

DATOS ABIERTOS DE INVESTIGACIÓN

Qué son los datos de investigación

Los datos son la materia prima de la investigación académica para el avance del conocimiento.

En los últimos años hemos sido testigos de cambios profundos en el proceso de investigación. Fenómenos como La Revolución de los Datos, los Datos Abiertos y la Ciencia Abierta nos llevan a replantear los procesos de gestión y publicación en la UNAM.

Los datos de investigación se recopilan, observan o crean para ser analizarlos y producir resultados originales. Estos datos se producen y son validados en las distintas áreas académicas, como las humanidades, las artes y la ciencia. Pueden ser cuantitativos o cualitativos, se resguardan en diversos formatos y se originan bajo diferentes metodologías de trabajo:

- observaciones
- experimentos
- simulaciones
- compilaciones (derivados)

Según su nivel de procesamiento, los datos de investigación pueden ser:

- ▶ **Primarios (sin procesar):** se colectaron pero no han sido procesados o analizados. Por ejemplo: observaciones, notas de campo o datos de experimentos.
- ▶ **Procesados:** han sido digitalizados, traducidos, transcritos, limpiados, validados, verificados y/o anonimizados.
- ▶ **Analizados:** modelos, gráficos, tablas, textos u otros que se procesaron y se pueden utilizar para presentar información útil o conclusiones y para tomar decisiones.

Por qué es importante abrir los datos de investigación

Los datos de investigación son un componente fundamental para la toma de decisiones y para hacer frente a problemas complejos globales como la pobreza, el cambio climático, la pérdida de diversidad genética o la acidificación de los océanos.

Para lograr un uso más efectivo de los datos, éstos deben ser abiertos y, por lo tanto, cumplir con los nuevos requerimientos de estandarización, limpieza y verificación con el fin de que se integren con diferentes proyectos y disciplinas. De esta manera, se promueve

- su uso y reutilización
- la generación de nuevo conocimiento

El objetivo de esta nueva forma de utilizar los datos es su integración y publicación en línea. En este sentido, los datos abiertos son una herramienta eficiente para compartir datos de investigación.

Como datos abiertos, los datos de investigación nos permiten:

- ▶ **Verificar** los resultados de una investigación en menor tiempo.
- ▶ **Cruzar datos** de otros proyectos de manera más sencilla y automatizada.
- ▶ **Fortalecer** la colaboración entre disciplinas.
- ▶ **Contribuir** a la toma de decisiones para resolver problemas globales.

Cómo se abren los datos de investigación

El potencial de los datos de investigación aumenta exponencialmente cuando se utilizan como datos abiertos. Mejor conocidos por su término en inglés, Open Data, los datos abiertos son digitales, de carácter público, que no tienen naturaleza reservada o confidencial; se encuentran en línea para ser utilizados, reutilizados y redistribuidos libremente por cualquier interesado siempre que se cite la fuente de origen. Su uso permite:

-  Descargarlos libremente de internet en formatos accesibles, estructurados y modificables.
-  Reutilizarlos y redistribuirlos bajo términos y condiciones que permitan su uso, reúso e integración con otros conjuntos de datos.

En la UNAM, los datos abiertos de investigación provienen de tres fuentes principales:

a) Acervos: fuentes de consulta, referencia o preservación, compuestas principalmente de datos primarios; por ejemplo, colecciones biológicas (Herbario Nacional, Paleontológica Nacional y Aves); fotográficas (Archivo Fotográfico Manuel Toussaint); mapas y capas geoespaciales (acervos de mapas históricos), entre otras.

b) Proyectos de investigación académica: datos generados en el desarrollo de proyectos de investigación (como experimentos o análisis estadístico) que sustentan los resultados de las publicaciones.

c) Publicaciones y repositorios: son datos que describen y clasifican los productos académicos terminados, como documentos, imágenes y otros objetos multimedia.

A través de la Coordinación de Colecciones Universitarias Digitales, éstos datos se integran y publican para su consulta en el *Portal de Datos Abiertos UNAM, Colecciones Universitarias* (datosabiertos.unam.mx).

Características de los Datos Abiertos

Sólo una parte de los datos generados por la investigación se pueden considerar abiertos. Para ello, deben cumplir con las siguientes características:

Accesibles: disponibles para la gama más amplia de usuarios, para cualquier propósito.

De libre uso: citan la fuente de origen como único requerimiento para ser utilizados libremente.

Gratuitos: se obtienen sin entregar a cambio contraprestación alguna.

Integrales: contienen el tema que describe a detalle y con los metadatos necesarios.

Legibles por máquinas: estar estructurados, total o parcialmente, para ser procesados e interpretados por equipos electrónicos de manera automática.

No discriminatorios: disponibles para cualquier persona, sin necesidad de registro.

Oportunos: actualizados periódicamente conforme se generan.

Permanentes: se conservan en el tiempo; las versiones históricas relevantes para uso público se mantendrán disponibles con identificadores adecuados al efecto.

Primarios: provienen de la fuente de origen con el máximo nivel de desagregación posible.

Los datos abiertos idealmente deben presentarse de manera estructurada y ser descargables en formatos libres de restricciones. El siguiente esquema nos permite identificar diferentes alternativas de publicación:

CALIFICACIÓN DE LOS DATOS ABIERTOS:		
★	No estructurados (por ejemplo, PDF)	
★★	Estructurados (por ejemplo, archivo Excel)	
★★★	Estructurados y en formatos no propietarios (como texto separado por comas "CSV")	
★★★★	Estructurados, en formatos no propietarios y con una URL única y estática	
★★★★★	Estructurados, en formatos no propietarios, con URL única y estática e interconectados (Linked Data) con otras bases de datos.	

ccud.unam.mx

Universidad Nacional Autónoma de México Coordinación de Colecciones Universitarias Digitales

Directorio

Dra. Tila María Pérez Ortiz
Coordinación General

Biól. Joaquín Giménez Héau
Coordinación Técnica

M. en C. Ariana Chávez Méndez
Líder de Proyecto

Lic. en C.C. Rubén Sáenz González
Desarrollo Tecnológico

Biól. Daniel Pérez Castillo
Informática y Estadística

Lic. Edurne Uriarte Santillán
Inventario de Colecciones y Comunicación

Informes:

Inventario de Colecciones y Comunicación: 5623 7723

Conmutador: 5622 8971

Correo: contacto@ccud.unam.mx



ccud.unam.mx