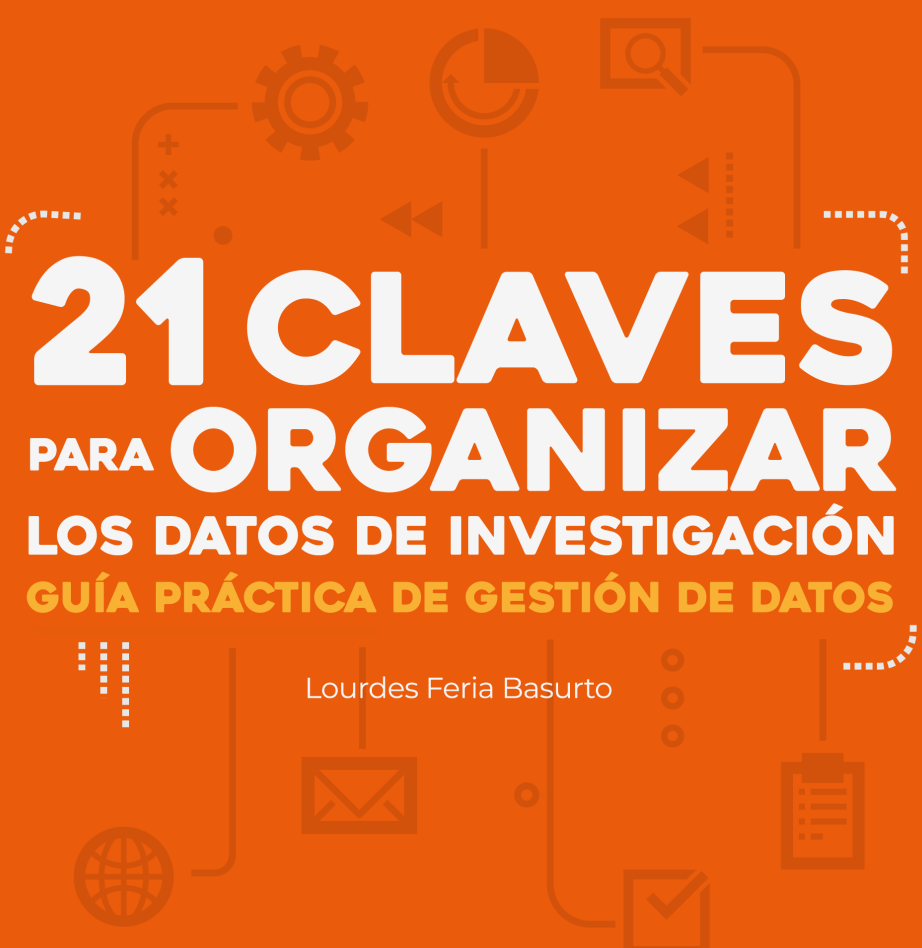




21 CLAVES

PARA **ORGANIZAR**

LOS DATOS DE INVESTIGACIÓN

GUÍA PRÁCTICA DE GESTIÓN DE DATOS

Lourdes Feria Basurto

21 claves para organizar los datos de investigación

Guía práctica de gestión de datos

buen^{En}plan

UNIVERSIDAD DE COLIMA

Dr. Christian Jorge Torres Ortiz Zermeño, Rector

Mtro. Joel Nino Jr., Secretario General

Mtro. Jorge Martínez Durán, Coordinador General de Comunicación Social

Mtra. Ana Karina Robles Gómez, Directora General de Publicaciones

21 claves para organizar los datos de investigación

Guía práctica de gestión de datos

Lourdes Feria Basurto



UNIVERSIDAD
DE COLIMA



PUERTABIERTA
EDITORES

© UNIVERSIDAD DE COLIMA, 2024
Avenida Universidad 333
Colima, Colima, México
Dirección General de Publicaciones
Teléfonos: 312 316 1081 y 316 1000, ext. 35004
Correo electrónico: publicaciones@uocol.mx
<http://www.uocol.mx>
ISBN electrónico: 978-607-8984-50-3
DOI: 10.53897/LI.2024.0050.UCOL
SE.11/317000/032/2024 Edición de publicación no periódica

© PUERTABIERTA EDITORES, S.A. DE C.V.
Ma. Refugio Morales No. 583, Col. El Porvenir, Colima, Col.
Tel: 312 312 1133
Para México: www.puertabierta.com.mx
Para España: www.puertabiertaeditores.com
ISBN electrónico: 978-607-8961-89-4

Derechos reservados conforme a la ley
Publicado en México | *Published in Mexico*



Este libro está bajo la licencia de Creative Commons, Atribución – NoComercial – CompartirIgual 4.0 Internacional (CC BY-NC-SA 4.0)
Usted es libre de: Compartir: copiar y redistribuir el material en cualquier medio o formato. Adaptar: remezclar, transformar y construir a partir del material bajo los siguientes términos: Atribución: Usted debe dar crédito de manera adecuada, brindar un enlace a la licencia, e indicar si se han realizado cambios. Puede hacerlo en cualquier forma razonable, pero no de forma tal que sugiera que usted o su uso tienen el apoyo de la licenciante. NoComercial: Usted no puede hacer uso del material con propósitos comerciales. CompartirIgual: Si remezcla, transforma o crea a partir del material, debe distribuir su contribución bajo la misma licencia del original.

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

You are free to: Share: copy and redistribute the material in any medium or format. Adapt: remix, transform, and build upon the material under the following terms: Attribution: You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use. NonCommercial: You may not use the material for commercial purposes. ShareAlike: If you remix, transform, or build upon the material, you must distribute your contributions under the same license as the original.

Imágenes ilustrativas: freepik.com
(slidesgo - macrovector - storyset y vectorjuice)

Proceso editorial certificado con normas ISO desde 2005
Dictaminación doble ciego y edición registradas en el Sistema Editorial Electrónico PRED

Registro: LI-005-24
Recibido: Febrero de 2024
Publicado: Noviembre de 2024

Agradecimientos:

Al Dr. José Román Herrera-Morales por la presentación de este libro.
A la Lic. Elizabeth Nashielli Márquez Sánchez por su colaboración para desarrollar el contenido del Capítulo 4.
A eScire por su colaboración, acompañamiento y apoyo.

Índice

▶ ¿Qué contiene este libro? -----	4
▶ Presentación -----	6
▶ Introducción -----	10
▶ Capítulo 1. Los datos -----	11
Clave 1. ¿Qué son, exactamente, los datos de investigación?	
Clave 2. Del <i>Big Data</i> al <i>Right Data</i>	
Clave 3. La manera de hacer investigación cambió: investigación intensiva en datos	
Clave 4. ¿Cuáles datos? Identificación de datos cualitativos y cuantitativos	
▶ Capítulo 2. La elegancia del orden -----	21
Clave 5. No te pierdas: elabora una Base de conocimiento	
Clave 6. Buenas prácticas vs. malas prácticas	
Clave 7. Léeme: la importancia de un fichero <i>README</i>	
Clave 8. Querido diario: tu Cuaderno de investigación	
Clave 9. ¿Metadatos en un caramelo?	
Clave 10. ¿Qué son los Libros de código?	
▶ Capítulo 3. Tu Plan de gestión de datos -----	48
Clave 11. Plan de gestión de datos: ¿qué es y para qué sirve?	
Clave 12. ¿Qué contiene un Plan de gestión de datos?	
Clave 13. ¿Cómo elaboro mi Plan de gestión de datos?	
Clave 14. Plan de gestión de datos y Plan de análisis de datos	
▶ Capítulo 4. Abrirse al mundo -----	67
Clave 15. Ciclo de vida de los datos científicos	
Clave 16. Datos abiertos y datos FAIR	
Clave 17. Formatos libres vs. formatos “prisioneros”	
Clave 18. Identificadores persistentes	
Clave 19. Licencias abiertas	
Clave 20. Reutilizar y citar datos	
Clave 21. Bibliotecas de datos de investigación (Repositorios)	
▶ ¿Qué aprendiste? -----	99
▶ Bibliografía complementaria -----	100

¿Qué contiene este libro?

Esta guía práctica te ayudará a responder las siguientes preguntas, de una forma clara y sencilla.

- **Capítulo 1** ¿Por qué organizar los datos es el desafío principal en la investigación, qué tipos de datos existen y cómo manejarlos adecuadamente?
- **Capítulo 2** ¿Cuáles son las ventajas de establecer un orden desde el principio, cómo mantenerlo, actualizarlo y no morir en el intento?
- **Capítulo 3** ¿Cómo ahorrar tiempo y esfuerzo en el proceso de investigación al estructurar, desde el inicio, un Plan de datos?
- **Capítulo 4** ¿Qué herramientas están disponibles gratuitamente para la gestión de tus datos y cómo acceder a datos científicos abiertos en el mundo?

“Las 21 claves
presentadas no son
simplemente directrices,
son estrategias para
aprovechar al máximo el
manejo de datos
en beneficio de la
investigación”.

Román Herrera-Morales

Presentación

Es un privilegio presentar este texto: *21 claves para organizar los datos de investigación. Guía práctica de gestión de datos*, escrito por una profesional excepcional a la cual tengo el honor de conocer y haber colaborado con ella desde hace más de dos décadas. Mi relación con la Dra. Lourdes Feria no sólo se basa en una conexión profesional, sino también en una profunda admiración por su dedicación, pasión y visión en el ámbito de las Tecnologías de la Información y la Gestión de Datos en el contexto de México y Latinoamérica.

La Dra. Feria, con su experiencia acumulada a lo largo de los años, destaca en esta obra un principio fundamental: “La gestión de datos comienza contigo”. Este enfoque personalizado refleja no sólo su perspectiva profesional, sino también su ética y compromiso con la jerarquización de la información. Las 21 claves presentadas no son simplemente directrices, son estrategias para aprovechar al máximo el manejo de datos en beneficio de la investigación.

Este libro resulta muy relevante en el contexto actual de una sociedad en constante transformación, donde la información no verificada se masifica, porque proporciona herramientas para la planificación de la gestión de datos, promoviendo la adopción de un enfoque estratégico en su uso. La autora resalta la importancia de los datos abiertos, los repositorios digitales y los principios “FAIR” (*Findable, Accessible, Interoperable, Reusable*), subrayando cómo fortalecen la transparencia y el valor de los datos en el ámbito de la investigación, favoreciendo el manejo responsable de la información, ayudando a encontrar con agilidad las fuentes, autores/as y anotaciones, manteniendo con pulcritud los vínculos con los objetivos de la investigación y generando estrategias claras para avanzar ordenadamente en el desarrollo de una tesis, una experimentación académica o un reporte científico.

Así mismo, frente a la exigencia de que en las instituciones se desarrollen directrices, recomendaciones y acciones académicas, legales y pedagógicas para que el plagio sea erradicado en todas sus formas y manifestaciones, las prácticas de datos responsables y pulcras constituyen una parte integral del trabajo científico, para garantizar que la información que se genere refleje un compromiso consciente con la veracidad, originalidad y rectitud que demanda la labor científica.

En resumen, esta obra meticulosamente estructurada es mucho más que una guía práctica, en la que cada capítulo, diseñado con precisión, invita a los lectores a responder preguntas y realizar ejercicios que no sólo amplían su comprensión para enriquecer el ámbito de la gestión de datos y favorecer su desarrollo profesional, sino que impulsan la conciencia de un compromiso ético más profundo en diversas etapas del ciclo científico.

Dr. J. Román Herrera-Morales



“La gestión de
datos empieza
contigo”.

Lourdes Feria Basurto

Introducción

Las 21 claves de este libro tienen por objeto mostrarte paso a paso la diferencia entre los tipos de datos y su ciclo de vida en los proyectos de investigación. Al revisarla, conocerás componentes básicos para organizar tus archivos, datos crudos, datos procesados, lecturas digitales, componentes de *software*, bitácoras y otros elementos, desde la primera investigación que realices.

Se trata de una guía que puedes leer de principio a fin o ir directamente al tema de tu interés. Su estructura te lo permite y la idea es ayudarte de la manera más sencilla posible en el manejo de datos, cada vez que lo requieras.

Es importante tener en cuenta que no es lo mismo la **Gestión de datos** para un proyecto de investigación que el **Análisis de los datos** que sustentan una investigación. “A pesar de su aparente simplicidad (o tal vez debido a ella), la gestión de datos a menudo se pasa por alto en las fases de planificación y recopilación de datos de la investigación. Y la literatura sobre el tema es limitada” (Guest, Namey y Mitchell, 2013).

En este sentido, el presente libro surge como respuesta a un vacío documental en la bibliografía especializada, sobre todo en español, donde las 21 claves que lo conforman son una guía para el adecuado manejo de los datos de toda investigación.



Capítulo 1.

Los datos

¿Qué encontrarás en este capítulo?

El primer paso para comprender algo es comenzar por su definición, los grandes conceptos nos llevan a grandes aprendizajes. Así que, como en toda historia, hay que empaparse de los antecedentes y conocer qué hay detrás, en este caso, de los datos. Para ello, en este capítulo se responde la pregunta de qué tratan los datos de investigación, las diferencias entre *Big Data* y *Right Data*, seguido de lo que implica una investigación intensiva, además de la identificación de datos cualitativos y cuantitativos, estructurados, no estructurados y mixtos.

Clave 1

¿Qué son exactamente los “Datos de investigación”?

El dato es el componente básico de toda investigación.

Vamos a poner un ejemplo: imaginemos que estamos construyendo un castillo con legos, a lo largo la construcción nos encontraremos con diferentes “bloques”, todos de un color y tamaño diferente, pero que al ensamblarse de cierta forma terminan por formar una estructura específica, en ese caso los “bloques” representarían el dato mientras que el “castillo” sería la estructura en la que fueron ordenados y analizados, es decir, la investigación completa en sí.

La Oficina de Administración y Presupuestos de Estados Unidos (OMB, por sus siglas en inglés), citada por la *National Science Foundation*, define a los datos científicos como material fáctico registrado, generalmente aceptado en la comunidad científica para validar los resultados de la investigación (NSF, 2020).

Pueden ser datos históricos, pero también pueden ser datos cruciales que acabas de levantar la semana pasada, ya sea para tu tesis doctoral, para un estudio que estés desarrollando o para acreditar un examen final.





Se entiende por datos todo lo que te proporciona información para sustentar tu trabajo académico.

En el mundo de los datos manejamos alguno de estos tipos: texto, cifras, imágenes, video, audio, *software*, algoritmos, animaciones, modelos, etc., y se trabajan en algunos formatos de los cuales aquí van sólo algunos ejemplos: PDF, JPG, CVS, MP3, MP4.

Sin buenos datos
no hay buena
investigación.

