

Por un cambio de cultura en la gestión de los datos en España: Una propuesta de reforma

EsadeEcPol Insight #17 Octubre 2020

AUTORES

Miguel Almunia

Profesor, CUNEF

Pedro Rey-Biel

Profesor titular, ESADE

Apoyan la iniciativa

Samuel Bentolilla (CEMFI), Antonio Cabrales (Carlos III), Sara de la Rica (Fundación ISEAK y Universidad del País Vasco), Juan José Dolado (Universidad Carlos III), Natalia Fabra (Universidad Carlos III), Lidia Farré (Universidad de Barcelona y IAE-CSIC), Florentino Felgueroso (Fedea), Rafael Frutos Vivar (Estadístico del Estado), Jesús Gonzalo (Universidad Carlos III), Lucas Gortázar (World Bank, EsadeEcPol), Manuel Hidalgo (Universidad Pablo Olavide, EsadeEcPol), Nagore Iriberry (Universidad del País Vasco), Víctor Lapuente (Universidad de Göteborg), Mónica Martínez Bravo (CEMFI), Andreu Mas-Colell (Universitat Pompeu Fabra), Toni Roldán (EsadeEcPol), Marta Suárez-Varela (EsadeEcPol, UAM)

RESUMEN EJECUTIVO

- **La gestión de la información durante la pandemia ha puesto al descubierto graves debilidades para gestionar y coordinar datos en España.** Estas debilidades no afectan solamente al ámbito sanitario, sino que se extienden al resto de la administración. Mientras que, en los últimos años, numerosos países de nuestro entorno han realizado avances significativos en la digitalización y apertura de la administración, España se está quedando rezagada. Debemos abordar dos retos principales: (1) mejorar nuestra capacidad de recolectar y procesar datos y (2) expandir la capacidad de analizarlos.
- **Una mayor transparencia y disponibilidad de datos tendría efectos enormemente positivos para todos:** un uso más eficiente de los recursos públicos, una mejor investigación científica, una mejor capacidad de respuesta ante los retos económicos y sociales a los que nos enfrentamos y una mayor calidad democrática. La buena noticia es que el Plan de Recuperación para Europa nos ofrece una oportunidad única para una reforma en la dirección correcta, con una dotación prioritaria de fondos destinada a “reformas para mejorar la eficiencia de las administraciones públicas a través de la digitalización”.
- **Respecto a cómo mejorar nuestra capacidad de recolectar y procesar datos, proponemos:** (1) establecer mecanismos de coordinación entre distintos niveles de la administración para hacer los sistemas informáticos compatibles entre sí y fomentar el intercambio de datos, (2) realizar una inversión ambiciosa en capital humano, contratando informáticos, ingenieros y expertos en ciencia de datos, y (3) un compromiso firme de mantener esta iniciativa al margen de la disputa política, al tratarse de un proyecto de país.
- **Para dar respuesta al segundo reto y expandir nuestra capacidad de analizar datos,** nuestras propuestas se basan en las mejores experiencias internacionales y responden a cuatro preguntas fundamentales: ¿Qué institución debe custodiar los datos? ¿Cómo se garantizará la confidencialidad? ¿Cuáles deberían ser los criterios de acceso a los datos? ¿Cómo se accedería a los datos? En concreto, proponemos que se formalice un consorcio entre instituciones y organismos públicos (INE, la Agencia Tributaria, la Seguridad Social, el Banco de España, la AIREF y los Gobiernos Autonómicos) y cuáles deberían ser los criterios para el acceso a esos datos que garanticen el respeto a la privacidad y un acceso ágil a la comunidad científica.

