



POLÍTICA DEL REPOSITORIO DE DATOS DE INVESTIGACIÓN DEL TEC

MSc. Maritza Morales Roldan.
MATI Jason Enríquez Fallas.

<https://creativecommons.org/licenses/by/4.0/>

TABLA DE CONTENIDO

Glosario.....	3
1.Repositorio TECdatos.....	6
2. Datos aceptados en el repositorio TECdatos.....	7
<i>2.1 Condiciones para el depósito de datos de investigación.....</i>	<i>8</i>
<i>2.2 Pautas para nombrar al archivo Readme.....</i>	<i>9</i>
3. Depósito y descripción de datos de investigación.....	9
4. Denominar ficheros (archivos).....	10
5. Propiedad intelectual.....	11
<i>5.1 Legislación costarricense para la protección de datos de investigación.....</i>	<i>11</i>
<i>5.2 Licencias para el uso de datos.....</i>	<i>11</i>
<i>5.3 Compromiso de citar datos.....</i>	<i>13</i>
<i>5.4 Límites del acceso abierto.....</i>	<i>13</i>
<i>5.5 Datos confidenciales.....</i>	<i>14</i>
Anexos.....	15
<i>Anexo #1 Presentación de datos tubulares.....</i>	<i>16</i>



▶▶▶▶▶▶▶▶ GLOSARIO

ANONIMIZACIÓN DE DATOS: es una técnica de tratamiento de datos que elimina o modifica los datos personales identificables para obtener datos anónimos que no se pueden asociar con ninguna persona.

ARCHIVO README: Usualmente son creados como readme.txt o readme.md, suelen contener información importante sobre el sistema, proyecto, conjunto de datos o software al que se refieren. Para que los usuarios puedan encontrar fácilmente el archivo, se recomienda ubicarlo en el nivel superior de las carpetas.

DATACIONES: es el acto de atribuir un tiempo o una fecha a un suceso o a un objeto.

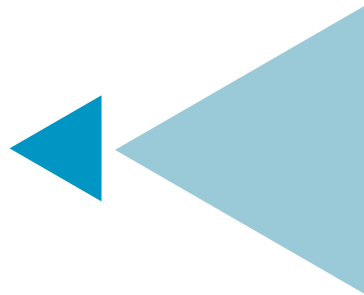
DATASET: Un dataset es un conjunto de datos relacionados con un tema o sector concreto. Los datasets o conjuntos de datos incluyen diferentes tipos de información.

DATAVERSE: Es un repositorio de datos de investigación que se encuentra alojado en la Universidad de Harvard, utiliza un software de código abierto que permite compartir datos.

DATOS FAIR: Son datos abiertos que cumplen con los principios de encontrabilidad, accesibilidad, interoperabilidad y reutilización.

FOTOGRAMETRÍAS: es la técnica cuyo objeto es estudiar y definir con precisión la forma, dimensiones y posición en el espacio de un objeto cualquiera, utilizando esencialmente medidas hechas sobre una o varias fotografías de ese objeto.

POLÍTICA DEL
REPOSITORIO
 DE DATOS
DE INVESTIGACIÓN DEL TEC



1.Repositorio TECdatos

Plataforma institucional del Tecnológico de Costa Rica que tiene como objetivo depositar, preservar y divulgar los datos en acceso abierto generados en las investigaciones que se producen en la Universidad aumentando la visibilidad de sus contenidos. Este repositorio permite a los investigadores depositar los datasets (conjunto de datos) asociados a artículos de revista, proyectos de investigación, informes de graduación u otras publicaciones siempre y cuando no sean confidenciales y que no existan limitaciones por propiedad intelectual y/o seguridad.

Beneficios de depositar datos en este repositorio:

- ▶ Se alinean con los datos FAIR (localizables, accesibles, interoperables, reutilizables).
- ▶ Los datos son resguardados de forma íntegra y segura a lo largo del tiempo.
- ▶ Se cumple con los requerimientos de agencias de financiamiento y revistas que solicitan la publicación de datos de investigación en un repositorio.
- ▶ Se acelera la producción académica y científica.
- ▶ Se reducen costos al evitar duplicar esfuerzos en recolección de datos.
- ▶ Se mejora la visibilidad de las investigaciones y los investigadores.
- ▶ Se da paso a nuevos tipos de colaboraciones entre los creadores y los usuarios de los datos.
- ▶ Aumenta la transparencia y la rendición de cuentas de los proyectos.