Clave 2

Del *Big Data* al *Right Data*

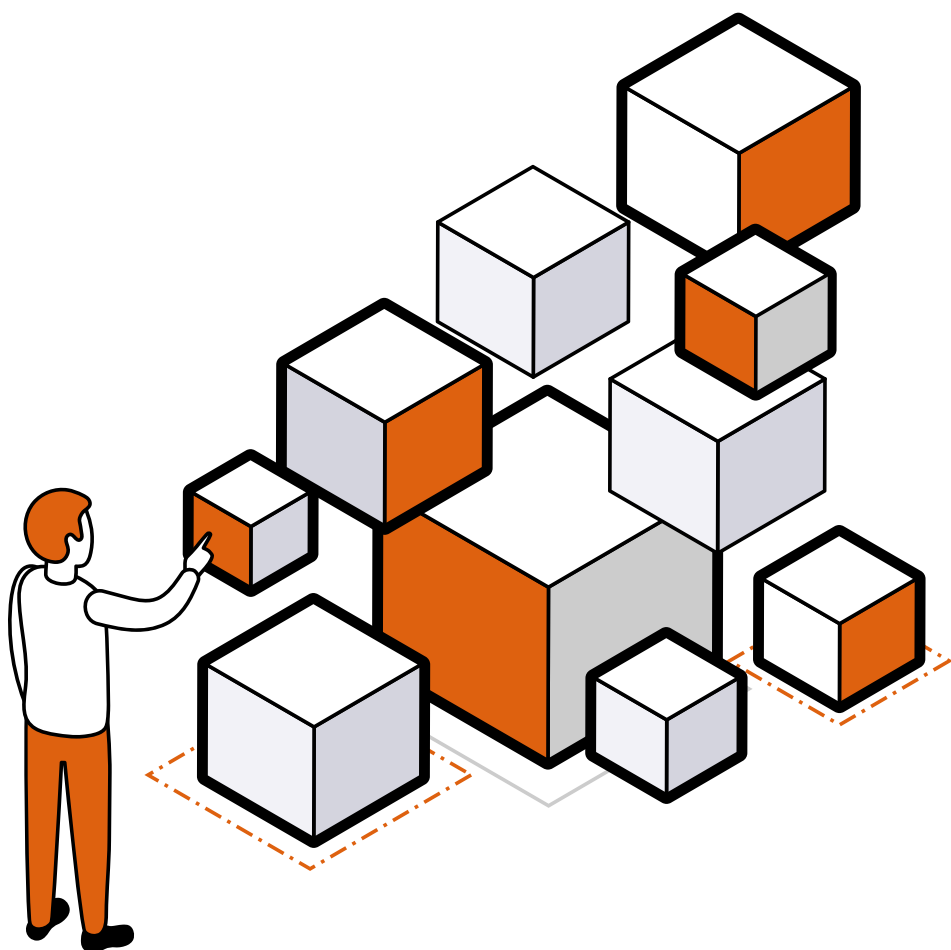
¿Qué es el *Big Data*? Es el término que se usa para referirse a la enorme cantidad de actividades que hacemos usando tecnología, de ahí nace el término, que se refiere a la recopilación de diversas y complejas colecciones de datos que se generan a partir de diversas fuentes, así como a la capacidad que tenemos de utilizar estos datos.

De hecho, al usar tu teléfono móvil estás creando, desde tu entorno, datos para alimentar el universo *Big Data*: al conversar en una videollamada o por Zoom, al enviar un mensaje de WhatsApp, hacer una compra en Amazon o elegir una serie en Netflix.

Todo eso genera datos, información y metadatos; por ejemplo, nombres de ciudades, horarios de uso de redes, productos que más se consumen y... ¡hasta emociones! ¿Quiénes pueden estar deprimidos? ¿Quiénes están felices? ¿Cuándo cumplen años? Y, por supuesto, datos sobre el movimiento del dinero, las formas de pago, los tipos de pago.

Cada movimiento
digital genera
Big Data.

El reto entonces es saber cómo puedo darle sentido y validar la veracidad en este mar de información. La respuesta es: haciendo investigación basada en buenos datos científicos; generando datos fidedignos y confiables; recabando, analizando y descubriendo los datos correctos, lo que suele llamarse Right Data.





Clave 3

La manera de hacer investigación cambió: investigación intensiva en datos

La investigación se ha apoyado siempre en datos que luego, con tu análisis, reflexión y discernimiento se irán transformando en información. En el pasado, la cantidad de datos disponibles no requería el uso de tecnología; sin embargo, hoy en día el *software* y las herramientas que permiten extraer conocimiento de enormes cantidades de datos producidos a través de experimentos y algoritmos va en aumento. La investigación intensiva en datos abarca una amplia variedad de métodos científicos y se caracteriza porque el procesamiento de esos datos se hace mediante análisis automatizados.



Ya sea que estés desarrollando un trabajo académico, tu tesis de grado o posgrado, o quizás una investigación postdoctoral, vas a recabar datos tanto mediante la revisión de la bibliografía, observaciones de laboratorio, datos de campo, entrevistas y otros métodos o mediciones.

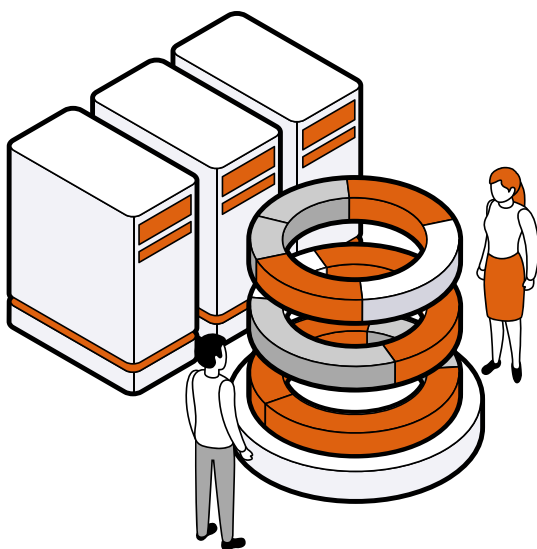
La investigación intensiva en datos demanda también una organización intensiva de los mismos y con esta guía vas a aprender cómo hacerlo.



Clave 4

¿Cuáles datos? Identificación de datos cualitativos y cuantitativos: datos estructurados, no estructurados y mixtos

Estamos rodeados de datos desde que despertamos hasta que nos vamos a dormir; datos meteorológicos, financieros, de redes sociales, geográficos, bibliográficos, datos personales, datos que arrojan las aplicaciones del celular y muchos otros.



En investigación se identifican sobre todo dos tipos de datos: cuantitativos y cualitativos, los primeros son numéricos y tienen que ver con información estructurada (por ejemplo, una estructura podrían ser las filas y columnas de una tabla); los segundos consisten en textos, narrativas, entrevistas, conversaciones, discursos, es decir, información no estructurada.

Los datos de investigación son los que recopilas, registras o creas para tu trabajo científico y se encuentran en soportes como los siguientes (NLM, 2020):

- Documentos, memorandos, notas, evaluaciones y protocolos de estudio
- Cuadernos y diarios de laboratorio y de campo
- Cuestionarios, transcripciones y Libros de código
- Cintas de audio, cintas de video y archivos .mp3
- Fotos, películas y archivos TIF
- Proteínas, secuencias genéticas y ADN clonado
- Respuestas de pruebas
- Contenido de las bases de datos
- Metodologías y flujos de trabajo
- Modelos, algoritmos y guiones
- Información técnica
- Procedimientos y protocolos operativos estándar
- Contenido de una solicitud
- Diapositivas, artefactos, especímenes, muestras, líneas celulares y organismos
- Hojas de cálculo

De todo ese
listado, ¿qué tipo
de datos utilizas?

Referencias

National Science Foundation (NSF). (2020, 27 de marzo). *Dear colleague letter: open science for research data*. <https://www.nsf.gov/pubs/2020/nsf20068/nsf20068.jsp>

National Library of Medicine (NLM). (2020). *Data literacy for the busy librarian: intro to data literacy and the research and data lifecycles* [Youtube]. www.youtube.com/watch?V=KP-T5Rbd22f8

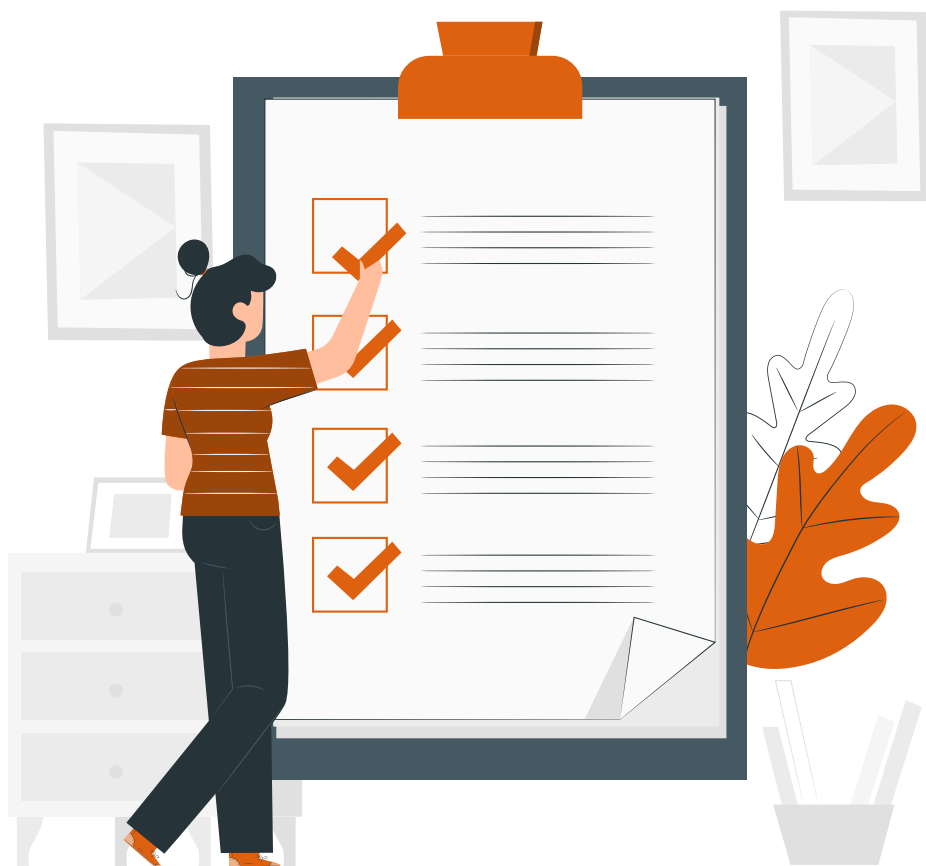


Capítulo 2.

La elegancia del orden

¿Qué encontrarás en este capítulo?

Mantener tus datos y documentos digitales organizados es fundamental en todo proceso de investigación. Es obvio, sí, pero por lo mismo no se le suele dar la importancia debida. Tener los archivos electrónicos ordenados aplica para tus documentos en la computadora, teléfono celular, en la nube y/o en un *software*; ya sea que estés trabajando de manera individual o con otras personas, es imprescindible tomar cartas en el asunto. Por eso, en este capítulo veremos cómo estructurar tu Base de conocimiento, cómo construir un Cuaderno de laboratorio, y cómo integrar ficheros *README* y Libros de código.



Clave 5

No te pierdas: elabora una Base de conocimiento

La información que vas recolectando sobre tu tema de investigación, los procesos de levantamiento de datos, las lecturas especializadas y otras fuentes constituyen tu activo más valioso.

Este tesoro de archivos de texto, tablas de Excel, PDF, correos electrónicos, imágenes y audios digitales y otro tipo de componentes ¡es tu Base de conocimiento! Y por tal motivo debe estar bien documentada desde el prin-

cipio para que siempre que lo necesites puedas encontrarla fácilmente.

A la Base de conocimiento también se le conoce como Árbol de carpetas o Mapa de archivos. Dicho de otro modo, es un sistema de clasificación.

Resulta muy útil porque reduce tu carga de trabajo, te permite reutilizar tus ideas y las de tu equipo, reciclar tus datos y te ahorra tiempo, sobre todo en el futuro. No hay una manera universal de hacer un Árbol de carpetas, pero sí hay elementos que es necesario considerar para el que piensas realizar. Desde luego, puedes encontrar tu información mediante el algoritmo de búsqueda, pero es mucho más eficiente dirigirte puntualmente a una carpeta o archivos específicos.

Hay quienes clasifican sus archivos por tipo de material, por ejemplo:

- Documentos en Word
- Documentos PDF
- Hojas de cálculo
- Vídeos
- Audios

Sin embargo, si vas a hacer una tesis o a escribir un artículo científico se recomienda organizarlos de acuerdo con tu índice, como se observa en el siguiente ejemplo:

Tesis

- Introd_y_Antecedentes
- Cap_1_Importancia_del_sector_agropecuario
 - Producción_Mundial
 - Principales_productos
- Cap_2_Evaluación_del_Sector
 - Subsidios
 - Barreras
 - Marco_institucional
 - Acuerdos_previos
 - Regulación_anterior
 - Estado_de_Tecnología
 - Biotecnología
 - Maquinaria
 - Fertilizantes
 - Riego
- Cap_3_Proceso_negociador
 - Programa
 - Declaración_Punta_del_Este
 - Propuestas
 - Reunión_Bruselas
 - Acuerdo_Blair_House
 - Acta_final
- Cap_4_Resultados
 - Panorama_General
 - Compromisos
 - Acceso_a_mercados
 - Subsidios_internos
 - Subsidios_exportación
 - Medidas_sanitarias_y_fitosanitarias
 - Productos_específicos
 - Efectos_del_Acuerdo
 - Países_desarrollados
 - Países_en_desarrollo
 - Otros
- Cap_5_Conclusiones_y_Propuestas
- Anexos
- Bibliografía



Clave 6

Buenas prácticas vs. malas prácticas

Al hacer un Árbol de archivos también se pueden tener malas prácticas. Ejemplo de éstas, sin duda, es el nombrar un archivo “mis datos” (en sentido estricto todos son “mis datos”). En otros casos sucede que nombramos al archivo como “mis datos buenos” porque los originales ya dejaron de ser “los buenos”, pero independientemente de que parece divertido es una mala práctica muy común.

Otro ejemplo de mala práctica es anotar: “año y data”, pero eso es demasiado escueto y para tu “yo del futuro” puede ser difícil de interpretar. A continuación, en la Tabla 1 podrás observar en la columna de la izquierda las malas prácticas más comunes en el nombrado y versionado de archivo. ¿Cómo puedes reemplazarla por una práctica correcta? Eso lo verás en la columna de la derecha.

Tabla 1. Lo bueno y lo malo

Mala práctica	Buena práctica
Gestionar tus datos hasta el final de la investigación (de lo contrario se pueden perder, contaminar, revolver).	Establecer un Plan de gestión de datos. Abarca desde el surgimiento de tu idea hasta el análisis y publicación de los resultados de la investigación, listo para compartirse y archivar adecuadamente para su conservación.
Nombrar los archivos y variables sin una estructura.	Crear una guía para nombrar y versionar tus carpetas y archivos, así como las variables e incluso los valores ausentes (<i>missing values</i>). Todo ello formará parte de una documentación general a conservar desde el inicio de tu investigación, con un Diccionario para dar consistencia.
Omitir documentar todo, incluidos los ajustes a tu plan original.	Preservar hasta el dato más insignificante (en apariencia). Puedes hacer un archivo <i>README</i>, un Libro de código o llevar un Cuaderno de investigación. En tu documentación admite datos acerca del reclutamiento, criterios de inclusión, recolección, procedimientos aplicados, seguridad de los datos, control de calidad y anonimización de los datos (limitar la exposición de datos sensibles) y, por encima de todo, los meta-datos.
No hacer una copia maestra (<i>master</i>) de tus datos.	Conservar tu grupo de datos original, con los datos crudos.
No trabajar con datos <i>tidy</i> o <i>long</i> .	Evitar que tus datos crezcan horizontalmente. Trabajar con datos <i>tidy</i> o <i>long</i> te permitirá organizar tus datos en forma de variable por columna y observación por fila.
No establecer una preparación para replicar tu proyecto de datos desde el inicio.	Proyectar una publicación que incluya el código utilizado y los datos ejecutables para su replicación.

