

# ¿Dónde encontrar datos?

Encontrar datos puede no ser una tarea fácil. El grado de dificultad varía mucho según el tema, la geografía y otras dimensiones. El siguiente es un documento que puede ayudar a facilitar este proceso. Proporciono algunos enlaces a lugares donde puede encontrar conjuntos de datos y luego proporciono algunos consejos sobre otras formas de encontrar datos fuera de la búsqueda en la web.

## Motores de búsqueda de datos

Los siguientes sitios web pueden considerarse como "Motores de búsqueda de datos"; es decir, ingresas una consulta y el sitio web te mostrará un conjunto de opciones de posibles datos disponibles (ya sea de forma pública o privada) relacionados con esa consulta.

En algunos casos, los datos no están disponibles para el público. Sin embargo, ¡no te desanimes! A veces, las barreras de entrada son tan pequeñas como enviar un mail, mientras que en otras situaciones, deberán completar un formulario de solicitud de datos (esto es más complejo por tema tiempos). No tema preguntarle a una organización o institución cómo puede acceder a sus datos.

- [ICPSR](#) - Un servicio del Instituto de Investigaciones Sociales de la Universidad de Michigan. Este motor de búsqueda de datos es un gran recurso y le proporcionará muchas buenas opciones. Si bien puede encontrar información sobre otros países, esta fuente se inclina mucho hacia los Estados Unidos. Las consultas pueden ser del tipo "encarcelamiento" o "salud mental". Una vez que haya encontrado un conjunto de datos que desee descargar, deberá crear una cuenta y tal vez completar un formulario sobre por qué necesita los datos. Piense en esto como una inversión única en lugar de un costo por datos.
- [ResearchDataGov](#) - En 2019, ICPSR recibió un contrato de la Oficina del Censo para crear [ResearchDataGov.org](#), este sigue siendo un sitio web de prueba de concepto. Entonces, RDG contiene solo datos de acceso restringido, y solo de aproximadamente 5 agencias. Están en el proceso de construir el portal real que involucrará datos restringidos de 17 agencias / programas federales. (¡Gracias Lynette Hoelter!)
- [Data.gov](#) - es el lugar donde se supone que todas las agencias gubernamentales (en cualquier nivel, desde municipal hasta federal) deben compartir sus datos de uso público, donde como [researchdata.gov](#) si se trata de datos de acceso restringido. (h / t a [@willwheels](#) y Lynette Hoelter!)
- [World Bank Microdata Library](#) - un motor de búsqueda de datos alojado por el banco mundial
- [Dataverse](#) - Esta página es un depósito de datos de artículos que ya existen. En muchos casos, si está interesado en las fuentes de datos utilizadas por un documento específico, Dataverse puede proporcionarle esta información (así como archivos do-files para esos datos). Al igual que ICPSR, puede escribir una palabra de búsqueda y probar suerte. (Debe tener en cuenta que la propia revista puede haber pedido a los autores que proporcionen un kit de replicación, que estaría disponible en la página en línea del artículo o en la página del autor).
- [Quand](#) Este otro motor de búsqueda de datos, se centra más en datos relacionados con las finanzas.
- [FRED](#) Un servicio de la Fed de St. Louis. Este es un buen sitio web para encontrar datos relacionados con variables macroeconómicas (desempleo, tasas de interés, índices de desigualdad, etc.). Si bien tiene información sobre otros países para ciertas variables, se centra principalmente en Estados Unidos.