2. Datos aceptados en el Repositorio TECdatos

Debido a la diversidad de datos e instrumentos utilizados por nuestra comunidad universitaria, no existen restricciones actuales sobre los formatos de archivo que los usuarios pueden cargar en el TecDatos.

Para fines de preservación e interoperabilidad a largo plazo y para garantizar un uso más amplio, recomendamos la publicación de datos en formatos abiertos, o que se transformen en abiertos si esto no implica que los datos se alteren.

A continuación, se muestra una lista orientativa de formatos recomendados para datos de investigación:

- ▶ **Imágenes:** TIFF, JPEG 2000, PDF, PNG, GIF, BMP, SVG
- ▶ **Datos tabulares:** CSV, TXT, XLSX
- ▶ **Videos:** MPEG, AVI, MXF, MKV
- ▶ **Audio:** WAVE, AIFF, MP3, MXF
- ▶ **Texto:** XML, PDF/A, HTML, ASCII, UTF-8
- ▶ **Bases de datos:** XML, CSV, JSON
- ▶ **Datos Geoespaciales:** SHP, DBF, GeoTIFF, netCDF, GeoJSON, Rinex
- ▶ **CAD:** .dxf, .sat, .igs, .stp
- ▶ **Nube de Puntos (PointCloud):** LAS, LAZ, XYZ, PTX
- ▶ **Estadísticas:** ASCII, DTA, POR, SAS, SAV
- ▶ **Sismología:** SEED
- ▶ **Código:** (tcl files, py files) Jupyter Notebook
- ▶ **Archivo web:** WARC
- ▶ **Contenedores:** TAR, GZIP, ZIP

2.1 Condiciones para el depósito de datos de investigación

- 1 Solo se pueden gestionar y publicar datos relacionados con investigaciones del Tecnológico de Costa Rica o con trabajos publicados en el Portal de Revistas del TEC.
- 2 Tienen que ser públicos y en ningún caso, su publicación podría suponer una violación legal.
- 3 Se permite el embargo de los datasets (conjunto de datos) para los casos que lo ameriten.
- 4 Para publicar datos de investigación se debe de enviar el formulario “Autorización de Publicación y Descripción de datos de investigación”.
- 5 Cuando la presentación de los datos sea en formato tabular (p.ej. hojas de cálculo) es imprescindible que los campos, filas y columnas aparezcan identificados con una denominación que permita reconocer a qué se refieren los datos. (Ver Anexo 1).
- 6 No se aconseja el envío de archivos comprimidos, con excepción de archivos muy grandes (preferir opciones como OneDrive Institucional, memoria USB, entre otros).
- 7 Evite enviar archivos muy grandes al Repositorio TecDatos, el tamaño recomendado es de 100 MB como máximo.
- 8 Se recomienda incluir un archivo Readme por conjunto de datos. Este archivo es una forma de documentación que usualmente está en un archivo de texto plano como por ejemplo un bloc de notas formato txt.

- 9 Suba únicamente datos sobre los que usted y su grupo de trabajo tengan autoridad para distribuir.
- 10 El o los datasets deben ser originales.
- 11 No infringir ningún derecho de terceros, sea de propiedad industrial, intelectual, secreto comercial o cualquier otro derecho susceptible de protección jurídica.

2.2 Pautas para nombrar al archivo Readme

En caso de que se desee incluir el archivo Readme se deberá de nombrar de tal manera que sea fácilmente identificado en el conjunto de datos.

Ejemplo:

- ➡ Readme.txt
- ➡ Readme_plasma.txt

Si el ítem está compuesto de varios dataset y preferimos crear un único archivo Readme que englobe todos, debemos asignarle un nombre significativo, es decir, que todos los dataset puedan verse reflejados en él, ejemplo:

- ➡ Readme_general

Contenido del archivo Readme:

1. Información general sobre el o los dataset:
2. Información metodológica
3. Información de los archivos
4. Información específica de los datos

3. Depósito y descripción de datos de investigación

El depósito de los dataset, lo llevará a cabo la Unidad de Procesos Técnicos del SIBITEC.