Mala práctica	Buena práctica
Asumir la exactitud en los datos (olvidar que los datos vienen con valores ausentes, duplicados, contaminados, no extrapolables y valores no esperados).	Establecer un diseño de datos que permita gestionar y privilegiar su calidad desde el comienzo. Elo incluye revisar cada fila y columna, remover duplicados, separar columnas que fusionan más de una variable, anonimizar datos, renombrar variables acorde con el diccionario correspondiente, revisar tipos de variables, corregir códigos y todo aquello que te permitirá brindar un trato específico y estandarizado a tus datos. Documenta todos los cambios.

Fuente: elaboración propia, con base en Milanés (2023).

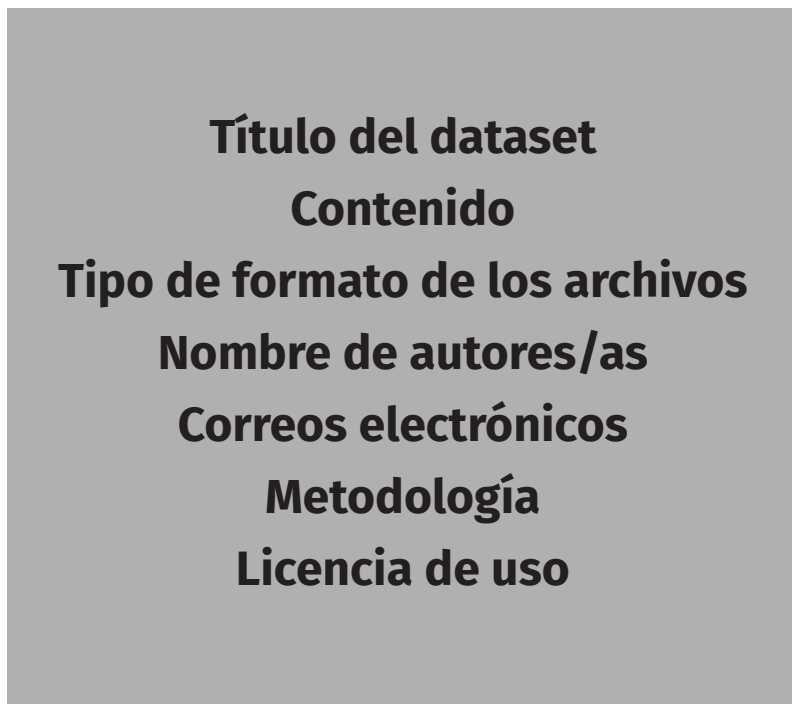
Clave 7

Léeme: la importancia de un fichero *README*

Nadie puede leer la mente ni adivinar la estructura de tu Base de conocimiento o las instrucciones para entender lo más básico. Para quienes trabajan con archivos de datos de investigación científica, todo Mapa de archivos, toda Base de conocimiento, tienen que contar primero que nada con un archivo denominado *README*.

El archivo *README* es un documento digital de texto, tú lo redactas y sirve para describir la estructura de tu Base de conocimiento. El *README* sirve de guía y en cualquier organización de archivos es la primera carpeta que tiene que aparecer. Se trata de un archivo de texto y lo recomendable es que su formato sea “.TXT”, donde estarás describiendo el contenido.

No necesariamente tiene que estar redactado ampliamente, basta con que incluyas los puntos fundamentales de los archivos que contiene tu Árbol de carpetas. Aquí se incluyen como ejemplo los elementos que sugieren las Biblioguías de la Universidad de Navarra (2021).



A continuación, te muestro un ejemplo ficticio donde aparecen enlistados los archivos. Se plantea un diseño experimental que se va a trabajar, pero tú puedes escribir el archivo *README* básicamente como quieras, no hay una norma rígida en términos de formato, pero debes cumplir con el contenido que te propones.

Título del dataset:

La agricultura y las negociaciones de la Ronda Uruguay

Contenido:

5 carpetas digitales (uno por cada capítulo) y sus subcarpetas.

1 carpeta de anexos conteniendo 3 archivos:

Indicadores de la producción agrícola 1980-1986.

Indicadores de superficie cultivada décadas 1980 y 1990.

Levantamiento de datos sobre el aprovechamiento de tierras por hectárea, 1992.

1 carpeta conteniendo el archivo de la bibliografía.

Tipo de formato de los archivos:

Los cinco archivos correspondientes a los capítulos y el archivo de bibliografía son todos “.doc”

El archivo “Anexos” contiene los distintos sets de datos utilizados en la tesis, todos ellos en formato “.csv”

Nombre y correo de los autores/as:

Juan Pérez, jperez@correo.edu

Metodología:

Se desarrolló en dos etapas, de las cuales la primera correspondería a una zonificación a nivel 1 de reconocimiento y utilizando material cartográfico a escala 1:50000. La segunda incluyó el estudio detallado de una o varias zonas agrícolas, constituyendo un área piloto representativo de algunas de las principales condiciones.

Licencia de uso:

Se autoriza la reproducción total o parcial de los sets de datos siempre y cuando se cite la fuente.

Fuente: elaboración propia.



En resumen: los archivos *README* son ficheros en los que se describe la información necesaria para que los conjuntos de datos sean comprensibles y reutilizables, pueden hablar de la autoría; si eres tú, puedes poner: “todos los archivos incluidos aquí son de mi autoría” o, en caso contrario, explicar quién es el responsable de los textos, el título, los nombres que van a usar para los archivos, la descripción (muy somera), la metodología de su investigación (también de manera muy breve), los proyectos financiadores, qué cobertura temporal y geográfica tiene su investigación, los derechos de uso de sus datos y los de privacidad, la descripción del contenido de las carpetas (muy breve), así como un listado de variables.

Clave 8

Querido diario: tu Cuaderno de investigación

Un Cuaderno de investigación es un documento físico o digital que en el que irás anotando el seguimiento de tu proyecto. Se conoce también con el nombre de Bitácora de datos, Diario de investigación o Cuaderno de laboratorio. Hay varios tipos de cuadernos de investigación y se pueden utilizar variedad de formatos, desde un cuaderno físico tradicional hasta un programa de computadora. Aunque parezcan redundantes, en realidad son muy útiles para organizar pensamientos y realizar investigaciones que puedan ser replicables.

¿Qué vas a registrar en el Cuaderno de laboratorio? Todos los datos y la información que vas generando en tu investigación; esto es: los procedimientos y acciones que realizas, los reactivos que utilizas, las observaciones que haces, y las ideas y pensamientos relevantes



que puedan ayudar a otro científico o a tu yo del futuro para reproducir los experimentos. Esto generalmente incluye una explicación del por qué se realizaron los experimentos, los antecedentes y las referencias necesarios, así como los resultados.

El Cuaderno no es un simple accesorio. Debes saber que es tu soporte legal porque documenta paso a paso tu trabajo en caso de que tu investigación contribuya a generar una patente; además, en el remoto caso de que se presentase algún reclamo de fraude contra un trabajo publicado, el Cuaderno de laboratorio te permitirá validar tus hallazgos y tu autoría intelectual.

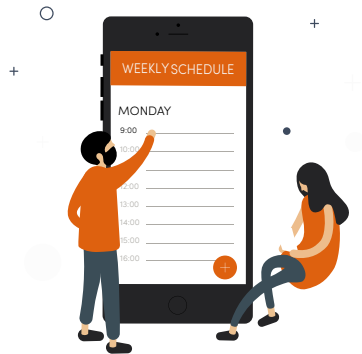
A continuación verás un ejemplo muy básico tomado de mi propio Diario de investigación, en el que escribí:

Hoy 14/02/2019 se obtuvieron en total 173 fichas bibliográficas, elaboré una tabla sencilla con los metadatos temáticos, la cita de los artículos encontrados, la mención de repetidos y los elegidos para leer (Feria Basurto, 2019).

No hay un formato específico para elaborar un Cuaderno de investigación porque dependerá de la especialidad en la que trabajas, pero en todos los casos se recomienda al inicio crear una portada con el título del proyecto, tu nombre y el nombre del investigador a cargo, así como la fecha de inicio.

Todo esto lo puedes hacer en Word u otra aplicación digital de texto como Evernote, One Note, Google Docs o incluso la opción de Notas del celular. Además, puedes usar alguna App, pues algunas disciplinas y áreas de conocimiento ya las están empezando a utilizar.

Por ejemplo, los agrónomos emplean Hortisys, la cual no solamente les deja escribir su libreta digital, sino que les controla en tiempo real sus cultivos, en sitio o a distancia, para que puedan ser monitoreados. Te recomiendo buscar en internet si existe alguna aplicación especializada para tu disciplina.



Contar con un Cuaderno de este tipo es una maravilla, sobre todo llegado el momento de comenzar a redactar la tesis o el paper, ese instante en el que te enfrentas a “la hoja en blanco” cuando ya vas a empezar a escribir y tienes la pantalla lista para recibir el primer párrafo y pues... no sabes ni qué poner; entonces, gracias al Diario de investigación ya tienes componentes-base para empezar el proceso de redacción, partiendo de tus propias notas, de tu propia información, a partir de los datos que has recabado o creado.

En algunas instituciones los cuadernos se consideran como el legado de los científicos de ese laboratorio. Incluso llega a suceder que, aún mucho después de que el investigador se haya ido a otra universidad, su cuaderno permanece y sigue siendo referenciado: así es como se convierte en un material bibliográfico referenciable y pueden actualizar permanentemente lo que se había desarrollado.

Clave 9

¿Metadatos en un caramelo?

Cuando tomas una fotografía con tu teléfono móvil de inmediato queda registrada en un sistema automático que le asigna un nombre de archivo, hora y fecha de captura, resolución, datos de iluminación, ubicación o georreferencia, si se usó flash o se tomó con luz natural, entre otros. Estas especificaciones se conocen con el nombre de metadatos.

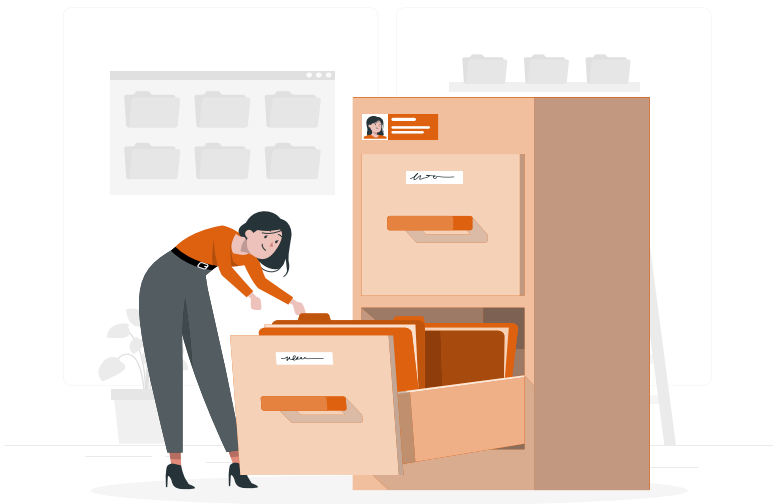
La mayoría de los productos que encuentras en el supermercado también presentan sus metadatos: ¡incluso un caramelo, en la envoltura te los muestra! (Por ejemplo, peso: 5 gramos; ingrediente: azúcar o fructosa; sabor: menta).



Así, tus fuentes de análisis, tus carpetas, documentos digitales y tus archivos deberían tener metadatos. Las revistas, libros, textos, páginas web, blogs y videos también cuentan con esta información.

Los metadatos se asignan a lo largo de todo el ciclo de vida de tu investigación, desde que haces tu recopilación previa de datos hasta que los analizas y, finalmente, los publicas.

Los metadatos son,
entonces, datos
acerca de un objeto
físico o virtual.



Los metadatos son, entonces, etiquetas acerca de un objeto físico o virtual y su finalidad es describir y recuperar información en el mundo de la presencialidad, en tu computadora o en la web, porque hay una cantidad impresionante de datos en el espacio que nos rodea. Por ello, es imperativo que al trabajar con tus fuentes de información las puedas recuperar de manera correcta y útil; es decir, que la tengas sistematizada.

A medida que vas trabajando con tus datos de investigación codificarás y analizarás tus lecturas, tu referencial teórico, ya sea con base en archivos de texto o con conjuntos de datos. El nombre que asignas a cada uno de tus archivos es un metadato, pero también los que se enlistan a continuación.

Metadatos son las palabras descriptivas sobre tus fuentes.

Ejemplo: nombre, ciudad, puesto, institución.

Metadatos son los nombres que usas para las clasificaciones de archivos.

Ejemplo: nombre del archivo, autor/a, programa utilizado, fecha, resolución.

Metadatos son todos y cada uno de los elementos de un registro bibliográfico (usando normas APA u otras).

Ejemplo: título, autor/a, año, editorial, páginas, DOI o URL general.

Metadatos son los términos que asignas a tus variables en una tabla de Excel.

Ejemplo: sexo, género, estado civil, condiciones de vivienda.

Metadatos son las palabras clave con las que vas señalando tus anotaciones, tus comentarios en lecturas, en redes sociales, artículos, noticias o encuestas.

Ejemplo: fecha, ubicación, tipo, lenguaje, categoría.

Fuente: elaboración propia.

Con toda probabilidad, ya estás capturando metadatos básicos mínimos sobre tu investigación. Tus cuadernos de laboratorio y archivos de investigación contienen mucha, si no es que toda, esta información:

- Nombre del investigador/a
- Fecha
- Proyecto
- Detalles del experimento/análisis que se está ejecutando (incluidos el propósito y los métodos utilizados)
- Fuentes de otros datos utilizados en el experimento/análisis

La clave es recopilar toda la información necesaria (metadatos) mientras trabajas y luego vincular esos metadatos a los propios archivos de datos. Ten presente siempre que los metadatos robustos son:

- **Consistentes**
- **Significativos** para ti y para tu equipo
- **Fáciles** de localizar e identificar

Y, por último, si vas a ser tú la única persona que usa los datos, es posible que los metadatos no necesiten estar muy estructurados para que sean útiles. Sin embargo, existen estándares y normas formales para trabajos que así lo requieran.

Clave 10

¿Qué son los Libros de código?

Conocido también por su nombre en inglés *codebook*, el Libro de código es una bitácora, una lista de claves.

