

EL ACCESO A DATOS ADMINISTRATIVOS PARA LA INVESTIGACIÓN

Olympia Bover

Directora del Departamento de Análisis Estructural y
Estudios Microeconómicos

Fundación Areces
10 de diciembre 2019

DEPARTAMENTO DE ANÁLISIS ESTRUCTURAL Y ESTUDIOS MICROECONÓMICOS



ÍNDICE

1. Cuestiones generales
2. Protocolos de acceso a universos de datos administrativos para la investigación
3. El BELab (laboratorio de datos del BdE)
4. Complementariedad entre encuestas y datos administrativos en la investigación



- **Los datos administrativos tienen un potencial inmenso para la investigación empírica**
- **También para la elaboración de las estadísticas económicas**
- **En teoría los datos administrativos son preferibles a las encuestas en tres dimensiones**
 - Contienen muchas más observaciones. Se necesitan grandes muestras para estudiar los eventos raros (por ejemplo 0,999 percentil de la distribución de la renta)
 - Tienen una dimensión longitudinal. Seguir a los individuos en el tiempo es necesario para estudiar transiciones y efectos a largo plazo
 - Por lo general proporcionan información de más calidad que la de las encuestas (ausencia de no respuesta o de respuestas parciales)
- **La frontera de la investigación se ha movido al uso de datos administrativos. Estos datos ya han permitido obtener resultados muy influyentes (por ejemplo, trabajos de Piketty y Saez)**

2. SISTEMAS DE ACCESO A LOS DATOS ADMINISTRATIVOS

1) Anonimizar vs. compartir la custodia

- El énfasis en la anonimización tiene límites, especialmente cuando se trata de datos de empresas.
- Más allá de la anonimización, se trata de *compartir la custodia* de estos datos, ejercida hasta ahora exclusivamente por los funcionarios responsables de las agencias respectivas, con miembros autorizados de la comunidad científica
- El que los investigadores accedan a estos datos se justifica por el beneficio público que se deriva de sus investigaciones

2) Características de un sistema de acceso viable

- Según Card, Chetty, Feldstein y Saez (2010) en su libro blanco dirigido al NSF, un programa de acceso a los datos sostenible y eficiente debe reunir las condiciones siguientes:

- Competencia abierta por el acceso a los datos basada en el mérito científico
- Suficiente ancho de banda para acomodar proyectos múltiples de forma simultánea
- Inclusión de investigadores jóvenes y estudiantes de postgrado en los equipos de investigadores con acceso a los datos
- Modo de acceso: acceso directo a los micro datos anonimizados in situ o preferiblemente a través de conexiones remotas
 - *El acceso a datos procesados en vez de los registros brutos o el envío de códigos ejecutables a las agencias no resuelve el problema*
 - *El envío de ejecutables elimina en la práctica la posibilidad de análisis exploratorio o interactivo con los datos*

- La posibilidad de fusión de datos de distintas fuentes es esencial para la investigación.

2. SISTEMAS DE ACCESO A LOS DATOS ADMINISTRATIVOS

3) Problemas de accesibilidad previsibles: la viabilidad política

- **Dado que los datos administrativos no se recogen principalmente con fines de investigación, el público puede tener dudas sobre su privacidad, la vinculación de sus datos de diferentes fuentes y el uso por parte de los investigadores.**
- **La desconfianza en el gobierno afecta la capacidad de implementar programas eficientes de acceso a los datos administrativos.**
- **Para la sostenibilidad del acceso a los micro datos administrativos en la perspectiva política, la divulgación es esencial. La mayoría de gestores y ciudadanos no son conscientes de los beneficios para la sociedad de la investigación que utiliza datos administrativos.**
- **En el caso de cruces de datos de distintas fuentes (de distintas administraciones sectoriales o territoriales), la desconfianza entre administraciones puede ser un gran problema.**

4) Problemas de replicabilidad

- **La replicabilidad de la investigación es fundamental para garantizar buenos estándares en la investigación y para el progreso científico.**
- **En el caso de los datos administrativos, las preocupaciones con la confidencialidad pueden imposibilitar la replicabilidad.**
- **Esto es un problema que por ahora no tiene buena solución pero pone de manifiesto la importancia de que la capacidad de acceso sea amplia.**