Contenidos

- **Resumen ejecutivo**
- **01. Introducción**
 - Reto 1:
Mejorar nuestra capacidad de recopilar y procesar datos administrativos
 - Reto 2:
Mejorar nuestra capacidad de analizar los datos administrativos
- **02. Copiar a los mejores: ¿Qué modelos de otros países deberían inspirarnos?**
 1. El modelo nórdico
 2. El modelo continental
 3. El modelo anglosajón
- **03. El acceso a datos administrativos en España**
- **04. Propuestas para modernizar la gestión de la información administrativa en España**
 - Propuestas para el Reto 1:
Modernizar la gestión de los datos en las administraciones públicas
 - Propuestas para el Reto 2:
Expandir nuestra capacidad de analizar políticas públicas
- **Conclusiones**
- **Referencias**

01. Introducción

La recolección y procesamiento de datos son actividades esenciales que los estados deben realizar para cumplir muchas de sus funciones básicas. Las administraciones públicas necesitan registros de la actividad económica de los contribuyentes para recaudar impuestos, información sobre los contagiados (y sus contactos), ingresados y fallecidos para afrontar una pandemia como la actual, o registros de afiliación a la seguridad social para gestionar los sistemas de previsión social, por mencionar solo algunos ejemplos. No es casualidad que el término “estadística” provenga originalmente del latín *statisticum*, “relacionado con el estado”.

Los enormes avances en las tecnologías de la información y la comunicación (TIC) permiten el almacenamiento de una cantidad cada vez mayor de datos a un coste cada vez más bajo. Esto ha llevado a los gobiernos de muchos países a crear grandes registros administrativos sobre, por ejemplo, tributación, salud, educación y programas sociales. A su vez, la mejora de nuestra capacidad de procesamiento y análisis de datos supone una oportunidad de expandir el conocimiento utilizando los registros administrativos. La posibilidad de realizar evaluaciones de las políticas públicas y diseñar nuevas políticas basadas en la evidencia existente depende crucialmente de la disponibilidad y el aprovechamiento de los datos.

Nuestra administración se enfrenta a **dos retos: (1) mejorar su capacidad de recolectar y procesar datos, y (2) expandir la capacidad de analizarlos.** Los eventos recientes muestran claramente que nuestro país se está quedando rezagado en estas dos dimensiones, por lo que es urgente identificar los problemas y proponer soluciones. En el resto de este informe analizamos en profundidad estos dos retos, describimos los sistemas de gestión y acceso a datos que han puesto en marcha países de nuestro entorno y, finalmente, realizamos una serie de propuestas para modernizar la gestión de datos en España.

Reto 1:

Mejorar nuestra capacidad de recopilar y procesar datos administrativos

APRENDER DE LOS ERRORES EN LA GESTIÓN DE LA PANDEMIA

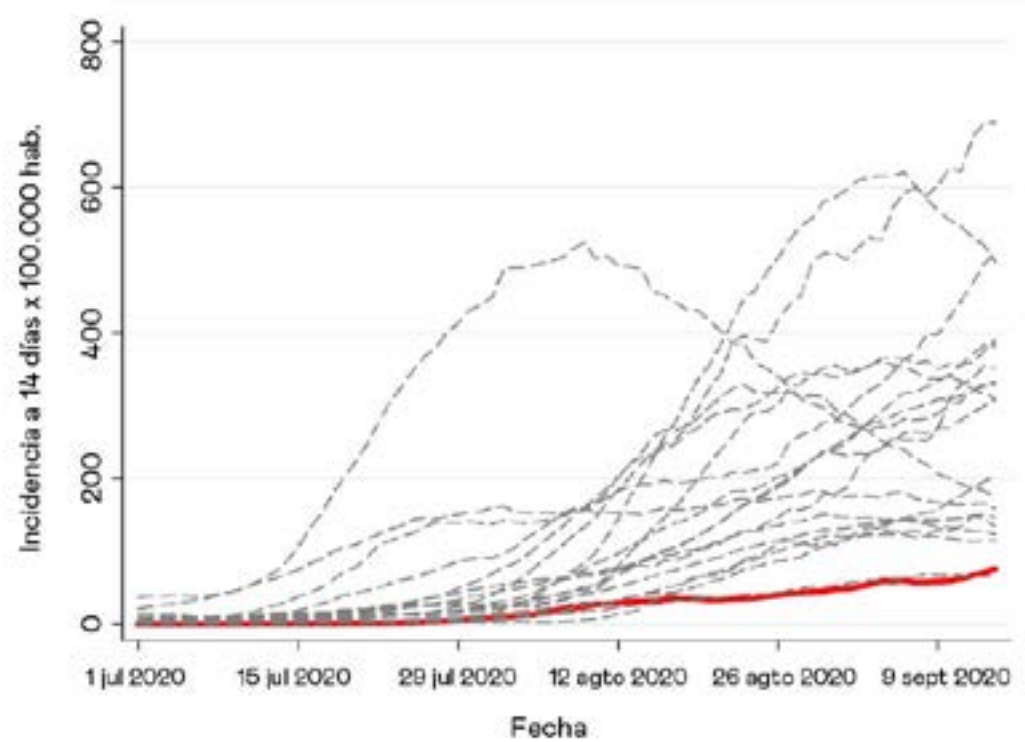
La capacidad de las administraciones públicas de recopilar y procesar datos administrativos de manera eficiente tiene un efecto directo sobre su capacidad de gestionar las políticas públicas. La gestión de la pandemia de covid-19 ha puesto de manifiesto las limitaciones de la gestión de los datos por parte de nuestras administraciones públicas. Hemos sido incapaces de frenar la expansión del virus dos veces consecutivas en el espacio de pocos meses debido a que los sistemas de detección y rastreo no han estado a la altura. La experiencia de los países que han tenido más éxito en frenar la expansión del virus (por ejemplo, [Corea del Sur](#), [Alemania](#) o [Vietnam](#)) ha demostrado que las estrategias más efectivas para el control de la pandemia han sido aquellas centradas en la recopilación de datos (tests de diagnóstico y rastreo de contactos) para frenar las cadenas de infección. Para que estas estrategias sean efectivas es necesario disponer de sistemas de recogida de información ágiles y coordinados, que deben estar financiados con suficientes recursos.