Se utiliza en la investigación social, así como en las ciencias básicas y aplicadas y su finalidad es proporcionar una estructura ordenada a tus datos para mantener claridad durante todo el proceso.

Esta herramienta contiene el origen y los criterios para incluir, o no, los datos de una investigación en curso, es decir, genera un documento que contiene información detallada de una fuente de datos, tales como nombres de variable, etiquetas de variable, de valores y/o valores perdidos, entre otros.

De esta manera, el Libro de código permite conocer el contenido, estructura y diseño de un estudio, lo cual facilita la comprensión de las variables en la base de datos de tu investigación. Un Libro de código es, en este sentido, una “hoja de ruta” que permite a los investigadores identificar patrones, tendencias y relaciones entre segmentos de datos mediante la asignación de etiquetas o códigos significativos.

En general, los Libros de código contienen los siguientes elementos:

Nombre de variable: nombre o numeración signada a cada variable en la base de datos, que pueden ser abreviaciones mnemotécnicas o patrones alfanuméricos.
--

Definición: descripción breve que permite identificar la variable.

Tipo de variables: cualitativas, numéricas, alfanuméricas, categóricas, otras.

Instrucciones de recopilación: se refiere a las pautas que se habrán de tener en cuenta para la recolección de los datos.
--

Fuente: Libro de código para variable en base de datos (ISSP, 2009).

Es importante considerar que cada Libro de código se adapta a las necesidades de cada investigación. A continuación verás ejemplos para tres disciplinas distintas, uno para antropología, otro para biología y uno más para informática.

Finalmente, un Libro de código es un reto para los investigadores ya que siempre se puede adaptar según avancemos en nuestro objeto de estudio, permitiendo actualizar los valores según las necesidades de la investigación (véase OAI, 2023).

Ejemplo de un *codebook* para un estudio antropológico

Variable	Definición	Tipo de variable	Instrucciones de recolección
Prácticas de caza y pesca	Descripción detallada de las técnicas, herramientas y creencias asociadas con la caza y la pesca en la comunidad.	Cualitativa (descriptiva)	Realizar entrevistas semiestructuradas y observaciones participantes para documentar las prácticas tradicionales. Registrar técnicas, mitos y su impacto en la comunidad.

Variable	Definición	Tipo de variable	Instrucciones de reco-pilación
Ritos de paso	Descripción de los rituales y ceremonias relacionadas con eventos significativos como nacimientos, matrimonios y funerales.	Cualitativa (descriptiva)	Entrevistas a informantes clave y participación en ceremonias para documentar rituales. Registrar símbolos, canciones y creencias.
Cosmovisión	Creencias, mitos y valores que explican la relación de la comunidad con el universo, la naturaleza y lo sobrenatural.	Cualitativa (descriptiva)	Entrevistas en profundidad con ancianos y líderes espirituales para comprender la cosmovisión. Registrar mitos y creencias.

Variable	Definición	Tipo de variable	Instrucciones de recolección
Sistemas de parentesco	Estructuras y términos que describen las relaciones familiares y sociales dentro de la comunidad.	Cualitativa (descriptiva)	Utilizar árboles genealógicos y entrevistas para entender los sistemas de parentesco. Registrar términos y roles familiares.
Uso de plantas medicinales	Plantas utilizadas con propósitos medicinales, describiendo su preparación y aplicación.	Cualitativa (descriptiva)	Entrevistas con curanderos y observaciones participantes para documentar el conocimiento sobre plantas medicinales. Registrar nombres, preparación y condiciones tratadas.

Fuente: elaboración propia con OAI (2023).

Ejemplo de un *codebook* para Biología

Variable	Definición	Tipo de variable	Instrucciones de reco-pilación
Longitud del alar (en milímetros) de la mariposa Monarca	La medida desde el extremo de un ala hasta el otro.	Numérica (continua)	Utilizar un calibrador digital para medir la longitud del alar de mariposas seleccionadas aleatoriamente. Medir tres veces para obtener un promedio preciso.
Coloración del abdomen	Descripción del color y patrón en el abdomen de la mariposa.	Categórica	Asignar un código numérico según la escala de colores establecida para describir la coloración del abdomen (1 = marrón claro, 2 = marrón oscuro, etc.).

Variable	Definición	Tipo de variable	Instrucciones de recopilación
Genotipo de la mariposa Monarca	Combinación específica de alelos en genes seleccionados.	Alfanumérica	Utilizar técnicas de biología molecular para determinar el genotipo de las mariposas en relación con genes específicos asociados con la migración y la reproducción.
Altura del hábitat de cría (en metros sobre el nivel del mar)	La altitud del lugar donde las mariposas Monarca ponen sus huevos y crían a sus larvas.	N Numérica (discreta)	Utilizar un dispositivo de GPS para registrar la altitud del hábitat de cría de cada mariposa observada.

Variable	Definición	Tipo de variable	Instrucciones de reco-pilación
Temperatura ambiente (en grados Celsius)	La temperatura del entorno donde se encontró la mariposa.	Numérica (continua)	Utilizar un termómetro digital para medir la temperatura ambiente en el lugar de observación de cada mariposa.

Fuente: elaboración propia con OAI (2023).

Ejemplo de un *codebook* para Informática

Variable	Definición	Tipo de variable	Instrucciones de reco-pilación
Integridad de la red	Indica si la red ha sido comprometida o no.	Binaria	Determinar la integridad mediante herramientas de seguridad informática. Registrar como 1 si no se detectan intrusiones, 0 si se detecta una intrusión.

Variable	Definición	Tipo de variable	Instrucciones de reco-pilación
Velocidad de la conexión	La velocidad de transferencia de datos de la red.	Numérica (continua)	Utilizar herramientas de prueba de velocidad para medir la velocidad de conexión en diferentes momentos del día. Registrar los resultados en Mbps.
Nivel de encriptación	Indica el tipo y nivel de encriptación utilizado en la red.	Categorica	Determinar el nivel de encriptación de la red inalámbrica a través de configuraciones del router y <i>software</i> de seguridad. Registrar el tipo y nivel de encriptación utilizado.

Variable	Definición	Tipo de variable	Instrucciones de recopilación
Dispositivos conectados	Número total de dispositivos conectados a la red.	Numérica (discreta)	Utilizar herramientas de administración de red para contar el número de dispositivos conectados en un momento específico del día.
Frecuencia de actualización de contraseñas	La frecuencia con la que se actualizan las contraseñas de la red.	Numérica (discreta)	Entrevistar a los administradores de red para determinar la política de actualización de contraseñas. Registrar el período de tiempo (en días) después del cual se requiere actualizar las contraseñas.

Fuente: elaboración propia con OAI (2023).

Referencias

OAI. (2023). *Chat GPT. Texto generado: Ejemplos de libros de código para antropología, biología e informática.*

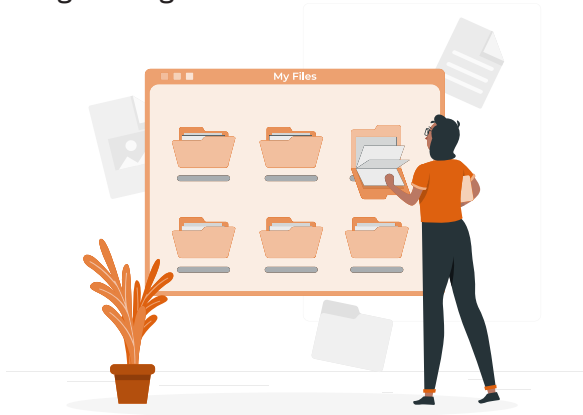
Feria Basurto, L. (2019). *Diario de investigación postdoctoral. Instituto de Investigaciones Bibliotecológicas y de la Información.* UNAM.

The International Social Survey Programme (ISSP). (2009). *Social Inequality.* <https://issp.org/>

Milanés, Y. (2023, 11-13 de enero). Errores vs. buenas prácticas en gestión de datos de investigación. En Ministerio de Salud de Perú (Coord.), *Curso Internacional de Bibliotecas Médicas y Promoción Científica.* Perú. <https://zenodo.org/record/7540861#.Y-PqoS-ZNPY>

Moreno, M. A., Egan, K y Brockmann, L. (2011). Development of a Researcher Codebook for Use in Evaluating Social Networking Site Profiles. *Journal of Adolescent Health, 49*(1), 29-35. <https://doi.org/10.1016%2Fj.jadohealth.2011.04.015>

Universidad de Navarra (2021). *Datos de investigación, gestionar datos. Biblioguías.* <https://biblioguias.unav.edu/datosdeinvestigacion/gestion>





Capítulo 3.

Tu Plan de gestión de datos

¿Qué encontrarás en este capítulo?

Ya conoces la importancia de organizar los datos y las ventajas de mantener un orden para su cuidado, ahora prepárate para conocer qué es un Plan de gestión de datos (PGD), cómo se elabora y cómo puede ser de utilidad para tu labor dentro de una investigación, tesis, etcétera. Un PGD es prácticamente un salvavidas para ti y para quienes retomen tus pasos en busca de nuevos conocimientos a partir de tu investigación.

Clave 11

Plan de gestión de datos: ¿qué es y para qué sirve?

¿Qué es?

El Plan de gestión de datos (PGD) es un documento formal elaborado por el investigador, el tesista o grupo de investigación. El PGD describe todos los aspectos de la gestión de datos (CEPAL, 2020).

¿Para qué sirve?

Es para estimar tiempos, evitar imprevistos, prever recursos y costos, así como garantizar la buena clasificación y el almacenamiento ordenado para una posible reutilización y compartición de los datos.

Hacemos un PGD para nuestro uso personal, considerando además que las agencias financiadoras y las universidades le dan cada vez más valor, por la integridad científica y transparencia que representa, a la par de propiciar la accesibilidad y reutilización.

Es importante revisar los requerimientos de tu universidad o del organismo que apoyará tu investigación, porque cada una podría contener distintos requisitos.



Clave 12

¿Qué contiene un Plan de gestión de datos?

El contenido va a depender del tipo de datos que recolectarás. Generalmente describirá qué clase de datos vas a reunir, cuáles serán tus fuentes, ¿recolectarás los datos en algún laboratorio, mediante experimentos, aplicarás simulaciones informáticas, lo harás mediante observación directa en alguna comunidad?

Cabe aclarar que cada universidad, organismo, institución tiene su propio modelo de Plan de gestión de datos.

Figura 1. Formato PGD

Plan de gestión de datos	
Nombre del proyecto (título de la tesis o artículo científico)	
Nombre de la escuela o facultad	
Nombre de la universidad o institución	
Fecha	
Generalidades y resumen	
Investigador(a)	
Título del proyecto	
Tutor(a)	
Duración del proyecto	
Resumen	
Definición de los datos	
Procedencia	
Frecuencia de obtención de nuevos datos	
Cantidad de datos a ser generados	
Formatos de los datos	
Responsable(s) de los datos	

Tipo de datos que se generarán			
Tipos de datos de investigación	Especificaciones	Opciones de software	Tipo de archivos (extensión)

Procesamiento y almacenamiento de datos	
Software que se usará para crear, analizar y visualizar datos	
Cantidad aproximada de los datos y de su crecimiento estimado	
Compartición y propiedad	
Documentación y metadatos	
Almacenamiento a corto plazo	

Fuente: elaboración propia.

Por lo general, un PGD incluye, además del encabezado, una diversidad de información que puedes observar en la Figura 1, integrada en cuatro secciones: Generalidades y resumen, Definición de los datos, Tipo de datos que se generarán, Procesamiento y almacenamiento de datos.

Clave 13

¿Cómo elaboro mi Plan de gestión de datos?

Vamos a describir cada uno de los elementos del formato (Figura 1). Comenzaremos con el Encabezado en el que, además de decir explícitamente que se trata de un Plan de gestión de datos también se incluye el título del proyecto de investigación (trátase de una tesis, artículo científico u otra variante), el nombre de la facultad y, finalmente, el nombre de la institución que alberga el proyecto.

A continuación, observarás cuatro grandes secciones:

- a) Generalidades y resumen.
- b) Definición de los datos.
- c) Tipo de datos que se generarán.
- d) Procesamiento y almacenamiento de datos.

Detengámonos en los elementos que muestra cada una de ellas.

a) Generalidades y resumen

En este punto vas a indicar la información básica, podríamos decir que es la cápsula, el plan “en una nuez”. Por ello incluirás tu nombre, nuevamente el título del proyecto, el nombre de tu tutor (si es que lo hay), la duración estimada del proyecto y un resumen que destaque lo más significativo de lo que será tu jornada de levantamiento y análisis de los datos. Puedes incluir los detalles referentes a tu experimento, los métodos que utilizarás y el contexto en el que se enmarca tu investigación; dicho de otro modo, se refiere al qué, cómo, dónde, cuándo y por qué recolectarás los datos.

Ejemplo:

Investigador (a)	Carmona R., Alicia; Domene, Víctor
Título del proyecto	Participación de jóvenes en espacios de innovación en bibliotecas públicas en comunidades semi-urbanas del municipio de Santa Teresita.
Tutor(a)	Martínez Rivera, Damián

Duración del proyecto	8 meses
Resumen	Con el objetivo de reconocer el impacto de la apropiación social de la biblioteca en el municipio, particularmente entre las y los jóvenes, se estructuró una metodología que buscará enlazar en un mismo sistema lógico tanto el conocimiento especializado como la observación, medición, análisis y sistematización de información. A partir de ello, se definió el esquema de trabajo que considera cuatro dimensiones: Investigación bibliográfica en bases de datos científicas; Investigación cualitativa en campo mediante la técnica de entrevista a profundidad; Observación participante; y Aplicación de encuestas.

Fuente: elaboración propia.

b) Definición de los datos

Aquí vas a anotar, primero, de dónde proceden tus datos, de tal manera que te respondas preguntas como



las siguientes: ¿Vas a tomarlos de fuentes bibliográficas? ¿Vas a salir a campo a reunirlos? ¿Serán mediciones y observaciones de laboratorio?

A continuación, vas a indicar la frecuencia, es decir, la periodicidad con la que vas a estar haciendo el levantamiento de datos; posteriormente, harás un cálculo del espacio digital que necesitarás para tus archivos, teniendo presente, como dice William Michener (2015), que el volumen es básico: ¡no es lo mismo 10 Tb que 10 Mb!

En la penúltima fila vas a registrar los formatos de tus datos, ya sean de texto (doc, txt, pdf, xml, rtf); numéricos (csv, xls); video (mpeg-4, quicktime); imagen (tiff, jpg, gif, JPEG2000); audio (wav, mp3); o algún formato especializado. Por último, identificarás las responsabilidades. ¿Por qué? Porque especialmente en las investigaciones en equipo siempre se sabrá a quién dirigirse para la información puntual en torno a los datos, de lo contrario, dicha comisión se diluye.

En este punto anotarás quién va a ser la persona que se haga cargo de los datos y los roles que cada cual desempeña; por ejemplo: la recolección, el registro, el control de calidad de los datos y metadatos, los respaldos y la incorporación de los sets de datos en un repositorio.



