

FS 2.1: 2023 Ampliación del conjunto de datos de la encuesta de hogares de OLAS

La Ampliación del Conjunto de Datos de Encuestas de Hogares de OLAS fue un proyecto implementado en colaboración entre OLAS y la División Social (SCL) del Banco Interamericano de Desarrollo (BID) entre diciembre de 2022 y julio de 2023. Los objetivos del proyecto eran múltiples, con un enfoque en mejorar la información disponible sobre el acceso al agua y al saneamiento y, al mismo tiempo, agilizar el proceso de generación de datos.

Antecedentes

Antes de este proyecto, había dos conjuntos de datos regionales producidos a partir de encuestas de hogares nacionales: el Conjunto de datos de encuestas socioeconómicas de hogares de SCL y el Conjunto de datos de encuestas de hogares de OLAS.

El conjunto de datos de encuestas socioeconómicas de hogares de la SCL incluye información sobre temas como educación, trabajo, ingresos, educación y condiciones de vida; sin embargo, antes de este proyecto el conjunto de datos incorporaba pocos indicadores sobre agua y saneamiento, la mayoría de los cuales utilizaban definiciones y términos obsoletos. Por el contrario, el conjunto de datos de la encuesta de hogares de OLAS incluía muchos indicadores valiosos con definiciones relevantes para el sector, pero carecía de muchos de los beneficios del conjunto de datos de la encuesta de hogares de SCL, como su gran cobertura temporal, variedad de dimensiones de desglose y recursos y procesos de datos establecidos para la actualización. El conjunto de datos. La fusión de los conjuntos de datos ofreció beneficios claros y eliminó la duplicidad.

Tabla 1: Características de ambos conjuntos de datos, que muestran la ventaja de fusionar los proyectos. (Elaboración propia del auto las autoras)

Conjunto de datos del OLAS ENAHO 2020		Conjunto de datos Indicadores SCL:	
✓	Basado en los datos propios de los países	✓	Basado en los datos propios de los países
■	Actualizaciones tomaban mucho tiempo por falta de infraestructura existente	✓	Proceso e infraestructura de actualización ya existente
■	No permitía análisis temporal (1 o 2 años por país, con años de 2014-2020)	✓	Permite análisis temporal (promedio de más de 13 años por país, con años de 2003-2022)
■	2 dimensiones de desglose (zona y quintil de ingresos)	✓	8 dimensiones de desglose (zona, quintil, sexo, edad, etnicidad, estatus migratorio, estatus de discapacidad, y nivel educativo)
✓	32 de indicadores de AyS	■	4 indicadores de AyS
✓	Indicadores de AyS basados en el marco de JMP	■	Indicadores de AyS limitados y desactualizados (definiciones ODM)

Ejecución

El proyecto se llevó a cabo entre enero de 2023 y julio de 2023 por un pequeño equipo compuesto por miembros del equipo de SCL Data y OLAS. Se evaluaron y armonizaron más de trescientas encuestas y sus conjuntos de datos y cuestionarios asociados de acuerdo con las definiciones deseadas de los indicadores de agua y saneamiento, abarcando 23 países entre 2003 y 2022. Estos conjuntos de datos armonizados se utilizaron luego para generar indicadores a nivel de país y se desglosaron en varias dimensiones socioeconómicas. Este esfuerzo se dividió en dos sprints, el primero abarcó encuestas de 2013-2022 y el segundo armonizando información de 2003-2012, después de lo cual se realizaron cinco rondas de validación, lo que

permitió ajustes y correcciones iterativas. Los resultados del proyecto se presentaron a los puntos focales de país de OLAS el 22 de agosto de 2023.

Metodología

El proceso de creación del conjunto de datos consta de tres pasos:

- Armonización de los microdatos de cada encuesta,
- Generación de indicadores a partir de los datos armonizados, y
- Unificación, limpieza y validación del conjunto de datos final.

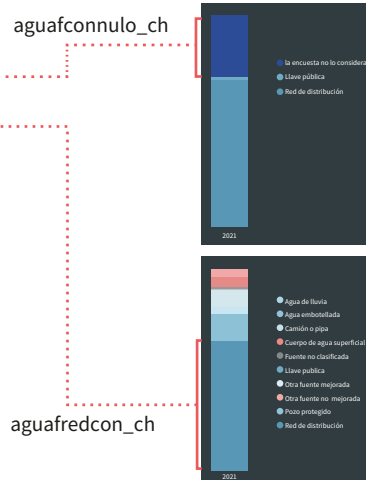
Figura 1: Normas para armonizar la información sobre fuentes de agua potable. Los valores de esta variable intermedia se utilizaron luego para generar los indicadores finales.