2. SISTEMAS DE ACCESO A LOS DATOS ADMINISTRATIVOS

5) Ejemplo de procedimientos de acceso: universo de afiliados a la SS en Europa

- **El acceso al universo de afiliados requiere una propuesta de investigación que debe ser aprobada por el órgano competente.**
- **Existen diferentes vías de acceso. En algunos casos (Alemania, Portugal) el investigador se desplaza a un centro de datos**
 - Estos terminales pueden estar físicamente en las oficinas del organismo que da acceso a los datos o este organismo puede llegar a acuerdos con otros centros para dotarles de un terminal con estas características.
- **En otros casos (Francia y países escandinavos), se permite el acceso remoto desde un terminal personal a una carpeta en red que tiene datos.**
- **Una vez autorizado, el investigador tiene que firmar un acuerdo de confidencialidad.**

6) Formación de usuarios de datos confidenciales

- **Se trata de un aspecto importante sobre el que existe experiencia internacional que se puede aprovechar para evitar errores que cometieron otros**
- **Formación a usuarios de datos confidenciales en el RU (Ritchie, Green, Newman, and Parker 2017)**
- **Inicialmente la formación tenía por objetivo asustar, instruir, autorizar y transferir la responsabilidad**
- **Se vio que esta estrategia no funcionaba con los científicos**
- **La estrategia actual se basa en educar, concienciar, ser francos acerca de los riesgos de filtraciones y compartir responsabilidad.**

3. BELAB (LABORATORIO DE DATOS DEL BDE) (I)



Lanzamiento en julio pasado del BELab

En una primera fase, solo acceso presencial a los datos: *data room* en las instalaciones del BdE

Pero se está estudiando acceso remoto



- **Por ahora, datos de Central de Balances pero el BdE tiene el propósito de ir añadiendo nuevas bases de datos**
- **Esta plataforma nace con el deseo de ser una vía de acceso no solo a datos propios del BdE sino a otras fuentes de datos de interés para los investigadores (por ejemplo, datos de la Seguridad Social alemana; conversaciones avanzadas con German Institute for Employment Research)**
- **La posibilidad de tener más de un mecanismo de acceso a una fuente de datos puede ser atractiva desde el punto de vista de**
 - Ampliar la capacidad del sistema en conjunto
 - Permitir experimentar con distintos protocolos
 - Facilitar enlaces entre fuentes que estén disponibles dentro de una plataforma

Aspectos que es poco probable que estén contenidos en los datos administrativos:

- Uso del tiempo: consumo, búsqueda, tareas en el empleo, en el ocio, en la producción del hogar, cuidado de niños y mayores
- Variables auto-reportadas: salud, riqueza, competencias adultas (financieras y otras)
- Expectativas: renta, mortalidad, entorno económico (por ejemplo, expectativas probabilísticas de precios de la vivienda y renta junto con variable de gasto en la EFF, expectativas de inflación Survey NY Fed)

Combinación de datos de encuestas y datos administrativos:

- Vinculación de encuestas y datos administrativos: una forma de introducir variables de alta calidad en muestras de encuestas sin preguntar a los encuestados (ejemplo: Registro de crédito y datos de encuesta a empresas para separar factores de demanda y factores de oferta en la cantidad de crédito Jiménez, Ongena, Peydró, and Saurina: "Hazardous Times for Monetary Policy...", Econometrica 2014)
- Registros administrativos como base para marcos de encuestas vinculadas más pequeñas.
- Encuestas como el PSID en EE.UU. o la EFF en España tienen la característica especial de proporcionar datos longitudinales conjuntos sobre riqueza, ingresos y gastos a nivel de hogar. Tener estas dimensiones juntas es importantes en los estudios de desigualdad. Los datos administrativos de renta proporcionan información complementaria.

GRACIAS POR SU ATENCIÓN

DEPARTAMENTO DE ANÁLISIS ESTRUCTURAL Y ESTUDIOS
MICROECONÓMICOS