Para ello, en algunos casos es de utilidad contratar a un asistente de investigación comprometido y dedicado con el proyecto.

Ejemplo:

Procedencia	Bases de datos de las escuelas de la comunidad, Google Scholar, Gestores bibliográficos, Cuestionario cuantitativo y Observación de campo a través de grabaciones de entrevistas y grupos focales que serán posteriormente transcritas.
Frecuencia de obtención de nuevos datos	Semanal, durante los primeros cuatro meses de la investigación.
Cantidad de datos a ser generados	1 gigabyte
Formatos de los datos	MP3, Docx, Excel, Nvo
Responsable(s) de los datos	Audios - Carmona R., Alicia Docs - Carmona R., Alicia Excel - Domene, Víctor

Fuente: elaboración propia.

c) Tipos de datos que se generarán

En esta sección indicarás qué tipo de datos vas a generar, sus especificaciones, el *software* con el que los procesarás y, como producto de ello, los archivos que resultarán.

Ejemplo:

Tipo de datos de investigación	Especificaciones	Opciones de <i>software</i>	Tipo de archivos (extensión)
Cualitativos	Se entrevistará a 26 jóvenes y se harán las transcripciones	Aplicación del celular para grabar audio • Word • NVivo	.mp3 .docx .nvo
Cuantitativos	Se encuestará a través de <i>Survey Monkey</i>	Excel	.csv

Fuente: elaboración propia.

d) Procesamiento y almacenamiento de los datos

Este apartado es vital, ya que definirás cómo organizarás tus datos y si éstos estarán disponibles o no. Para este punto ya tienes tu Base de conocimiento (ver Clave 5) y tu fichero *README* (ver Clave 7); habrás nombrado y versionado correctamente tus archivos.

El *software* que se usará tal vez sea el mismo para crear, analizar y visualizar los datos, pero en ocasiones suele ser diferente para cada caso.

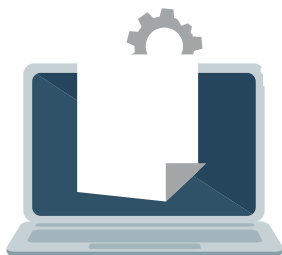
También anotarás la cantidad de datos que producirás, así como un cálculo de su posible crecimiento. Los detalles de este proceso los encontrarás en Definición de los datos (ver Clave 13), pero no está por demás repetirlo.

En la sección referente a Compartir tus datos tendrás que tomar una decisión, basada en las políticas de los proyectos de datos que maneja tu universidad u organización. Este recuadro deberá responder a las preguntas siguientes: ¿Tu información quedará confidencial? ¿Permitirás la diseminación en un proyecto personal o en una web? ¿Distribución de tus datos sólo por correo electrónico previa solicitud? ¿Los publicarás en un repositorio abierto? ¿Enviarás tus conjuntos de datos (todos o sólo una parte) como apéndices o suplementos mediante algún artículo de revista científica? ¿Publicarás los datos, metadatos y código como data paper? (Michener, 2015).

Documentación y metadatos

Los metadatos serán tu llave para localizar (descubrir), usar y citar correctamente tus datos. Además, incrementan las oportunidades de que les asegures larga vida y de que tú y/u otras personas los reutilicen.

En el capítulo 2 vimos que los metadatos son muy importantes porque ayudan a saber cómo nombraste, estructuraste y almacenaste tus archivos, datos y documentos. En lo posible, indicarás los metadatos que utilizarás para tu trabajo con tablas, hojas de cálculo y otras herramientas. Para este punto ya tienes tu Base



de conocimiento (Clave 5) y tu fichero *README* (Clave 7), y habrás nombrado y versionado correctamente tus archivos.

Almacenamiento a corto plazo

Desde un inicio prevé dónde vas a guardar tus datos. En el corto plazo podrías conservarlos sólo en tu equipo de cómputo personal, en la nube (especificando el URL), o en una memoria externa.

Almacenamiento de largo plazo

En este punto define tu estrategia de preservación describiendo cuánto tiempo estarán accesibles los datos una vez terminada la tesis o el proyecto, anota cómo los irás guardando mientras dure la investigación y cómo los conservarás para usos posteriores. Esto ayudará a que tu “yo del futuro” no tenga que adivinar: “¿Dónde quedó el archivo digital de mi tesis?” Además, permíteme recordarte, aunque resulte obvio, que se nos olvida la importancia de hacer respaldos, hasta que un día ocurre un imprevisto como:

- El disco duro se contamina
- Los documentos se pierden
- Las URL se rompen
- Los medios físicos de almacenamiento se degradan (Michener, 2015)

Hasta aquí el formato para el llenado de Plan de gestión de datos. Por último, aunque no todos los formatos de gestión de datos de investigación lo solicitan, hay dos aspectos importantes para tener en cuenta: los costos y los aspectos éticos.

Ejemplo:

Software que se usará para crear, analizar y visualizar los datos	Transcripciones en NVivo. Analizar los datos cualitativos en NVivo. Capturar datos cuantitativos a partir de <i>Survey Monkey</i> e integrarlos en Excel.
Cantidad de los datos y de su crecimiento estimado	1 gigabyte (aproximado).
Compartición y propiedad	Todos los datos serán libres para ser compartidos de manera abierta a quienes pudieran estar interesados en reutilizarlos, si es posible, para usar el modelo y mejorarlo.
Documentación y metadatos	Los datos se sistematizarán en dos entregables que serán depositados en el repositorio Zenodo como conjunto de archivos digitales y documentos de texto. Para los datos cuantitativos se asignarán metadatos en una plantilla previa en Word a los sets de datos (al momento de elaborar este plan aún no se han definido las etiquetas en su totalidad). Al subir los datos al repositorio ahí se capturarán los metadatos requeridos basándose en la plantilla y complementando los que sean necesarios.
Almacenamiento a corto plazo	Los datos se guardarán en formato docx, nvo y csv, y se respaldarán en una memoria externa que exclusivamente contenga esa documentación. Los archivos se organizarán a partir de una carpeta denominada Doc 20190301 con su respectivo archivo <i>README</i> .
Almacenamiento en el largo plazo	Al concluir la investigación, los datos se almacenarán en el repositorio de la institución (si existiese), de lo contrario, se mantendrán en el repositorio de datos abiertos Zotero.

Fuente: elaboración propia.

Costos

Todo levantamiento de datos tiene un costo, aunque sea mínimo. Pregúntate cuál va a ser la cantidad que se requiere invertir para ello, aunque no siempre sea obvia. Incluye gastos en *software* o aplicaciones, equipo, transportación y otros.

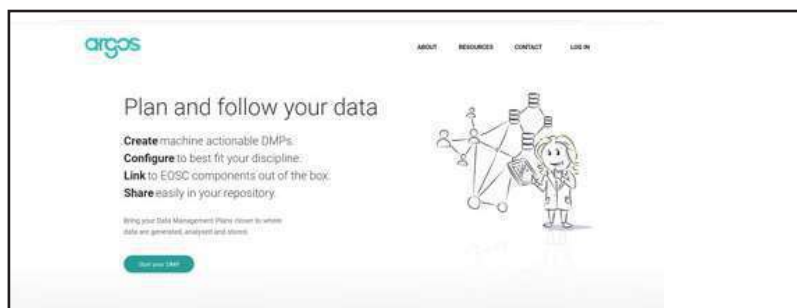
Ética

El comportamiento ético siempre va a fortalecerse profesionalmente. Por ello, considera y haz explícita tu postura al elaborar tu Plan. Aquí deberás tener en cuenta los aspectos de licenciamiento e intercambio de tus materiales, tus planes para limitar el uso de los mismos y, particularmente, todo lo que tiene que ver con el acceso a datos sensibles que pudieran poner en riesgo a los sujetos de tu investigación.

Por último, si quieres ver los formatos que usan otros organismos y universidades puedes dar una visita a los siguientes sitios:

Ejemplo 1

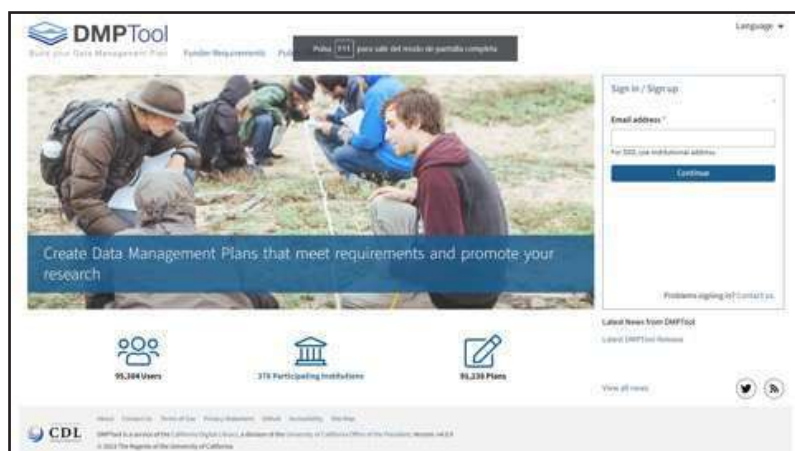
Argos. Es un servicio abierto que simplifica el manejo, validación, monitoreo y mantenimiento de planes de gestión de datos.



Fuente: argos.openaire.edu

Ejemplo 2

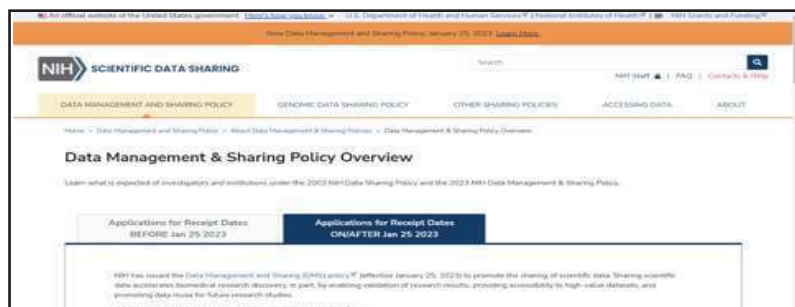
DMP Tool. Es una aplicación en línea, de libre acceso, dispuesta para que investigadores puedan crear sus planes de datos.



Fuente: dpmtool.org

Ejemplo 3

United States National Institutes of Health (NIH). Visita esta unidad para conocer cómo los institutos de salud en Estados Unidos manejan su política de manejo y distribución de datos.



Fuente: sharing.nih.gov

Clave 14

Plan de gestión de datos y Plan de análisis de datos: ¿que alguien me explique!

¿Es lo mismo un **Plan de gestión de datos** que un **Plan de análisis de datos**?

No, son dos documentos distintos que se utilizan en diferentes etapas.

Como has leído hasta ahora, el Plan de gestión se orienta a administrar los datos durante todo el proyecto, mientras que un Plan de análisis se enfoca en cómo se analizarán los datos una vez que se hayan recopilado.

Cada tipo de plan se utiliza en diferentes etapas del proyecto de investigación.

Un Plan de gestión de datos de investigación se elabora al principio del proyecto y describe cómo se recolectarán, almacenarán, procesarán, preservarán y compartirán los datos durante y después del proyecto. En otras palabras, se trata de un plan para gestionar los datos de investigación de manera efectiva y eficiente. Este plan también suele incluir información sobre las políticas y los requisitos institucionales o legales relacionados con la gestión de datos.

Por otro lado, un Plan de análisis de datos de investigación se elabora después de haber recopilado los datos y describe cómo se analizarán y presentarán para responder a las preguntas de investigación. En otras palabras, se trata de un plan para analizar los datos de manera sistemática y rigurosa. Este plan también

puede incluir información sobre las herramientas y técnicas de análisis que se utilizarán, así como sobre los plazos y las responsabilidades para su realización.

En resumen, mientras que un Plan de gestión pone énfasis en cómo se administrarán los datos durante todo el proyecto, un Plan de análisis se enfoca en cómo harás las observaciones, el razonamiento y la reflexión en torno a los datos una vez que se hayan recopilado. Ambos planes son importantes para garantizar la calidad y la integridad de los datos de investigación, y deben ser elaborados de manera clara y precisa para asegurar su efectividad.



En esta guía no abordaremos cómo elaborar un Plan de análisis, ya que su realización depende de cada disciplina y está en función del tipo de datos que se manejen; por ejemplo, es muy distinto el análisis para los datos cualitativos que para los cuantitativos y mixtos.

Por ahora, baste sólo decir que un Plan de análisis de datos es importante porque es la receta paso a paso que te irá guiando en el procesamiento de la información que recabas para tu investigación. En el siguiente cuadro puedes ver algunas otras razones.

Importancia del Plan de análisis de datos

Ayuda a establecer objetivos claros	Un Plan de análisis de datos ayuda a definir los objetivos y metas de un proyecto. Esto es importante porque si no se tiene una meta clara, es difícil saber qué datos se necesitan para alcanzarla.
Facilita la identificación de los datos necesarios	Un Plan de análisis de datos ayuda a identificar qué datos se necesitan para alcanzar los objetivos del proyecto. Esto es prioritario porque si se recolectan datos que no son importantes, el análisis resultante puede ser irrelevante.
Ahorra tiempo y recursos	Al tener un Plan de análisis de datos bien definido se puede ahorrar tiempo y recursos al no tener que recolectar y analizar datos que no son necesarios. Esto permite enfocar los esfuerzos en los datos que son importantes y relevantes para el proyecto.
Ayuda a tomar decisiones informadas	Un Plan de análisis de datos ayuda a tomar decisiones basadas en datos.

Fuente: elaboración propia.

En conclusión, un Plan de análisis de datos es importante porque ayuda a establecer objetivos claros, identificar datos necesarios, ahorrar tiempo y recursos y tomar decisiones informadas y precisas.

Para finalizar este capítulo te recomiendo revisar las 10 recomendaciones para el Plan de datos que William K. Michener (2015) hace en su texto: *Ten Simple Rules for Creating a Good Data Management Plan*. Si lo prefieres, puedes consultar la síntesis de este documento en una presentación en español disponible en:

https://es.slideshare.net/Lourdes_Feria/10-recomendaciones-plan-de-datospdf



Referencias

Comisión Económica para América Latina y el Caribe (CEPAL). (2020). *La gestión de datos de investigación*. <https://biblioguias.cepal.org/gestion-de-datos-de-investigacion>

Michener, W. K. (2015). *Ten Simple Rules for Creating a Good Data Management Plan*. <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004525>

Feria, L. (2015). Reseña y traducción del artículo de W.K. Michener “*Ten Simple Rules for Creating a Good Data Management Plan*”. https://es.slideshare.net/Lourdes_Feria/10-recomendaciones-plan-de-datospdf



Capítulo 4.

Abrirse al mundo: *Open Data*

¿Qué encontrarás en este capítulo?

Ahora que has conocido este mundo de los datos, deberás saber que el objetivo es que dichos datos sean abiertos, es decir, libres para ser encontrados y utilizados por colegas y público en general. Para ello, existen una serie de elementos a fin de cumplir con los requerimientos que diversos repositorios y expertos han establecido.