“ La gestión de la pandemia ha puesto de manifiesto las limitaciones de la gestión de los datos por parte de nuestras administraciones públicas

En buena parte, nuestro fracaso se ha producido por los problemas de coordinación entre distintos niveles de la administración. La falta de estadísticas homogéneas y sistemas de información compatibles entre sí ha dificultado el seguimiento de la situación epidemiológica en tiempo real en todo el territorio. Por otro lado, el sentido de propiedad sobre los datos generados por cada administración, y los intentos de su uso con fines políticos, ha impedido aprovechar el valor de estos como un bien común. Estos problemas han sido especialmente dañinos ante un virus que no entiende de fronteras territoriales y genera externalidades negativas.

No todas las administraciones lo han hecho igual de mal. En algunas comunidades autónomas, como [Asturias](#), se invirtió pronto para expandir la capacidad de realizar tests y en el rastreo, lo que ha resultado en una incidencia menor que en el resto del país (ver Figura 1-A). Otras, como Castilla y León, han destacado por la transparencia en la difusión de datos sobre la evolución de la epidemia (ver Figura 1-B y esta [página web](#)). Estos casos de éxito no están alineados con una ideología concreta, ya que en la primera gobierna el PSOE con IU y en la segunda el PP con Ciudadanos. Lo que diferencia la actuación de estas dos comunidades autónomas es que han diseñado sus políticas de acuerdo con la evidencia existente y han dedicado los recursos necesarios para mejorar su gestión de los datos.

Figura 1-A
Evolución de la
incidencia de
contagios por
CCAA. La línea
roja corresponde
a Asturias



Fuente: Instituto de Salud Carlos III, M^o de Sanidad

Figura 1-B
Página web con
datos sobre
covid-19 del
Gobierno de
Castilla y León



MÁS ALLÁ DEL ÁMBITO SANITARIO

Los problemas en la gestión de los datos en España se extienden más allá del ámbito sanitario. La administración ha mostrado también poca agilidad para evaluar las solicitudes del ingreso mínimo vital, la reforma estrella del sistema de protección social. Desde su aprobación a finales de mayo hasta el 20 de agosto, solo se habían aprobado 6.000 de las 750.000 solicitudes (un 0,8%), aparte de las 74.000 aprobadas de oficio. También se han producido retrasos en el cobro de los ERTE, por el colapso del Servicio Público de Empleo Estatal (SEPE). De cara al futuro, es urgente hacer una revisión general de los sistemas de recolección y procesamiento de datos de nuestras administraciones para mejorar su capacidad de respuesta a este tipo de situaciones.



