Observatorio Económico Social UNR

XX Jornadas de Ciencias, Tecnologías e Innovación
29 y 30 de octubre de 2026 · Rosario

PUBLICACIÓN DE MICRODATOS DE LA EPH COMO DATOS ABIERTOS EN UN REPOSITORIO DATAVERSE

Germán Tessmer · german.tessmer@unr.edu.ar

Observatorio Económico Social, Vicerrectoría, Universidad Nacional de Rosario (UNR)

Palabras clave: microdatos; Encuesta Permanente de Hogares; datos abiertos; repositorio institucional.

Accedé a los datos

Introducción

La **Encuesta Permanente de Hogares (EPH)** del Instituto Nacional de Estadística y Censos (INDEC) es la fuente más usada para conocer el empleo, los ingresos y la pobreza en la Argentina. Releva cada trimestre 31 aglomerados urbanos, cerca del 62 % de la población urbana. Pero sus microdatos llegan en dos bases codificadas que solo se entienden con un manual de códigos. **Un dato disponible todavía no es accesible**: para ser un bien público tiene que poder encontrarse, citarse y reutilizarse.

Objetivo

Publicar la EPH como **datos abiertos** en el Repositorio de Datos Académicos de la UNR (plataforma Dataverse): una versión **curada, ampliada y documentada**, de uso libre y persistente, legible aun sin ser especialista. El principio rector es que **el trabajo técnico se hace una sola vez**, en el Observatorio, y no lo repite cada usuario.

La serie y sus quiebres

- **2016** Inicio de la serie (4° trimestre).
- **2019** Nuevas proyecciones de población.
- **2020–2022** Encuesta telefónica (pandemia).
- **2021** Ponderadores diferenciados.
- **2024** **Quiebre:** nuevo cuestionario, 76 variables.
- **2026** Versión 8.0, serie comparable.

Figura 1. Hitos de la EPH entre 2016 y 2026.

Metodología

1. **Bases del INDEC.** Hogares e individuos, trimestrales, desde el 4° trimestre de 2016.
2. **Curado y ampliación en R.** Base única por persona, cada variable y valor etiquetados, y variables propias.
3. **Armonización del rediseño de 2024.** Se reconstruyen las variables discontinuadas y se corrigen errores de origen.
4. **Publicación en Dataverse.** Cada trimestre se exporta a Stata y R y se sube al repositorio. La plataforma genera la versión tabular y el libro de códigos (*codebook*) en el estándar internacional DDI (Iniciativa de Documentación de Datos).
5. **Datos abiertos.** Identificador de objeto digital (DOI) y licencia Creative Commons (CC BY 4.0), para usar desde R, Stata y Python, con dos programas públicos en R para reconstruir o separar las bases.

Resultados y discusión

416 variables documentadas — **320 de la EPH** | **96 propias**

84 archivos publicados

+21.000 visualizaciones

Lectura directa. Cada variable y cada valor tienen nombre y etiqueta comprensibles: la base se entiende **sin abrir el manual de códigos** del INDEC. Todo se documenta en un **Diccionario de 416 fichas** y en un **Manual Metodológico**, con número internacional de publicación seriada (ISSN). El diccionario también está en formato JSON, para consultarlo desde R, Python o con inteligencia artificial.

Parte A. Armonización del rediseño de 2024

A fines de 2024 el INDEC rediseñó el cuestionario: por primera vez mide la **formalidad de los trabajadores independientes**, separa los ingresos no laborales por fuente y elimina agregados históricos. El repositorio **reconstruye las variables eliminadas** con tablas de verdad de 64 escenarios categóricos y 250 de ingresos, corrige errores de los archivos crudos y evita el doble conteo.

calificacion_ocupado — CREADA

Tipo: factor · **Base:** individuos

Significado: clasifica a los ocupados según los conocimientos y habilidades que exige su trabajo principal.

Valores: Profesional | Técnico | Operativo | No calificado | Ns/Nc | Anómalo

Figura 2. Una de las 416 fichas del Diccionario, en versión web (HTML) y PDF.

Parte B. Variables de elaboración propia

El índice de **Necesidades Básicas Insatisfechas (NBI)**, con sus cinco dimensiones de privación:

Hacinamiento | Vivienda | Sanitarias | Escuela | Subsistencia

También, **ingresos reales** deflactados por el Índice de Precios al Consumidor (IPC) de cada región, que permiten comparar el poder adquisitivo entre trimestres y regiones; el desglose de los clasificadores de ocupación y de actividad; la antigüedad laboral y el lugar de origen.

¿Quiénes lo usan?

Investigación | Organismos públicos | Organizaciones sociales | Consultoras y empresas | Periodismo

Conclusiones

La EPH puede consultarse hoy como un conjunto de datos **abierto, documentado y citable**: 84 archivos con DOI y licencia CC BY 4.0. La serie se mantiene **comparable a pesar del rediseño de 2024**, y cada trimestre nuevo se incorpora bajo el mismo estándar. Así, el Observatorio consolida una política de datos abiertos: el trabajo técnico se hace una sola vez y queda disponible para todos, en línea con la **Ley N.° 27.275 de Acceso a la Información Pública**.