Clave 15

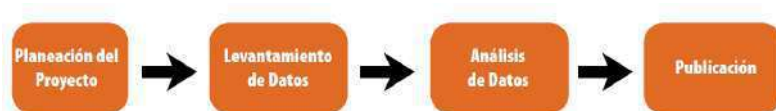
Ciclo de vida de los datos científicos

Tus datos de investigación son tu tesoro. A tus datos, ya sean números, imágenes o palabras les has invertido tiempo, desvelos y atención. Por ello, debes cuidarlos antes, durante y después del proceso; vigilar que estén limpios, bien organizados y en buen estado para garantizar que tu investigación sea confiable y que los resultados sean correctos para poder reutilizarlos y, ¿por qué no?, compartirlos con otros colegas.

Tradicionalmente, cuando se hacía investigación se seguían estos cuatro pasos que tal vez tú ya has puesto en práctica o lo harás cuando prepares tu tesis o algún artículo científico:

- Planeación del proyecto (anteproyecto de investigación)
- Recopilación de datos
- Análisis de datos
- Publicación

**Figura 2. Ciclo de vida de los datos.
Modelo base**

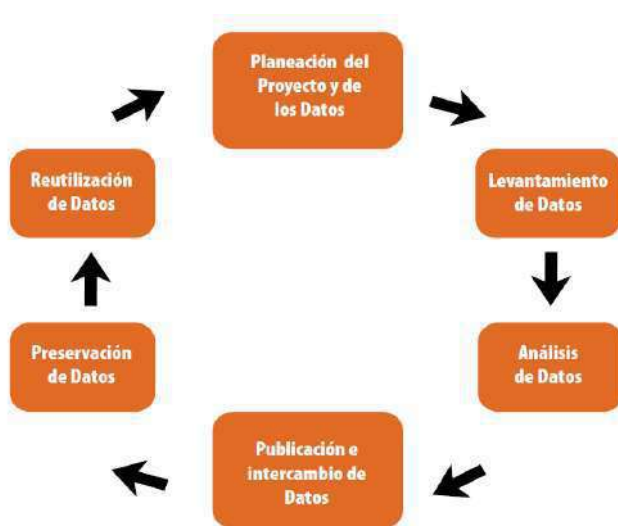


Fuente: elaboración propia con base en Briney (2015).

En la medida en que nos hemos ido familiarizando más con el entorno digital, han aparecido tres elementos más (los he puesto en negritas para resaltarlos); la planeación del proyecto ahora se enriquece con el Plan de datos y el ciclo de vida ha pasado de ser lineal a circular:

- Planeación del Proyecto y de los Datos (incluye no sólo anteproyecto de investigación sino el **Plan de datos**)
- Levantamiento de Datos
- Análisis de Datos
- Publicación e **Intercambio de Datos**
- Preservación
- **Reutilización de Datos**

Figura 3. Ciclo de vida de los metadatos.
Modelo actual

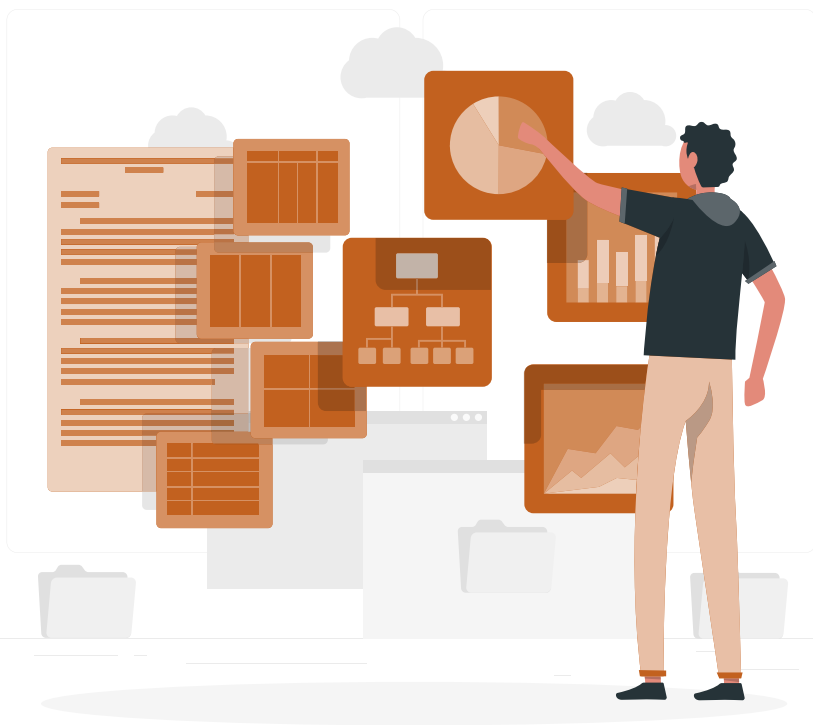


Fuente: elaboración propia con base en Briney (2015).

Vamos a ver de manera muy breve estos puntos:

Planeación del Proyecto y de los Datos. Se refiere a la elaboración de tu “mapa de ruta”, donde identificas cómo vas a organizar y guardar tus datos para que sean fáciles de entender y compartir de manera responsable. Puedes volver a leer el capítulo 3 para recordar los detalles.

Levantamiento de Datos. En esta etapa realizas lecturas científicas, sales a campo, ejecutas trabajo de laboratorio, haces experimentos, observaciones o aplicas otras metodologías de adquisición de datos. De manera simultánea vas organizándolos y documentándolos adecuadamente, como se explica en el capítulo 2.



Análisis de datos. Una vez que has recabado la información, aplicarás las herramientas y técnicas específicas para reflexionar, razonar, discernir y procesar tus datos. Esto puede incluir el análisis sistemático de bibliografía, la aplicación de métodos estadísticos, la creación de gráficos para visualizar y descubrir patrones o incluso la realización de modelos computacionales. Todo ello para la interpretación de esos datos y, a partir de ello, obtener resultados y conclusiones.

Publicación e Intercambio de Datos. Llegado el momento es importante comunicar tus datos de manera clara y precisa en impreso, digital o ambos formatos, a través de publicaciones científicas, presentaciones en conferencias u otros medios. Esto permite que tus colegas y pares puedan tener acceso y conocer los hallazgos de tu investigación y que, de esa manera, el conocimiento científico siga creciendo.

Preservación. Para que tus datos científicos no se pierdan y para asegurar su integridad y disponibilidad en el futuro, deberás respaldarlos y protegerlos para el largo plazo. En el entorno de los datos digitales masivos esto implica almacenarlos de manera segura y tener la certeza de que están protegidos contra pérdidas y/o daños, así podrás reutilizarlos, transparentarlos y contribuir a la reproducibilidad de la ciencia.

Reutilización de Datos. Este concepto, llamado también “reproducibilidad,” es importante en la ciencia porque ayuda a que los resultados de una investigación sean confiables y que otros investigadores puedan replicar los mismos resultados, con lo cual tendrán más credibilidad y certeza; además, se podrán construir nuevas observaciones sobre los hallazgos previos.

Clave 16

Datos abiertos y Datos FAIR

El movimiento conocido con el nombre genérico de Ciencia Abierta implica compartir conocimiento, datos, resultados y recursos de investigación de manera accesible, transparente y colaborativa. Abrir tus datos es una manera de participar de esta dinámica que permite generar ideas, nuevos hallazgos, aprender de diferentes personas e incluso trabajar en equipo.

“Datos abiertos” son conjuntos de información (*data-sets*) que se ponen a disposición del público en forma gratuita y en formatos accesibles para su uso y reutilización. Por ejemplo, tus tablas de Excel, tus entrevistas completas o incluso tu Cuaderno de investigación.



La decisión de transparentar los datos puede dar mayor visibilidad a tu trabajo científico, hacer que otros profesores, medios de comunicación y la sociedad en general te conozcan y se interesen en el tema que estás trabajando. Para ello, deberás cuidar que tus datos sean FAIR.

¿Qué es eso de FAIR?

Desde principios del siglo XXI la comunidad científica internacional adoptó las siglas FAIR (así, con mayúsculas compactas), para designar a los datos que sean Fácilmente Accesibles, Interoperables y Reutilizables (siglas de: *Findable, Accesible, Interoperable, Reusable*); cabe señalar que *Fair*, en inglés, significa “justo”. Dicho de otra manera, los datos se consideran FAIR cuando son localizables con facilidad, están organizados, estructurados y formateados de tal manera que se puedan usar y combinar con otros datos de manera eficiente.

La práctica de la apertura de datos para compartirlos fue un factor importante en la rápida respuesta científica frente a la pandemia de COVID-19, favoreciendo el desarrollo de vacunas efectivas en un tiempo récord.

Los datos abiertos permitieron a los científicos acceder a información relevante sobre la secuencia genética del virus, los ensayos clínicos y los resultados de las pruebas en forma rápida y colaborativa, lo que aceleró la comprensión del virus, la forma en que se transmite y se dispersa, su diagnóstico y tratamiento. Por supuesto, hay otros factores como la financiación, la inversión en investigación y desarrollo, la capacidad de producción, la comunicación entre expertos, etc., pero abrir los datos fue crucial.

Para solicitar fondos de apoyo a proyectos, las agencias internacionales cada vez están requiriendo más a los científicos la apertura de sus planes y conjuntos de datos. También las revistas especializadas empiezan a solicitar a las y los autores incluir, junto con su texto, los datos base como una condición para ser publicados, con lo cual se está generando la necesidad de participar en este tipo de prácticas.

Si estás considerando la posibilidad de compartir tus datos abiertamente es importante que conozcas la ruta que debes seguir para construir tus datos de manera FAIR. Ésta incluye el tratamiento ordenado, la adecuada selección, la posibilidad de permitir la descarga en un formato compatible y la elección del repositorio o biblioteca digital de datos confiable. De todo ello trata no sólo este capítulo, sino en general las 21 claves de esta guía.

A continuación, te dejo tres recomendaciones a tomar en cuenta al compartir datos en un repositorio abierto:

**1. Ser accesibles
y reutilizables
para reproducir
y comprobar
resultados**

Para que los datos sean abiertos tienen que estar disponibles para volver a ser usados con la finalidad de reproducir y comprobar resultados. Para esto, es fundamental que los datos que se comparten abiertos estén completos, tengan la información necesaria que les dé contexto y estén libres de cualquier restricción técnica.

2. Permitir su uso para cualquier otro fin científico	A partir de los datos originales debería ser posible hacer otro estudio, construir un <i>software</i> , hacer un análisis, publicar un gráfico, pero para eso también se requiere que el acceso sea completo, sin restricciones legales y que los datos abiertos tengan las licencias correspondientes.
3. Sin restricciones para su distribución y difusión	No debe haber ningún tipo de restricción financiera para que puedan ser utilizados con cualquier entidad para uso y en beneficio de la sociedad.

Fuente: elaboración propia.

Nota importante: no todos los datos científicos pueden ni deben ser compartidos públicamente, porque a veces refieren información privada o confidencial, como son los datos personales, la información de especies protegidas o en peligro de extinción; o puede darse el caso de que no haya suficientes recursos o tecnología para compartirlos de manera abierta.

Por esta razón, siempre es recomendable redactar lo que se conoce como “Declaración de disponibilidad de datos” (*data availability statement*) que acompañe a todo set de datos. Dicha declaración es un párrafo donde se informa acerca de cómo las personas interesadas pueden acceder a esa información; por ejemplo, dónde se encuentran almacenados, cómo acceder a ellos y, si hay restricciones de acceso, explicar cuáles son. Esta práctica varía de acuerdo a cada disciplina científica.

Clave 17

Formatos libres vs. formatos “prisioneros”

Los formatos libres son todos aquellos datos abiertos que puedes utilizar citando la fuente, mientras que los formatos “prisioneros” se refieren a los datos no disponibles para el público en general.

Cada tipo de archivo digital tiene su propio formato con características específicas. Usar formatos compatibles con diferentes dispositivos y programas de computadora es lo más recomendable cuando se trata de compartir, ya que el objetivo es procesar y visualizar correctamente y que sea legible en el *software* y sistemas más utilizados por tus colegas. A eso se le llama “Interoperabilidad” y es parte del concepto FAIR.

Algunos de los más conocidos son:

Formatos digitales

Formatos de texto	TXT, PDF, DOC/DOCX, RTF, HTML, XML; CSV
Formatos de imagen	JPEG, PNG, GIF, TIFF, SVG
Formatos de audio	MP3, WAV, FLAC, AAC
Formatos de video	MP4, AVI, MOV, MKV, WMV, FLV

Fuente: elaboración propia.

Trabajar con formatos abiertos es como usar un idioma que todos pueden entender, mientras que los formatos cerrados sólo pocas personas los manejan; aunque sean abiertos a nivel legal y a nivel de búsqueda, acceso, sin restricciones financieras y técnicas, de nada sirven si el contenido no puede ser reutilizado. Un ejemplo de ello son las tablas con datos numéricos en PDF.



Si se trata de una tabla sencilla quizá puedas darte el tiempo de transcribirla, pero si contiene decenas o incluso miles de registros no sería factible, aunque se cuente con el permiso legal, o sea la licencia, y aunque no tengan un cobro y sea totalmente accesible en términos económicos, técnicamente se considera “cerrado” porque usar esa información presenta limitaciones. Con ello en cuenta, debes cuidar que esto no suceda con tus datos.

En este punto me gustaría comentarte que la World Wide Web Consortium (W3C) es el organismo que rige y propone la mayoría de los estándares de internet. Esta entidad desarrolló el concepto de calificación con estrellas para los datos abiertos, con la idea es que entre más estrellas le puedas poner a tu *dataset* significará que es más libre.

Esquema de estrellas para *Open Data*

1 estrella	Ofrecer los datos en cualquier formato, aunque sean difíciles de manipular, como un PDF o una imagen escaneada.
2 estrellas	Entregar los datos de manera estructurada, como en un archivo Excel con extensión XLS.
3 estrellas	Entregar los datos en un formato que no sea propietario, como CSV en vez de Excel.
4 estrellas	Usar URIs (que es una dirección web de un dato que sirve para enlazarlo con otros datos) para identificar cosas y propiedades, de manera que se pueda apuntar a los datos. Requiere usar un estándar RDF.
5 estrellas	Vincular sus datos con los de otras personas, dotándolos de contexto. En la práctica, refiere a que la información entregada también apunte a otras fuentes de datos. Por ejemplo, si publico información sobre un libro dentro de mis datos, enlazar descripciones del libro que pertenezcan a otros publicadores (por ejemplo: DBPedia o Amazon).

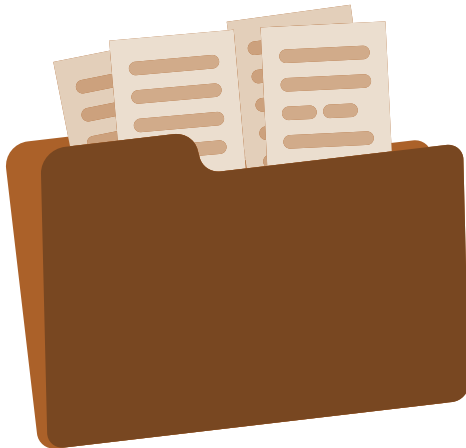
Fuente: Biblioteca del Congreso Nacional de Chile (s.f.).