Los problemas en la gestión de los datos en España se extienden más allá del ámbito sanitario

Otro ejemplo más lejano en el tiempo es la oportunidad desaprovechada con la implantación del DNI electrónico (DNI-e). El DNI-e se creó en el año 2006 con un alto coste para el erario público (314 millones de euros invertidos entre 2005 y 2009, según un informe de la Comisión Técnica de Apoyo a la Implantación del DNI-e). Su principal característica es que incorpora un chip en el que se puede almacenar información sobre su propietario. El potencial de esta iniciativa era enorme: estaba llamado a digitalizar la mayor parte de las interacciones de los ciudadanos con la administración, facilitar la firma electrónica e incluso incorporar datos relativos a la licencia de conducir o la tarjeta sanitaria (para reducir el número de tarjetas). Más de una década después, la única información que incluye el chip en la mayoría de casos es la que está escrita sobre el propio DNI, y apenas el 0,2% de los trámites con la administración se realizan utilizando el DNI-e. Esto no es del todo sorprendente si se tiene en cuenta que para poder firmar electrónicamente con el DNI-e, los ciudadanos tienen que acudir presencialmente a una oficina para obtener un certificado electrónico al menos una vez cada 4 años, un periodo más corto que el de vigencia del DNI. De nuevo, se ha desaprovechado una oportunidad de utilizar la tecnología para aumentar la eficiencia de la administración.

APROVECHAR LA OPORTUNIDAD: UN MARCO LEGAL FAVORABLE Y FONDOS EUROPEOS

El marco legal existente no es el problema, ya que prevé eliminar trabas administrativas a la coordinación y crear un entorno de transparencia, pero los resultados de la aplicación de estas leyes son poco alentadores. El artículo 149 de la Constitución Española establece que el Estado tiene competencia exclusiva sobre la estadística para fines estatales (punto 1, 31ª), lo que debería garantizar la cooperación entre administraciones. La Ley de transparencia,

acceso a la información pública y buen gobierno ([Ley 19/2013, de 9 de diciembre](#)) y la Ley sobre reutilización de la información del sector público ([Ley 37/2007, de 16 de noviembre](#)) apuntan también en la dirección correcta (para más detalles sobre el marco regulatorio, [ver Rivera y Garrido Chamorro, 2020](#)). Pese a la existencia de este marco legal, este mismo año ha fracasado un intento de uniformizar los sistemas informáticos utilizados para gestionar los historiales sanitarios en cada Comunidad Autónoma por falta de consenso para reformar el Sistema Nacional de Salud.

“ Para que las reformas que proponemos sean viables, será necesario dotar a las administraciones públicas de los recursos humanos necesarios

Para que esta legislación sea realmente aplicada y las reformas que proponemos sean viables, será necesario dotar a las administraciones públicas de los recursos humanos necesarios. La edad media de los trabajadores de la Administración General del Estado era de [52 años en 2019, casi 10 años más](#) que el conjunto de la población activa. Los funcionarios que accedieron a la administración hace décadas carecen, en su gran mayoría, de formación en el manejo más puntero de bases de datos. Esa falta debe suplirse con la modernización de la formación de quienes trabajan con datos administrativos y **la contratación de personal con nuevas capacidades –ingenieros de sistemas, informáticos, expertos en ciencia de datos– en todos los organismos públicos.**

La inversión necesaria para implementar estas reformas es sustancial, pero el retorno a largo plazo es potencialmente muy alto por las enormes posibilidades que ofrece. Del mismo modo que las empresas más productivas utilizan indicadores detallados de la actividad en cada una de sus fábricas y tiendas para aumentar su productividad (Bloom y Van Reenen, 2010), la administración puede explotar su enorme riqueza de datos para incrementar la eficiencia en la gestión de las políticas públicas. La disponibilidad de fondos del [Plan de Recuperación para Europa](#) asignados específicamente a “reformas para mejorar la eficiencia de las administraciones públicas a través de la digitalización” supone una oportunidad que no podemos dejar pasar.