Ejemplo de la armonización de una variable ASH

Variable intermedia

Variable intermedia	Descripción:	Reglas:	Categorías
aguaconsumo_ch	Fuente de agua utilizado para consumo humano en el hogar.	Si la encuesta no diferencia entre agua para beber versus agua para uso general, asigne 0. Si su categorización no es clara (ejemplo: "pozo" en vez de pozo protegido) asigne 10	0 no pregunta sobre agua para beber o no especifican que la fuente es potable
			1 red de distribución, llave privada
			2 red (llave pública, standpipe)
			3 agua embotellada
			4 pozo protegido
			5 agua de lluvia
			6 camión, cisterna, pipa
			7 otra fuente mejorada
			8 cuerpo de agua superficial
			9 otra fuente no mejorada
			10 pozo, manantial y otra sin clasificación clara

Indicadores finales



De estos tres pasos, la armonización de los microdatos es el más vulnerable porque determina qué valores y campos de las encuestas originales contribuirán a cada indicador. El proceso consiste en crear nuevos campos en el conjunto de microdatos de cada encuesta que representen la misma información en todos los países y años. Armonizar un gran número de diferentes preguntas y respuestas de encuestas presenta un desafío importante debido a la heterogeneidad de las encuestas en toda la región. Por lo tanto, las variables de armonización se diseñaron para permitir la categorización de datos ambiguos y lagunas de datos.

armonizaron todos los datos, se ejecutó el proceso de generación de indicadores, que es común a todos los conjuntos de datos, y se generaron los archivos de datos de los indicadores. Luego, estos archivos se compilaron, limpiaron y analizaron para identificar posibles problemas.

Puede encontrar información detallada sobre las tres etapas de la generación de datos en los documentos de respaldo y los repositorios de códigos asociados:

La Figura 1 proporciona un ejemplo de las categorías de armonización para fuentes de agua utilizadas para consumo humano. La creación de un valor (0) que representa a las personas a las que no se les preguntó sobre el tema, así como las categorías 7, 9 y 10 para las fuentes que no caen claramente en las categorías más específicas, permite que cualquier opción de respuesta disponible en una encuesta nacional ser categorizado. Cada una de estas clasificaciones se utiliza a su vez para crear un indicador final en el proceso de generación de indicadores.

[Guía de armonización de variables intermedias.](#)

[Proceso de armonización GitHub](#)

[Creación de indicadores GitHub](#)

[Diccionario de datos para los Indicadores SCL](#)

[Proceso de limpieza GitHub](#)

[Documento de datos y metodología de OLAS](#)
[Conjunto de datos de la encuesta de hogares](#)

El proceso de armonización implicó la revisión de más de 300 encuestas, cuestionarios y conjuntos de microdatos. Una vez que se

Resultados

El conjunto de datos resultante contiene 47 indicadores disponibles durante 20 años para 23 países de ALC, que a su vez pueden analizarse a través de varios lentes, como edad, ingresos, zona (urbana/rural), etnicidad y más. Este amplio conjunto de datos permite un análisis en profundidad del acceso al agua y al saneamiento con respecto a la desigualdad socioeconómica, lo que a su vez puede ayudar a los formuladores de políticas a abordar las brechas de accesibilidad. El conjunto de datos, así como los paneles que permiten un fácil análisis de los datos, están disponibles en OLAS. Puede encontrar un resumen de la información del nuevo conjunto de datos en:

[FS 1.3.b: Conjunto de datos de las encuestas de hogares de OLAS: Actualización 2023](#)

[FS 1.4.b: Actualización 2023 del conjunto de datos de la encuesta de hogares de OLAS: Indicadores de acceso al agua](#)

[FS 1.5.b: Actualización 2023 del conjunto de datos de la encuesta de hogares de OLAS: Indicadores de acceso al saneamiento](#)

Copyright © 2024 Banco Interamericano de Desarrollo (BID). Esta obra se encuentra sujeta a una licencia Creative Commons CC BY 3.0 IGO (<https://creativecommons.org/licenses/by/3.0/igo/legal-code>). Se deberá cumplir los términos y condiciones señalados en el enlace URL y otorgar el respectivo reconocimiento al BID.

En alcance a la sección 8 de la licencia indicada, cualquier mediación relacionada con disputas que surjan bajo esta licencia será llevada a cabo de conformidad con el Reglamento de Mediación de la OMPI. Cualquier disputa relacionada con el uso de las obras del BID que no pueda resolverse amistosamente se someterá a arbitraje de conformidad con las reglas de la Comisión de las Naciones Unidas para el Derecho Mercantil (CNUDMI). El uso del nombre del BID para cualquier fin distinto al reconocimiento respectivo y el uso del logotipo del BID, no están autorizados por esta licencia y requieren de un acuerdo de licencia adicional.

Note que el enlace URL incluye términos y condiciones que forman parte integral de esta licencia.

Las opiniones expresadas en esta obra son exclusivamente de los autores y no necesariamente reflejan el punto de vista del BID, de su Directorio Ejecutivo ni de los países que representa.