La calidad de la descripción es responsabilidad de los autores de los datos de investigación.

Se deberá entregar a la Biblioteca:

- ▶ Formulario de Autorización de Publicación y Descripción de datos de investigación.
- ▶ El o los datasets (conjuntos de datos).
- ▶ El archivo Readme.txt (opcional)

Esta información deberá compartirse en una carpeta al Onedrive datostec@itcr.ac.cr con el nombre del investigador responsable.

4. Denominar ficheros (archivos)

- ▶ Los nombres de los ficheros deben plasmar los contenidos de estos, además, tienen que ser unívocos.
- ▶ Crear nombres identificativos pero cortos.
- ▶ Conservar las extensiones de los códigos de aplicaciones como por ejemplo pdf, doc, xlsx, jpg entre otros.
- ▶ Usar el guión bajo (_) mejor que espacios en blanco entre palabras.
- ▶ Evitar el uso de caracteres especiales (. : ; / () \$ & | [] * < > “¿.)
- ▶ Reflejar tipo de archivo y el contenido del archivo. Ej. Tabla_Plantas_ornamentales
- ▶ Evitar llamar a los ficheros por un número, no es suficientemente descriptivo del contexto del registro.
- ▶ No superar los 25 caracteres.
- ▶ En registros que contienen distintas versiones, es recomendable diferenciar unas versiones de otras indicando v01, v02 etc. La versión definitiva, se puede llamar “versión FINAL”. Para estos casos, se aconseja una coherencia interna y un mismo criterio de denominación para cada fichero, ejemplo:
 - ▶ Tabla_Plantas_ornamentales_v01
 - ▶ Tabla_Plantas_ornamentales_v02
 - ▶ Tabla_Plantas_ornamentales_version_final

5. Propiedad intelectual

5.1 Legislación costarricense para la protección de datos de investigación:

- ➡ Ley 7975, Ley de Información No Divulgada
- ➡ Ley 8968, Ley de Protección de la Persona frente al tratamiento de sus datos personales
- ➡ Ley 7788, Ley de Biodiversidad de Costa Rica.
- ➡ Ley 9234, Ley Reguladora de Investigación Biomédica.

5.2 Licencias para el uso de datos

Para la protección de los datos abiertos que van a ser publicados se recomienda el uso de las licencias Creative Commons donde el dueño de los datos asignará el uso que quiere otorgar a los mismos, las que se recomiendan son las siguientes:

Reconocimiento CC BY 4.0

Asegura que siempre se dé el crédito adecuado a los autores, pero dificulta el uso simultáneo de varios conjuntos de datos al requerir que se siga una forma de citar predeterminada que puede ser diferente según el publicador.

Dominio Público CC0

Esta licencia hace disponible información sobre los derechos de autor y dedica la información al dominio público. Está especialmente recomendada para datos de investigación, ya que facilita la compilación y el uso masivo de la información.

CC BY-NC-ND 4.0

Se permite el uso de la obra siempre que se cite al autor, no se emplee en un proyecto comercial y no se modifique la obra original.

CC BY-NC-SA 4.0

Permite a los reutilizadores distribuir, remezclar, adaptar y construir sobre el material en cualquier medio o formato solo con fines no comerciales, y solo mientras se otorgue la atribución al creador. Si remezcla, adapta o construye sobre el material, debe licenciar el material modificado bajo términos idénticos.

CC BY-ND 4.0

Permite a los reutilizadores copiar y distribuir el material en cualquier medio o formato solo en forma no adaptada, y solo mientras se dé la atribución al creador. La licencia permite el uso comercial.

CC BY-SA 4.0

Permite a los reutilizadores distribuir, remezclar, adaptar y construir sobre el material en cualquier medio o formato, siempre que se dé atribución al creador. Permite el uso comercial. Si remezcla, adapta o construye sobre el material debe licenciar el material modificado bajo términos idénticos.

Reconocimiento CC BY NC

Esta licencia permite que los reutilizadores distribuyan, remezclen, adapten y desarrollen el material en cualquier medio o formato únicamente con fines no comerciales, y siempre y cuando se le otorgue la atribución al creador.

