilla GDI el tipo de licencia y alcance de acuerdo con la siguiente información.

#### **Tabla 2.**

#### *Licenciamiento creative Commons 4.0*

<b>Tipo de licencia</b>	<b>Descripción</b>
<b>Atribución (CC BY)</b>	La obra puede usarse siempre y cuando se cite al autor.
<b>Atribución-Compartir Igual (CC BY-SA)</b>	Puede usarse para cualquier proyecto, sea o no comercial, se edite o modifique o no, siempre y cuando se cite al autor y sea publicada bajo la misma licencia CC BY-SA 4.0.
<b>Atribución-No Derivadas (CC BY-ND)</b>	Sea o no con ánimo de lucro, siempre y cuando se cite al autor y no se haga una obra derivada de ella, es decir no editar la obra.
<b>Atribución-No Comercial (CC BY-NC)</b>	Se puede usar la obra si se cita al autor, pero no en proyectos comerciales.
<b>Atribución-No Comercial-Compartir Igual (CC BY-NC-SA)</b>	Se puede usar una obra siempre y cuando sea para un proyecto sin ánimo de lucro, se cite al autor y se publique el resultado bajo la licencia CC BY-ND-SA 4.0.

<b>Atribución-No Comercial-No Derivadas (CC BY-NC-ND):</b>	Use una obra siempre y cuando se cite al autor, sea para proyectos no comerciales y no se modifique de ninguna manera.
--	--

De esta forma se garantizará el alcance de publicación y acceso de los conjuntos de datos suministrados por el autor.

5. Esquema de metadata, interoperabilidad y organización de la publicación de conjunto de datos
6. Metadata y operabilidad

**Tabla 3.**

**Esquema de metadata e interoperabilidad RDI-Papyrus**

Ítem	Descripción
<b>Proveedor tecnológico</b>	Consortia Papyrus
<b>Estándares</b>	Diretrices de OpenAIRE para Administradores de Repositorios de Literatura 4.0 Principios FAIR
<b>Esquema</b>	DataCite Dublín Core
<b>Protocolo</b>	Citation metadata (DCMI Metadata) Geospatial Metadata (Iso 3166-1) Social Science & Humanities Metadata
<b>Identificador digital</b>	DOI ORCI D SCOP US ID
<b>Vocabulario semántico</b>	Lenguajes COAR (Confederation of Open Access Repositories)

## Glosario

**Data/Datos:** Colección de elementos básicos de información producidos, almacenados o procesados por un programa informático.

**Datafile/Archivo:** Archivo que contiene texto, gráficos, números u otro tipo de contenido.

**Dataset /Conjunto de datos:** Colección de datos relacionados, agrupados o almacenados juntos que fueron producidos dentro de un mismo proyecto, o se enfocan hacia un mismo objeto de estudio.

**Dataverse:** Colecciones que pueden contener datasets, datafiles o también otros dataverses.

**Datos abiertos:** Son considerados datos abiertos a los que cualquier persona puede acceder. Generalmente su uso está soportado en una licencia que indique el alcance del uso.

**Datos compartidos:** Este tipo de datos suelen estar disponibles con contenido de restricción sobre su uso o para un grupo o comunidad específica donde solo los usuarios autorizados tienen acceso.

**Datos de investigación:** Definidos según la CEPAL como insumos que se generan o recolectan para sustentar los resultados de una investigación.

**Datos restringidos:** Este tipo de datos tienen un tratamiento confidencial por diferentes causas, pero no para uso o libre descarga. Generalmente se incluyen en repositorios para descripción y preservación. DOI: Identificador de un objeto digital. Se usa en publicaciones electrónicas.

**Metadato:** Dato que describe el contenido, el estado, el origen, entre otras características de datos u otro tipo de información.

**Repositorio:** Plataforma donde se almacenan datos.

**Rol:** Conjunto predefinido de permisos para realizar ciertas acciones en el dataverse, dataset o datafile. Se puede asignar un rol tanto a un usuario como a un grupo.



## Bibliografía

*Confederation of open Access Repositories [COAR]. (s.f.). Guía de implementación para vocabularios controlados COAR.* <https://www.coar-repositories.org/implementation-guide-for-coar-controlled- vocabularies>