- [Google Public Data](#) - Este es un gran motor de búsqueda de datos para datos sobre indicadores agregados (piense en el IDH, el PIB per cápita, etc.) a nivel de país y a lo largo del tiempo. Pero es una excelente manera de ver lo que está disponible en un país determinado. La mayoría de los datos provienen de organizaciones como la ONU, el Banco Mundial, etc. Este también es otro gran recurso para explorar tendencias y crear gráficos interesantes.
- [OECD Data](#) - Similar a los datos públicos de Google, pero contiene más indicadores "económicos" y tiene un conjunto más amplio de temas. En su mayoría países de la OCDE. (HT / [@biancafrogner](#))
- [Health Data.Gov](#) - este sitio está dedicado a hacer que los datos se puedan descubrir y hacer que los datos gubernamentales valiosos estén disponibles para el público con la esperanza de obtener mejores resultados de salud para todos. En este sitio, puede encontrar datos sobre una amplia gama de temas, que incluyen salud ambiental, dispositivos médicos, Medicare y Medicaid, servicios sociales, salud comunitaria, salud mental y abuso de sustancias. Los datos se recopilan y suministran de agencias del Departamento de Salud y Servicios Humanos de EE. UU., Así como de socios estatales. Esto incluye los Centros de Servicios de Medicare y Medicaid, los Centros para el Control y la Prevención de Enfermedades, la Administración de Alimentos y Medicamentos y la Agencia para la Investigación y la Calidad de la Atención Médica, entre otros. Vea la lista completa de agencias que aportan datos a HealthData.gov. (HT / [@biancafrogner](#))
- [Medicare Data](#) - similar a la página anterior, pero centrado en Medicare. (HT / [@biancafrogner](#))
- [NYC OpenData](#) - (Comentario de Grant R. McDermott): Su misión es "hacer que la gran cantidad de datos públicos generados por varias agencias de la ciudad de Nueva York y otras organizaciones de la ciudad estén disponibles para uso público". Puede obtener datos sobre todo, desde datos de arrestos hasta la ubicación de puntos de acceso wi-fi, ofertas de trabajo en la ciudad, recuentos de población sin hogar, licencias de perros y un directorio de baños en parques públicos. (HT [@grant\\_mcdermott](#)). Grant ha escrito algoritmos para rasparlos y puedes encontrar [ellos en este enlace](#)).
- [JPAL Catalog of Admin Datasets](#) - este es un recurso realmente interesante de JPAL, documenta los procedimientos sobre cómo acceder a 35 (y contando) diferentes conjuntos de datos de administración. Los datos administrativos (admin) son muy útiles cuando se realizan experimentos, y han estado recopilando una base de búsqueda. Por ejemplo, ¿busca resultados de salud mental relacionados con la delincuencia? Lo llevarán a los datos de administración de NY o IL. (@JPAL\_NA)

## Lista de datos

Los siguientes son sitios web que tienen *una lista* de conjuntos de datos. El beneficio de la mayoría de los datos que se encuentran en estos sitios es que los datos son relativamente "más limpios" que los que se encuentran en los motores de búsqueda.

- [Census](#) - este es un PDF de la mayoría de los conjuntos de datos que se trabajan bajo el paraguas del censo. Algunos serán públicos y otros tendrán acceso restringido, pero deberían darle una idea de lo que tienen.
- [NBER](#) - estos son datos proporcionados por la Oficina Nacional de Investigación Económica. Cubre una amplia gama de temas: macro, industria y productividad, comercio internacional, encuestas de hogares, datos de atención médica, demografía,

patentes y otros. Use Control (Cmd) + F para buscar términos generales, en oposición a variables específicas.

- [IPUMS](#) - Esta fantástica institución ofrece alrededor de 10 encuestas (en su mayoría estadounidenses) que se han limpiado y armonizado. Los datos pueden ser difíciles de usar al principio, pero una vez que los dominas, son muy útiles (especialmente cuando intentas obtener una estimación rápida de una variable). Tiendo a usar con frecuencia el IPUMS-CPS (que es la versión limpia de la Encuesta de población actual) y el IHIS (que es la versión limpia de la Encuesta nacional de entrevistas de salud). Tenga en cuenta que no *todas las* variables que se encuentran en el conjunto de datos original están en los datos de IPUMS. Por lo tanto, si no encuentra una variable que esté buscando en IPUMS para un año específico, no significa que no exista en los datos originales.
- [Data is plural](#) es una lista de correo que le enviará un boletín con algunos conjuntos de datos extraños. ¡Suscribir! Pero también encuentre datos "extraños" sobre este [repositorio](#). Recomendando buscar en este repositorio datos que pueden ser más oscuros de encontrar, tienen una serie de temas, por lo que una búsqueda de control + F será de gran ayuda.
- [CDC Wonder](#) - Este sitio de los CDC tiene una lista de conjuntos de datos que patrocinan. Se trata principalmente de resultados de salud.
- [CDC Health Data \(Others\)](#) - esta es otra lista de otros conjuntos de datos relacionados con la salud.
- [BEA](#) - esta es otra lista de otros conjuntos de datos relacionados con la salud.
- [AEA Data List](#) - Esta es una lista de datos de la Asociación Económica Estadounidense.
- [Historical Statistics](#) - los datos históricos pueden ser difíciles de encontrar, pero este sitio web proporciona un excelente primer paso.