Las licencias Creative Commons no son adecuadas para los softwares. Existen otras licencias para este fin:

- ▶ MIT License: Esta licencia permite reutilizar el software, es posible usarla para licenciar tanto software libre como software no libre.
- ▶ Apache Licence: Es una licencia de software libre permisiva, requiere la conservación del aviso de derecho de autor y el descargo de responsabilidad, pero no es una licencia copyleft, ya que no requiere la redistribución del código fuente cuando se distribuyen versiones modificadas.

La licencia otorga al usuario la libertad de usar el software para cualquier propósito, redistribuirlo, modificarlo y distribuir versiones modificadas del software, bajo los términos de la licencia.

- ▶ GNU General Public Licence: Es una licencia de derecho de autor ampliamente usada en el mundo del software libre y código abierto, y garantiza a los usuarios finales (personas, organizaciones, compañías) la libertad de usar, estudiar, compartir (copiar) y modificar el software.

Su propósito es doble: declarar que el software cubierto por esta licencia es libre, y protegerlo (mediante una práctica conocida como copyleft) de intentos de apropiación que restrinjan esas libertades a nuevos usuarios cada vez que la obra es distribuida, modificada o ampliada.

Si usted requiere aplicar otro tipo de licencia para los datos depositados en el TECDatos, póngase en contacto a través del email tecdatos@itcr.ac.cr.

5.3 Compromiso de citar datos

El repositorio genera una cita estandarizada con identificadores únicos digitales (DOI). Esta cita debe ser colocada por los usuarios de los datos en la sección de referencias, tal como se citan otros tipos de trabajos.

5.4 Límites del acceso abierto

Existen circunstancias que limitan el acceso abierto a los datos, las más frecuentes son:

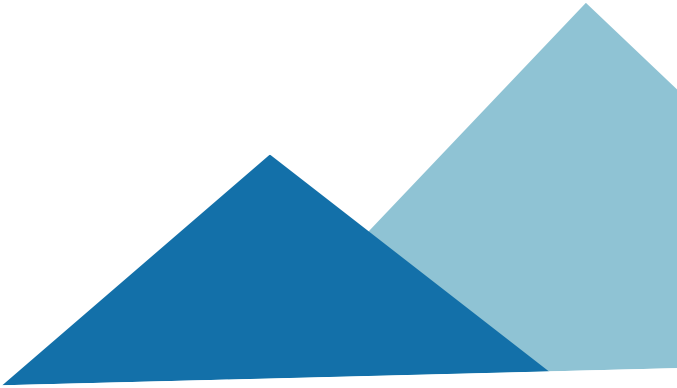
- ▶ Aquellos datos que se generen en el trámite de una patente de invención.
- ▶ Aquellos datos que se generen de un modelo de negocio.
- ▶ Aquellos datos que estén sujetos a compromisos contractuales, nacionales o internacionales que impidan su inclusión en el repositorio con vista en el plazo señalado en el contrato, convenio, compromiso entre otros

- ▶ Aquellos que tengan información de tipo confidencial y por tanto no puedan ser publicados o puestos a disposición del público.
- ▶ Aquellos que al divulgarlos puedan poner en riesgo la seguridad de los investigadores o sujetos de investigación.
- ▶ Datos de investigación que no cumplan con la calidad requerida Ej. Nitidez de imágenes. Titulación de tablas y gráficos, buena resolución de videos, entre otros)"
- ▶ Datos de terceros que están protegidos bajo secreto industrial o comercial o cuyos datos están en confidencialidad. (Ley 7975, Ley de Información No Divulgada)
- ▶ Datos sensibles que no estén debidamente autorizados para su publicación. (Ley 8968, Ley de Protección de la Persona frente al tratamiento de sus datos personales. Ley 9234, Ley Reguladora de Investigación Biomédica).
- ▶ Investigaciones que hagan uso de los elementos de la biodiversidad y que no cumplan con lo establecido en la ley 7788, Ley de Biodiversidad.

5.5 Datos confidenciales

Para minimizar los riesgos de divulgación de información confidencial, considere los siguientes factores cuando realice el diseño de su investigación:

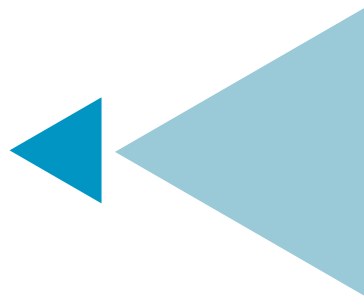
- ▶ Si es posible, recopile los datos necesarios sin utilizar información con identificación personal.
- ▶ Si se requiere información de identificación personal, retire la identificación de los datos después de la recolección (anonimización).