Para tener otra referencia, si publicas tus datos y los haces accesible sin costo ya son abiertos, pero en realidad sólo tienes una estrella porque vienen un PDF o vienen en un JPG, o en algún formato que a otra persona le va a implicar volver a capturar la información. Ahora bien, si los pasas a un formato que muestre los datos de manera estructurada y que al menos permita

hacer filtros, columnas y filas, aunque el *software* para usarlo requiere una licencia como puede ser Excel, Oracle, o algún sistema de este tipo, entonces sólo alcanzas 2 estrellas.

Si mantienes datos estructurados, pero en vez de tenerlos en un formato propietario como podría ser XLS los ofreces en otro ya totalmente abierto, como ocurre con un texto CSV separado por comas, esto tiene la ventaja de que ya no necesitas una licencia, porque lo puedes abrir en Unix, en Linux, Windows o en el sistema que sea, entonces te ganas otra estrella y así consecutivamente.

Inicia con una estrella. Ya verás que poco a poco irás aprendiendo cómo llegar al máximo nivel; por lo pronto, es importante que conozcas la escala de estrellas porque así tendrás una referencia para ir haciendo cada vez más libres tus conjuntos de datos.



Clave 18

Identificadores persistentes

Nadie en el mundo tiene la misma huella digital que tú. Por esa razón se crearon los identificadores persistentes. En el mundo de la investigación y la publicación científica un identificador persistente es una clave especial, una etiqueta única que se asigna a un objeto digital, a las y los autores en ciencia y a las instituciones.

Para los fines de esta guía nos concentraremos en el identificador ORCID (*Open Researcher and Contribution ID*). Se trata de una etiqueta para las personas con la cual podrás tener tu perfil en línea donde aparezca tu información, los artículos que has escrito, los proyectos en los que has trabajado y otros datos relevantes. Así, tu trabajo será fácil de encontrar y compartir con otros investigadores, sin temor a que te confundan con otra persona que podría llamarse igual que tú.

Como podrás imaginar, el problema de los nombres es antiguo; por ejemplo, un autor que se llame Pedro Pérez Ramírez, que podría aparecer también como Pedro Pérez-Ramírez, o Pedro Pérez R. Por eso surgió la iniciativa ORCID, para que el identificador te dé la certeza de que el Pedro Pérez Ramírez que estás buscando sea justamente el autor que necesitas. ORCID te ayuda a distinguirse de otros investigadores que puedan tener nombres similares, con lo cual se evita la confusión y los errores en la identificación de autoría en publicaciones académicas y otros documentos, lo que garantiza que el trabajo sea correctamente atribuido.

Tener tu ORCID es relevante por estas cinco razones:

- Es tu identificación única como investigador en cualquier parte del mundo.
- Te abre posibilidades de accesibilidad y visibilidad porque te permite tener tu perfil académico público en línea.
- Te permite conectarte con otras plataformas y sistemas de gestión de la investigación.
- Es válido en cualquier universidad o país.
- Genera confianza y credibilidad porque ha sido ampliamente adoptado por la comunidad académica y científica internacional.

¿Tú ya tienes tu ORCID? Si no es así lo puedes obtener de manera muy sencilla y sin costo en el sitio <https://orcid.org/>

Una vez que lo hayas conseguido podrás usarlo como identificador personal, la clave siempre será la misma y te permitirá rastrear todos tus trabajos sin que los sistemas los confundan.



Fuente: Orcid.org

Por último, sólo unas cuantas líneas para comentarte breve-

mente acerca de otros dos identificadores persistentes:

DOI. Se trata del *Digital Object Identifier* (DOI). Al asignar un DOI a un artículo de revista, a un video de YouTube o a una infografía, a un libro o capítulo de libro, así sabremos que al difundir ese DOI, ponerlo en nuestras citas, referenciarlo en algún lugar, etc., siempre vamos a llegar al destino donde se encuentra.

La gran ventaja que tiene un DOI, además de simplemente ser un identificador persistente es que no sólo viaja con una URL, sino que se difunde en internet acompañado de todos los metadatos que describen al set de datos. Entonces, el DOI no sólo se difunde como una dirección, sino que lleva implícito el registro de quién es la o el autor, el nombre del conjunto de datos, es decir, todos los demás metadatos. Eso, adicionalmente, ayuda a posicionar de mejor manera los trabajos, las personas y las instituciones. Aquí tienes algunos ejemplos:

DOI de libro	https://doi.org/10.47557/AWLT6398
DOI de capítulo de libro	https://doi.org/10.1016/J.JENVP.2008.10.004
DOI de artículo	https://doi.org/10.1017/aee.2018.32

ROR. *Research Organization Registry* es el directorio de organizaciones de investigación, la clave única que se asigna a las instituciones que experimentan también el mismo problema de los nombres de las personas, sólo en este caso alguien usa las siglas y las escribe con o sin puntos, otros ponen el nombre completo, quizás alguien lo pone en inglés ¡hasta agotar todas las variantes posibles!, pero utilizando el identificador persistente de las organizaciones se evitan confusiones. Mira aquí tres ejemplos:

☒ Active
☐ Inactive
☐ Withdrawn

<https://ror.org/0460jpj73>
University of Concepción

ORGANIZATION TYPE Education	LOCATION Concepción (GeoNames ID 3893894) Chile
OTHER NAMES UdeC, Universidad de Concepción	OTHER IDENTIFIERS GRID grid.5386.e ISNI 0000 0001 2298 9683 Crossref Funder ID 501100006255 Wikidata Q199948
WEBSITE https://www.udec.cl/pexterno/	

Fuente: Ror.org

7376 Organizations
Are we missing an organization you're looking for? [Submit a request to add it](#)

Record status
☒ Active
☐ Inactive
☐ Withdrawn

<https://ror.org/04znxe670>
University of Colima

ORGANIZATION TYPE Education	LOCATION Colima (GeoNames ID 4913516) Mexico
OTHER NAMES People's University of Colima, Universidad de Colima	OTHER IDENTIFIERS GRID grid.412687.0 ISNI 0000 0001 2375 8971 Crossref Funder ID 109819264 Wikidata Q2495731
WEBSITE http://www.ucof.mx/	

Fuente: Ror.org

7507 Organizations
Are we missing an organization you're looking for? [Submit a request to add it](#)

Record status
☒ Active
☐ Inactive
☐ Withdrawn

<https://ror.org/01tmp8f25>
Universidad Nacional Autónoma de México

ORGANIZATION TYPE Education	LOCATION Mexico City (GeoNames ID 3539597) Mexico
OTHER NAMES UNAM, National Autonomous University of Mexico	OTHER IDENTIFIERS GRID grid.9486.3 ISNI 0000 0001 2500 0681 Crossref Funder ID 5011000088739 Wikidata Q222738
WEBSITE http://www.unam.mx	
RELATIONSHIPS (1)	

Fuente: Ror.org

DOI	ROR
Sus siglas significan <i>Digital Object Identifier</i>	Sus siglas significan <i>Research Organization Registry</i>
Es un enlace estable y permanente para localizar objetos digitales y dónde localizarlos	Localiza organizaciones a través de metadatos
Se asigna a artículos de revistas electrónicas, libros, capítulos de libros, actas y comunicaciones de congresos, <i>software</i> , videos	Identifica qué organizaciones están asociadas con qué resultados de investigación
Es un código alfanumérico que sigue la norma NISO Z39.84, DOI Syntax	Es un identificador de organizaciones de investigación y que contiene metadatos asociados
Permite la interoperabilidad con otras plataformas, repositorios o motores de búsqueda	Permite usar la afiliación para describir cualquier relación formal entre un investigador y una organización asociada con investigadores
Se debe solicitar el número DOI a las agencias de registro correspondientes, por ejemplo, CrossRef	ROR es interoperable con otros identificadores. Las ID de ROR se asignan a GRID, ISNI, Crossref Funder ID y Wikidata

Fuente: elaboración propia.

Clave 19

Licencias abiertas

Tus conjuntos de datos son tu creación, como lo son también tus escritos, tus gráficos, tablas y todo el material que vas generando como producto de investigación con tus ideas. Tu trabajo y tu redacción son tuyos y tienes el derecho de decidir quién y cómo lo puede usar. Por ello, cuando se trata de compartir tus datos abiertos es recomendable conocer las licencias de propiedad intelectual.

Estas licencias son etiquetas basadas en acuerdos y normas internacionales que establecen cómo se pueden usar tus materiales y tú asignas los derechos de creación. Son los permisos, dentro de ciertos límites, que otorgas para compartir y reutilizar tus datos de investigación responsablemente.

¿Compartir? Sí, pero con reglas claras. Por ello, lo más recomendable es usar las licencias más reconocidas que son las de *Creative Commons*. Esta es una organización internacional sin fines de lucro que apoya a las personas para que participen del universo de los bienes comunes del conocimiento y la cultura compartidos.

Creative Commons surgió en 2001, se podría traducir como “Bienes comunes creativos”, ofrece seis tipos de licencias, desde la más permisiva hasta la más restrictiva.



En la siguiente tabla se describen:

Tipos de licencias *Creative Commons*

Reconocimiento (BY)
Se permite cualquier explotación de la obra, incluyendo una finalidad comercial, así como la creación de obras derivadas, la distribución de las cuales también está permitida sin ninguna restricción.
Reconocimiento – No Comercial (BY-NC)
Se permite la generación de obras derivadas siempre que no se haga un uso comercial. Tampoco se puede utilizar la obra original con finalidades comerciales.
Reconocimiento – No Comercial – Compartir Igual (BY-NC-SA)
No se permite un uso comercial de la obra original ni de las posibles obras derivadas, la distribución de las cuales se debe hacer con una licencia igual a la que regula la obra original.
Reconocimiento – No Comercial – Sin Obra Derivada (BY-NC-ND)
No se permite un uso comercial de la obra original ni la generación de obras derivadas.
Reconocimiento – Compartir Igual (BY-SA)
Se permite el uso comercial de la obra y de las posibles obras derivadas, cuya distribución se debe hacer con una licencia igual a la que regula la obra original.
Reconocimiento – Sin Obra Derivada (BY-ND)
Se permite el uso comercial de la obra, pero no la generación de obras derivadas.

Fuente: elaboración propia con base en <http://creativecommons.org>

¿Cómo crear una licencia CC?

- Entrar al sitio <https://creativecommons.org/>
- Eliges *Share your work* (Comparte tu trabajo).
- Vas al recuadro, presionas *Get started* (Empezar).
- Eliges las características.
- Ingresas la información opcional (puede ser únicamente el título de tu documento y tu nombre) y... ¡listo! Ya puedes usar el ícono para etiquetar tu trabajo.

Clave 20

Reutilizar y citar datos

La reutilización o *data sharing* es otra buena práctica que forma parte del ciclo de vida de los datos. Se refiere a la reproducibilidad que cada vez es más necesaria en la ciencia, pero ¿a qué se refiere la reproducibilidad? Como su nombre lo indica, a repetir con los mismos datos el mismo experimento. Un ejemplo de ello sería hacer disponibles los datos crudos* de nuestra investigación.



*¿Qué son los datos crudos? Son todos aquellos datos que recabaste en tu investigación, organizados, depurados y sistematizados, pero sin haber sido aún interpretados o procesados.

Si compartes, otros colegas podrán poner a prueba tus hallazgos y ver qué tan posible es replicarlos, con lo cual quedaría legitimada tu investigación o, en caso contrario, se podrían detectar las inconsistencias para abrir un debate al respecto y seguir construyendo conocimiento. Este tipo de prácticas buscan la transparencia.

Si has trabajado con grupos de investigación o en equipo, sabrás que la práctica de compartir datos no sólo es posible sino necesaria y, aunque no podemos obligar a nadie a que comparta datos, sí podemos argumentar razones a favor de hacerlo; una de las razones principales por las que cada vez más los investigadores en todo el mundo están adoptando esta práctica es porque al compartir sus datos se abre la posibilidad de ser más citados (¿quién no quiere verse citado?), con lo cual se incrementa la visibilidad y con ello hay nuevas oportunidades de colaboración, ya que se estaría fortaleciendo la veracidad y reproducibilidad de los resultados para seguir construyendo conocimiento. ¡Y de eso se trata la ciencia!

Cada conocimiento científico que se genera se construye sobre conocimiento previo, sobre literatura y datos anteriores; esa es la mecánica de la ciencia, que nos permite encabezar avances significativos en la comprensión y resolución de problemas complejos. Al utilizar datos que ya existen, pero que serán usados para responder las mismas o distintas preguntas de investigación, le damos un segundo valor a los datos; puedes profundizar, analizar e incluso, en algunos casos, ahorrar costos y evitar el volver a recopilar de nuevo los mismos datos que alguien más ya levantó.

Los datos que vas a reutilizar deben cumplir al menos las siguientes condiciones para que puedan ser de utilidad:

- Ofrecer la certeza de que sean de licencia libre: ¿la licencia con la que cuentan estos datos te da permiso para poder hacer lo que solicitas?
- Contar con los suficientes metadatos (autor/a, título, temas, identificadores) para reutilización, es decir ¿estos datos contienen la suficiente información adicional para que puedas entenderlos o aplicarlos a tu propia investigación?
- Garantizar que la fuente sea oficial, confiable, con altos estándares y que cumpla con políticas y regulaciones éticas que protejan los derechos.

¿Cómo citar los datos?

Ten presente siempre que la ética, la honestidad y el respeto son la base de la integridad académica. Por eso, cada vez que uses datos de otros debes citar la fuente. Esta práctica es clave porque permite, además de reconocer de dónde se tomaron, ubicar la precisión y verificar la autenticidad de esos datos y, por supuesto, darle el crédito a las y los autores originales sobre su propia obra.



Para citar datos, el manual *How to cite datasets and link to publications* menciona como elementos a tener en cuenta los siguientes:

Tabla 2. Elementos para citar los datos

Autor/Autora	¿Quién lo ha elaborado? ¿Quién creó ese conjunto de datos?
Fecha de publicación	Es la fecha en la que el procedimiento al que fueron sometidos los datos se completó y se puso a disposición.
Título	Nombre que se asignó al conjunto de datos.
Edición	Aquí se indica qué tan “crudos” o qué tan procesados o refinados se encuentran los datos.
Versión	El número de versión va aumentando cuando los datos cambian o son actualizados.
Características y URI	Esta información, más especializada, se incluye sólo si el conjunto de datos cuenta con ella.
Tipo de recurso	Se refiere al medio en el que se está presentando la información. Pueden ser: un conjunto de datos, base de datos, audio, video, entre otros.
Editor	La institución u organización que está resguardando los datos.
Huella numérica única (UNF)	Esta información es generada por un algoritmo. No aparece en todos los casos: si no existe no se registra en la cita.
Identificador	Es una etiqueta permanente con la que el repositorio identifica los datos de manera consistente (DOI: <i>Digital Object Identifier</i>).
Ubicación	Se trata de una URL permanente donde se encuentra alojado el conjunto de datos que estamos citando.