## Data específica de salud

- [Health Policy Tracking](#) - esta no es una lista de conjuntos de datos en sí. Ésta es una base de datos en sí misma de cambios en las políticas por estado. No es exhaustivo, por lo que puede haber un cambio de política que no esté registrado en esta base de datos, pero es un buen lugar para comenzar.
- [Health Policy Research Dataset](#) - este es un conjunto de datos muy útil que tiene mucha información sobre los cambios de políticas de los estados y el año en que ocurrieron. Proporciona una gran cantidad de "controles" rápidos a nivel de año estatal.
- [Area Health Resources Files](#) - El Area Health Resources Files (AHRF) incluyen datos sobre profesiones de atención médica, establecimientos de salud, características de la población, economía, capacitación de profesionales de la salud, utilización del hospital, gastos hospitalarios y medio ambiente a nivel del condado, estatal y nacional, desde más de 50 fuentes de datos y a lo largo del tiempo (2014-2016 disponible públicamente). ([@biancafrogerdesarrollo](#))



## Data específica de Desarrollo

Estas son encuestas que contienen información sobre varios países, recomendadas por [@DaveEvansPhD](#).

- [Demographic and Health Surveys \(DHS\)](#) - el programa ha recopilado, analizado y difundido datos precisos y representativos sobre población, salud, VIH y nutrición a través de más de 400 encuestas en más de 90 países.
- [Living Standards Measurement Study](#) - El estudio de medición de niveles de vida - Encuestas integradas sobre agricultura (LSMS-ISA) es un proyecto de encuesta de hogares

establecido con una subvención de la Fundación Bill y Melinda Gates e implementado por el equipo de LSMS.

- [Violence against Children Surveys \(VACS\)](#) - en la actualidad, más de 24 países de África, Asia-Pacífico, América Latina y el Caribe participan activamente en un trabajo fundamental para prevenir la violencia contra los niños y los jóvenes. A continuación se incluyen informes de países, cuestionarios de encuestas y datos de apoyo para VACS en cada país. Los conjuntos de datos de VACS para muchos países están disponibles para uso público, de acuerdo con los acuerdos con los países socios. El acceso al conjunto de datos de uso público, ícono externo, se coordina a través de nuestra asociación con Together for Girls. Together for Girls Resource Bankexternal también incluye los informes y herramientas más recientes de la asociación Together for Girls.

## Uruguay

- [INE](#) (pestaña Base de Datos, si es ECH, ¡consúltenme por una versión más amigable!)
- [BCU](#) (pestaña Estadísticas y estudios)
- [EFHU](#) – Encuesta Financiera de los Hogares Uruguayos (yo les puedo facilitar el acceso)

## Otros

- [International Ipums](#) - IPUMS-International se dedica a recopilar y distribuir datos censales de todo el mundo. Los objetivos del proyecto son recopilar y preservar datos y documentación, armonizar los datos y difundir los datos armonizados de forma gratuita. Este conjunto de datos en particular contiene datos de censos internacionales armonizados para las ciencias sociales y la investigación en salud. (@aiyaranaka)
- [Google BigQuery](#) – Sugerido por Grant Mcdermott - aún no está claro cómo usarlo, pero es posible que tenga más para [explicar aquí](#).
- [JEP Data watch Articles](#) – el Journal of Economic Perspectives (JEP) antes de 2000, tenía artículos en los que los autores resumían un conjunto de datos en particular que consideraban útil. Estos se denominaron “Data Watch”, explicaba con más información qué datos tenían. Entonces puede comenzar aquí o si desea saber más sobre la Encuesta Estadounidense sobre el Uso del Tiempo (ATUS), puede ir al artículo en particular que habla de ella. (@bartonwillage)
- [Ian Mccarthy's website](#) tiene un diseño agradable para proporcionar información sobre este sitio web y otros métodos.



## Tips:

- Si encuentra informes o artículos en los que ve una estadística o un “número” que necesita, busque en la nota al pie o en la referencia del documento para ver de dónde obtuvieron los datos los autores. En algún momento, es posible que deba ponerse en contacto o enviar un mail a alguien. Asegúrense de ser educado y utilizar su criterio personal (es decir, ¡no bombardear constantemente a una persona con muchos correos electrónicos!) Al redactar correos electrónicos. Algunos de esos documentos o informes pueden tener un archivo de réplica que sea de acceso público.
- Pregunte por fuentes de datos. Si tiene personas en su red que son expertas en un tema, pídale su consejo.
- Sean creativo. Si no puede encontrar la variable que está buscando, ¿cuál es otra variable que sería un buen proxy? ¿Puedes encontrar eso?

- Valide su fuente de datos. Si encuentra datos que le gustaron, vea si alguien más los ha utilizado y con qué propósito.