▶▶▶▶▶▶▶▶ **ANEXOS**

Anexo #1

Presentación de datos tabulares



Data_Heating_and_steaming_of_Hyronyma_alchornoides

Archivo Inicio Insertar Dibujo Diseño de página Fórmulas Datos Revisar Vista Automatizar Ayuda

Calibri (Cuerpo) 11 A N K S ab D [Color] [Bordes] [Estilos] Ajustar Combinar General \$

	A	B	C	D	E	F	G	H	I	J	K	L
1	Treatment	Tree	Height	Boards	Code	Width (mm)	Thickness (mm)	Bow	Crook	Twist	Check length (mm)	Split length (mm)
182	Untreatment	62	A	4	62A-4	91,05	26,68	10,69	4,89	.	.	.
183	Heating	63	A	1	63A-1	85,86	25,15	3,22	3,9	.	28,51	52,43
184	Heating	63	A	2	63A-2	84,69	26,03	1,58	.	.	.	93,61
185	Heating	63	A	3	63A-3	85,46	26,3	.	.	.	110,21	214,63
186	Heating	63	A	4	63A-4	84,75	23,62	.	.	2,9	.	170
187	Untreatment	64	A	1	64A-1	85,47	26,55	23,35	.	.	.	.
188	Untreatment	64	A	2	64A-2	36,35	27,1	8,15	5,15	2,81	397,45	.
189	Untreatment	64	A	4	64A-4	83,56	24,3	4,89	2,43	.	.	.
190	Untreatment	64	A	5	64A-5	86,15	26,44	9,32	.	1,93	.	.
191	Steaming	65	A	1	65A-1	82,91	24,96	31,03	.	.	61,53	.
192	Steaming	65	A	2	65A-2	82,89	24,74	8,31	5,31	.	123,52	47,16
193	Steaming	65	A	3	65A-3	82,46	25,59	.	8,36	.	.	98,61
194	Steaming	65	A	4	65A-4	82,25	25,23	4,01	7,12	.	30,91	125,65
195	Heating	66	A	1	66A-1	84,35	24,92	24,08	40,97	1,83	92,76	87,34
196	Heating	66	A	2	66A-2	84,32	23,45	11,48	10,95	.	.	89,095
197	Heating	66	A	3	66A-3	69,19	25,24	.	.	.	.	33,88
198	Heating	66	A	4	66A-4	69,19	24,43	4,52	.	14,39	161,41	200,3
199	Heating	67	A	1	67A-1	92,86	24,33	3,04	.	.	.	160,45
200	Heating	67	A	2	67A-2	78,31	21,2	7,46	6,62	2,21	41,51	63,835
201	Steaming	67	A	3	67A-3	94,15	25,61	3,97	.	.	85,94	.
202	Untreatment	68	A	1	68A-1	85,29	24,95	18,54	2,33	.	.	.
203	Untreatment	68	A	2	68A-2	86,13	25,77	6,99	.	.	.	.
204	Untreatment	68	A	3	68A-3	86,91	26,28	5,05	.	.	.	.
205	Steaming	69	A	1	69A-1	75,51	25,94	.	.	.	.	38,78
206	Steaming	69	A	2	69A-2	92,13	26,05	3,59	5,56	.	.	255,84
207	Steaming	69	A	3	69A-3	91,26	26,65	4,68	9,45	.	.	153,47
208	Steaming	69	A	4	69A-4	89,8	27,23	6,66	4,68	3,02	.	114,76
209	Steaming	70	A	1	70A-1	96,81	24,34	4,94	1,78	.	.	187,62
210	Steaming	70	A	2	70A-2	95,48	24,56	5,01	3,88	.	.	158,96
211	Steaming	70	A	3	70A-3	70,89	26,37	3,43	9,99	.	440,08	74,67
212	Steaming	70	A	4	70A-4	83,25	26,35	2,78	11,82	2,04	40,33	98,4