Fuente: elaboración propia con base en Ball y Duke (2015).

No es estrictamente necesario que existan todos estos elementos a la hora de citar un conjunto de datos, pero los que nunca deben faltar son: el autor/a, el título, la fecha, la ubicación, el identificador y la editorial.

Existen varios tipos de estilos en los que se pueden citar los datos digitales, pero para fines prácticos te recomendamos hacer tus citas en el formato de la Asociación Americana de Psicología (APA), que son de las normas más utilizadas.



Clave 21

Bibliotecas de datos de investigación (Repositorios)

Un Repositorio de datos es una biblioteca virtual donde los investigadores pueden guardar y compartir la información que han recopilado en sus estudios, la cual se organiza de manera ordenada y queda almacenada en un lugar seguro. Esto ayuda a que sea fácil de encontrar, usar y compartir con otras personas que están interesadas en el mismo tema.

Al igual que en una biblioteca física, los investigadores deben seguir ciertas reglas para depositar y acceder a los datos, como respetar los derechos de autoría, proteger la privacidad de las personas y seguir reglas éticas.

Estos repositorios de datos suelen contar con herramientas de búsqueda que permiten encontrar conjuntos de datos a través de palabras clave, ya que suelen incluir metadatos (etiquetas) que favorecen la localización de información adicional relevante sobre el origen de los datos. Los repositorios se fueron creando con base en las necesidades de los investigadores, es por esta razón que existen diferentes tipos, ya sea por disciplina (repositorios temáticos), de carácter más general (repositorios multidisciplinarios) o también los que han sido desarrollados por una organización en concreto (repositorios institucionales) y pueden estar abiertos al público o únicamente al servicio de la propia institución.

¿Quieres publicar tus datos en un repositorio?

¿Cuáles son las ventajas de hacerlo?

Usar repositorios de datos trae consigo diversas ventajas, entre otras:

- *Conservación*: como investigador podrás tener la certeza de que tus datos se preservan en un lugar seguro con base en políticas que te garantizan que tu aportación tendrá un buen uso.
- *Accesibilidad*: tu material y el de otros será fácilmente localizado en una plataforma confiable.
- *Reutilización*: otros colegas pueden compartir tu trabajo y tú podrás acceder al trabajo de otros investigadores. Todos podrán retomar sus datos (citándolos debidamente) y analizarlos desde la misma u otras metodologías.

¿Cómo elegir el repositorio correcto para ti?

Algunos puntos a tener en cuenta son:

- Requisitos de acceso y licencias de uso que se ajusten a las necesidades de acceso de cada disciplina.
- Debe ser intuitivo y fácil de manejar o bien contar con manuales, protocolos y guías o, en su defecto, contar con un soporte técnico que pueda atender sus dudas y necesidades.
- Sus términos y condiciones deben ser claros.
- Dentro de sus políticas ha de contar con garantías para la preservación y seguridad de los datos, así como también ofrecer un identificador persistente para cada conjunto de datos.
- Contar con una infraestructura sólida y segura que garantice que tus datos estarán seguros.
- Ser transparentes en sus procesos y políticas, aunado a una buena reputación, es decir, que otros colegas tengan buena experiencia.
- Saber de qué manera llevan a cabo los procesos de propiedad intelectual y derechos de autoría.

¿Quieres obtener datos de un repositorio para utilizarlos en tu investigación?

Si quieres o necesitas consultar los datos de otras investigaciones de manera libre tienes a tu disposición muchas opciones de repositorios. Los más utilizados, dado que cubren diversidad de temas, son: Re3data, Dataverse y Zenodo.

Re3data es una plataforma que ofrece información sobre otras, podría decirse que es un “repositorio de repositorios”. Proporciona detalles de fuentes de información de todo el mundo, incluyendo ubicación, temática, políticas de acceso, términos de uso y más. Para consultarla accede a: <https://www.re3data.org/>



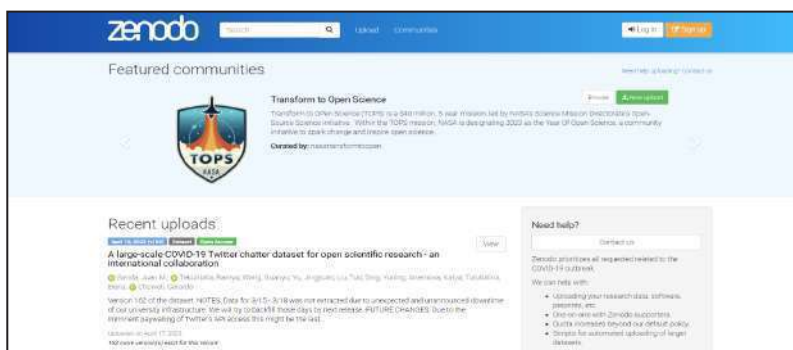
Fuente: Re3data.org

Dataverse es un repositorio dirigido a compartir, preservar, citar, explorar y analizar datos de investigación, facilitando que esta información esté disponible para otros. Investigadores, revistas, distribuidores de datos e instituciones afiliadas reciben crédito y visibilidad. <https://dataverse.org/>



Fuente: Dataverse.org

Y por último, Zenodo, que también es multidisciplinario, incluye tanto conjuntos de datos como publicaciones científicas y su cobertura contempla ciencias básicas, ciencias sociales, naturales y humanidades. Disponible en: <https://zenodo.org/>



Fuente: Zenodo.org

A continuación, te presentamos una tabla con la cual puedes guiarte para conocer otros repositorios de datos más especializados.

Tabla 3. Repositorios de datos especializados

Nombre	Descripción	Página
Gene Expression Omnibus	Este es un repositorio especializado en salud, su temática principal es la genética. Ellos mismos se definen como un repositorio de acceso público que almacena datos procedentes de la genómica funcional. Ofrece herramientas de consulta para descargar experimentos.	https://www.ncbi.nlm.nih.gov/geo/

Nombre	Descripción	Página
Qualitative Data Repository	Este repositorio contiene sólo datos cualitativos de Ciencias Sociales, es financiado por la <i>National Science Foundation</i> y hospedado en el <i>Center for Qualitative and Multi-Method Inquiry</i> , de la <i>Syracuse University</i> (Estados Unidos). Se centra en el almacenamiento y uso compartido de datos y metadatos, generados mediante la metodología cualitativa, tales como entrevistas semiestructuradas o no estructuradas, grupos focales o notas de campo, también ofrece herramientas de búsqueda.	https://qdr.syr.edu/about
Open Topography	Este repositorio es financiado por la <i>National Science Foundation Earth Sciences</i> y nace de un proyecto de ciberinfraestructura en Ciencias de la Tierra. Su función es liberar el acceso online a los datos topográficos en alta resolución orientados en temáticas de Ciencias de la Tierra, y que se recaudan mediante LIDAR u otras tecnologías. Dispone de varias herramientas, así como varios datos que se pueden compartir.	https://opentopography.org/
The European Space Agency	En este repositorio se pueden encontrar imágenes y datos procedentes de las diferentes misiones espaciales de la Agencia Espacial Europea (por su acrónimo en inglés, ESA).	https://www.esa.int/

Nombre	Descripción	Página
Open Data Cern	Este repositorio se especializa en datos de física de partículas, creado por el Centro Europeo para la Investigación Nuclear (CERN), el gran laboratorio europeo para la investigación nuclear.	https://opendata.cern.ch/
Kaggle.com	Este repositorio no se especializa en ningún conjunto de datos en particular, sino que contiene datos de diversas colecciones, desde la altura de las olas en las costas de Australia hasta venta de propiedades en Melbourne. Cabe resaltar que esta página paga por conjuntos de datos, y pone regularmente desafíos bien remunerados.	https://www.kaggle.com/
Datos.gob	Este repositorio de datos abiertos fue creado por el gobierno de México y contiene diversos conjuntos de datos oficiales en economía, salud y otros.	https://datos.gob.mx/

Fuente: elaboración propia.

Referencias

Ball, A. y Duke, M. (2015). How to cite datasets and link to publications. *DCC how-to guides*. <https://www.dcc.ac.uk/guidance/how-guides/cite-datasets#sec:elements>

Biblioteca del Congreso Nacional de Chile (s.f.) *Las cinco estrellas de los datos abiertos*. <https://datos.bcn.cl/es/informacion/las-5-estrellas>

Briney, K. (2015). *Data Management for Researchers: Organize maintain and share your data for research success*. Pelagic Publishing [edición Kindle].

García, J. (s.f.). *Los seis tipos de licencias Creative Commons*. <https://bit.ly/3L7isVl>

¿Qué aprendiste?

1. Define con tus propias palabras lo que se entiende por Datos de investigación.
2. ¿Cuál es la diferencia entre *Big Data* y *Right Data*?
3. Menciona 3 tipos de datos cualitativos y 3 cuantitativos.
4. ¿Para qué sirve y cómo se usa una Base de conocimiento?
5. ¿Qué sinónimos podrías usar para “Base de conocimiento”?
6. Explica la utilidad de un Archivo *README*.
7. ¿Cuáles son las ventajas de usar un Cuaderno de investigación?
8. Enlista los metadatos de la última imagen que guardaste en tu teléfono móvil.
9. Menciona algunos de los beneficios de usar Libros de código.
10. ¿Qué es un Plan de gestión de datos?
11. ¿Qué elementos contiene un Plan de gestión de datos?
12. ¿Cuál es la diferencia entre un Plan de gestión de datos y un Plan de análisis de datos?
13. Enlista los elementos que integran el ciclo de vida de los datos científicos.
14. ¿Qué es y cómo buscar *Open Data*?
15. ¿Qué es el formato libre (abierto)?
16. Explica qué es un formato cerrado.
17. Menciona los tipos más usuales de identificadores persistentes.
18. Explica la función de los repositorios y cómo se clasifican.
19. Los datos de investigación pueden encontrarse en varios formatos: menciona cinco.
20. ¿Cuál es el valor de la ética en el manejo de datos de investigación?
21. Da un ejemplo de una mala práctica en el nombrado/ versionado de archivos e indica cómo corregirla.

Bibliografía complementaria

Bhimrao Gajbe, S., A. Tiwari Gopalji y R. KumarSingh. (2021). *Evaluation and Analysis of Data Management Plan tools: A Parametric Approach*. <https://doi.org/10.1016/j.ipm.2020.102480>

Burnette, M. Williams, S. y Imker, H. (2016). Del plan a la acción: implementación exitosa del plan de gestión de datos en un proyecto multidisciplinario. *Revista de Biblioteconomía de Ciencias*, 5(1). <https://doi.org/10.7191/jeslib.2016.1101>

Cauchick-Miguel P.A., M. Ribera, S. R. Moro, y M. Amorim. (2020). *Data Management Plan in Research: Characteristics and Development*. https://doi.org/10.1007/978-3-030-50072-6_1

Corti L. Bishop L. Woollard M. y Van den Eynden V. (2014). *Managing and Sharing Research Data*. University of Essex.

Federer, L.M. y J. Qin. (2019). *Beyond the Data Management Plan: Expanding Roles for Librarians in Data Science and Open Science*. <https://doi.org/10.1002/pra2.82>

Geneva Graduate Institute. (2023, 14 de abril). *Research Data Management*. <https://libguides.graduateinstitute.ch/rdm/folders>

Guest, G., Namey, E. y Mitchell, M. (2013). *Collecting Qualitative Data: A Field Manual for Applied Research*. Publicaciones SAGE. Edición de Kindle.

Harvard University. (2023, 25 enero). *Data Management Plans*. <https://datamanagement.hms.harvard.edu/plan-design/data-management-plans>

Hey, T., Tansley, S. y Tolle, K. (Eds.). (2014). *El cuarto paradigma*. Universidad Autónoma Metropolitana. <https://casadelibrosabiertos.uam.mx/gpd-cuarto-paradigma-el.html>

LEARN Project. (2017). *Gestionar los activos del conocimiento para la investigación y la innovación en el siglo XXI: resumen ejecutivo*. LEARN - Comisión Europea. <http://learn-rdm.eu/wp-content/uploads/LEARN-Executive-Briefing-v2-Spanish.pdf>

Mayer-Schönberger, V. y Cukier, K. (2013). *Big Data. La revolución de los datos masivos*. Turner. <http://catedradatos.com.ar/media/3.-Big-data.-La-revolucion-de-los-datos-masivos-Noema-Spanish-Edition-Viktor-Mayer-Schonberger-Kenneth-Cukier.pdf>

Penn State. (2023, 1 de enero). *What is Metadata?* <https://www.e-education.psu.edu/dmpt/node/660>

Smarty. (2023, 23 abril). *Data Management: When Good Data Goes Bad*. <https://www.smarty.com/articles/data-management>

Toro, M. y Laniado, H. (2019). *Big data: historia, definición, herramientas y aplicaciones en la Industria*. *VirtualPro*, (2024), 2-5. www.revistavirtualpro.com.

UK Data Service. (2023, 1 de marzo). *Organising*. <https://ukdataservice.ac.uk/learning-hub/research-data-management/format-your-data/organising/>

21 claves para organizar los datos de investigación. Guía práctica de gestión de datos, de Lourdes Feria Basurto, fue editado en la Dirección General de Publicaciones de la Universidad de Colima, avenida Universidad 333, Colima, Colima, México, <http://www.ucol.mx>. La edición se terminó en noviembre de 2024. Se utilizó la familia Fira Sans en la composición tipográfica. Programa Editorial No Periódico: Eréndira Cortés Ventura. Gestión administrativa: María Inés Sandoval Venegas. Cuidado de la edición: Juan Manuel Ramírez Vélez y Leticia Bermúdez Aceves. Diseño de portada: Lizeth Maricruz Vázquez Viera.



Las 21 claves de este libro están concebidas para que las y los lectores aprendan a aprovechar algunas de las mejores estrategias para el manejo adecuado de todo tipo de datos en favor de su investigación. Al revisar esta guía conocerán los componentes básicos para organizar sus archivos, datos crudos, datos procesados, lecturas digitales, diversos componentes de *software*, bitácoras y otros elementos desde la primera investigación que realice.

Egresada de la carrera de Bibliotecología en la UNAM, Lourdes Feria Basurto es doctora en Ciencias de la Documentación por la Universidad Complutense de Madrid y cuenta con una estancia posdoctoral en "Gestión de datos científicos" en el Instituto de Investigaciones Bibliotecológicas (UNAM). Fue Coordinadora General de Tecnologías de Información de la Universidad de Colima y coautora del *software* SIABUC utilizado en miles de bibliotecas en Latinoamérica. Destaca su participación en proyectos en colaboración con la UNESCO, Cámara de Diputados, Embajada de México en Estados Unidos, sistema de bibliotecas públicas de la SEP, entre otros, así como en diversas instituciones educativas de México y el mundo. Actualmente es una reconocida formadora y consultora en tecnologías de información, gestión de conocimiento y proyectos de investigación.

ISBN: 978-607-8984-50-3



9 786078 984503

ISBN: 978-607-8961-89-4



9 786078 961894



UNIVERSIDAD
DE COLIMA



PUERTABIERTA
EDITORES