

ADMINISTRACIÓN DE DATOS ABIERTOS EN RIMISP

Manual de manejo
de datos

Septiembre, 2022



1. ¿QUÉ SON LOS DATOS ABIERTOS?

Durante el desarrollo de proyectos de investigación se realizan encuestas, se generan indicadores y se sistematiza información utilizando bases de datos relevantes y muchas veces innovadoras. Parte del quehacer investigativo es que dichos datos cumplan con principios éticos y universales. Estos principios implican que los datos sean: Encontrables, Accesibles, Interoperables y Reusables¹. Esto conlleva que los datos de una investigación sean abiertos. En este sentido:

“Los datos abiertos son datos que pueden ser utilizados, reutilizados y redistribuidos libremente por cualquier persona, y que se encuentran sujetos, cuando más, al requerimiento de atribución y de compartirse de la misma manera en que aparecen” (Open Data Handbook, 2022).

Que los datos sean abiertos significa entonces que éstos estén dispuestos en un repositorio o sitio web de libre y fácil acceso para su uso.

2. IMPORTANCIA DE LOS DATOS ABIERTOS

El mundo actual presenta grandes desafíos en cuanto a la generación y manejo de información. Con el avance de la conexión al internet es cada vez es más fácil acceder a todo tipo de datos. Sin embargo, aún existen grandes limitaciones con respecto al acceso a bases de datos cuantitativas, el desconocimiento de sitios en donde buscar información e incluso la carencia de datos de temas poco tratados o de territorios aislados. Esto sucede particularmente en el mundo rural en donde hay mayor dificultad de acceso y, por lo tanto, es muy costoso recopilar información.

En este sentido, la disponibilidad de datos abiertos permea dichas limitaciones a través de la generación de conocimiento y políticas basadas en la evidencia, así como la forma innovadora de promover la transparencia y la contabilidad de información en un mundo en donde distintas disciplinas se entrelazan para hacer investigación.

La apertura de los datos es importante, entonces, porque permite la reproducibilidad de las investigaciones, la difusión de información científica y la aplicación de buenas prácticas éticas y de transparencia. Que los datos sean abiertos se traduce en que éstos permitan que una mayor cantidad de gente, como investigadores o tomadores de decisión, puedan tomar mejores o disponer de material para reflexionar sobre nuevas ideas. Además, la interoperabilidad de los datos abiertos facilita que la información pueda ser integrada con otros datos, aportando a un diálogo multidisciplinar.

Por otro lado, seguir protocolos de administración de datos promueve sistemas de orden y organización para la presentación y estandarización de información en los procesos de investigación. Hoy en día muchas organizaciones promueven e incluso exigen disponer de las bases de datos de las investigaciones en repositorios o sitios web, asegurando así que la

¹ Por sus siglas en inglés, esto significa que los datos cumplen los principios de “Data FAIR”.

información cumpla con principios de ética y transparencia.

Dentro de éstas instituciones se encuentran, por ejemplo, Universidades, revistas académicas y fondos de investigación. Algunas de ellas se detallan en la figura 1.

Figura 1. Organizaciones que promueven principios éticos y transparentes de disponibilidad de datos (FAIR principles)



Fuente: Nature Research Academies (2019)

Finalmente, pero no menos importante, la apertura de datos contribuye al desarrollo y visibilización de los territorios, especialmente los rurales que no disponen de mucha información. En la medida en la que se cuenta con más y mejores datos, también se aporta a la discusión y el diseño de mejores políticas públicas, siendo este uno de los principales aportes e intereses de RIMISP como centro de investigación y desarrollo.

3. ¿CÓMO SE TRABAJAN LOS DATOS PARA QUE SEAN DE LIBRE DISPOSICIÓN?

Parte de los lineamientos que se siguen en las publicaciones de datos es que se respete la identidad de los encuestados, mediante la anonimización de los mismos y la estandarización de los formatos para el fácil uso por parte de terceros. En este sentido, algunos de los pasos que se siguen son:

1. Se anonimiza el nombre del encuestado/a, si tiene nombre
2. Si la encuesta es a nivel de hogares o unidades productivas, se recodifican con identificadores numéricos
3. Se resguarda la información de los encuestadores
4. Se elimina la georeferenciación, si es que la hubiera
5. Se realiza una limpieza de la base de datos para evitar valores atípicos o que no se encuentren en el rango de las respuestas posibles
6. Se recodifican los casos perdidos
7. Se recodifican las variables para que sean legibles

El objetivo de estas directrices es que se disponga de la información comprensible y ordenada para cualquier usuario pero resguardando la identificación de las personas encuestadas.

4. RIMISP DISPONE SUS DATOS AL PÚBLICO

Parte del compromiso de RIMISP con la comunidad científica es disponer información reproducible y de fácil acceso, por ello dispone de sus datos libremente dentro del repositorio de Harvard Dataverse en el portal creado para Rimisp: <https://dataverse.harvard.edu/dataverse/rimisp>

En dicho espacio se encuentran bases de datos de los diferentes proyectos de investigación desde el año 2020. Cada publicación incluye:

1. Base de datos: información de encuestas recopiladas dispuesta en formatos .csv,.xlsx (Excel), .dta (Stata) o .Rdata (R)
2. Cuestionario: documento en formato pdf en donde se indican las preguntas que se hicieron a los entrevistados.
3. Libro de códigos: documento en formato pdf en donde se detallan las variables, etiquetas y preguntas del cuestionario.
4. Ficha técnica o manual metodológico: documento en formato pdf que contiene la información general y técnica del diseño, muestreo y limpieza de la base de datos.

Adicionalmente, el portal conecta con un investigador/a de RIMISP que puede contestar dudas o consultas acerca del uso de la información disponible.

5. REFERENCIAS

Open Data Handbook (2022). “*What is Open data?*”. Disponible en: <https://opendatahandbook.org/guide/es/what-is-open-data/>

Nature Research Academies (2019). “*The Principles and Practices of Data Sharing Webinar series 2019*”.



www.rimisp.org

.....
Chile · Ecuador · Colombia · Centroamérica · México