DOS EJEMPLOS DE ÉXITO: HACIENDA Y LA DGT

El reto no es inasumible. Algunos organismos públicos han sabido modernizar sus sistemas para mejorar su eficiencia. Un ejemplo claro es la recaudación de impuestos. Si la Agencia Tributaria (AEAT) no tuviese abundante información sobre la actividad económica que desarrollan ciudadanos y empresas, le resultaría mucho más difícil recaudar impuestos. La información detallada que obtiene la AEAT al cruzar múltiples fuentes de información es lo que le permite obtener un alto nivel de cumplimiento fiscal sin tener que realizar costosas inspecciones a un porcentaje elevado de los contribuyentes. Esto ha sido posible gracias a la enorme inversión que ha hecho la AEAT en sistemas informáticos para recopilar y, especialmente, cruzar datos de distinta procedencia.

De manera similar, la labor de control del tráfico en carretera sería mucho más difícil si la Dirección General de Tráfico (DGT) careciera de toda una red de radares y cámaras que monitorizan el comportamiento de los conductores. Sin el uso de estas tecnologías para recolectar datos y un procesamiento ágil de los mismos, la DGT tendría muchas dificultades para sancionar a los conductores que no cumplen con normas como los límites de velocidad. Gracias a estas inversiones y a reformas acertadas como el carnet por puntos, la DGT ha logrado una notable reducción en las muertes por accidentes de tráfico en las últimas décadas.

“ La inversión necesaria para implementar estas reformas es sustancial, pero el retorno a largo plazo es potencialmente muy alto

Reto 2:

Mejorar nuestra capacidad de analizar los datos administrativos

PERMITIR EL ACCESO LOS DATOS PARA INVESTIGACIÓN: UNA ESTRATEGIA GANADORA PARA TODOS

Este segundo reto estriba en la capacidad limitada de las administraciones públicas para analizar la ingente cantidad de datos que poseen, lo que supone un despilfarro tremendo de los recursos destinados a recolectarlos y una ineficiencia respecto al enorme potencial que el uso de esos datos podría tener para el desarrollo y la evaluación de las políticas públicas.

La administración no puede, ni debe, tener el monopolio del análisis de los datos que posee. No puede, en primer lugar, porque no es su función principal. La prioridad de la administración es gestionar las políticas públicas, no realizar un análisis profundo de los datos de los que dispone. En segundo lugar, porque los recursos públicos disponibles para este fin son reducidos. Aunque la administración cuenta con algunos organismos especializados en producir información estadística, como el Instituto Nacional de Estadística (INE) o el Centro de Investigaciones Sociológicas (CIS), los recursos económicos de los que disponen para evaluar políticas públicas son limitados. Por último, la administración no debe tener el monopolio del análisis de las políticas públicas porque puede tener un conflicto de interés al hacerlo. La creación de una entidad independiente como la AIReF ha significado un paso hacia delante en este sentido, pero su mandato está restringido a la evaluación de políticas fiscales.

“ La administración no puede, ni debe, tener el monopolio del análisis de los datos que posee

Permitir que la comunidad científica tenga acceso, de manera controlada, a registros administrativos resultaría en un enorme incremento de nuestra capacidad para analizarlos. Esto expandiría nuestro conocimiento sobre la efectividad de las políticas públicas, aumentaría la transparencia de la actuación de las administraciones y, por lo tanto, profundizaría la calidad de nuestra democracia (Arellano, 2018). Los incentivos están alineados: la administración se beneficiaría porque su capacidad de analizar los datos es limitada, mientras que la comunidad investigadora tendría la oportunidad de explotar esta ingente fuente de información para realizar investigación puntera.

Los nuevos conocimientos obtenidos se podrían utilizar para diseñar políticas públicas basadas en la evidencia, algo que desafortunadamente es poco habitual en nuestro país. Estos objetivos solo pueden ser temidos por aquellos que hacen un uso partidista de la información y no comparten que la democracia se fortalece cuanto mayor sea el conocimiento que tienen los ciudadanos sobre el efecto de las políticas que se les aplican.



Los nuevos conocimientos obtenidos se podrían utilizar para diseñar políticas públicas basadas en la evidencia

VENTAJAS Y LIMITACIONES DEL USO DE DATOS ADMINISTRATIVOS PARA INVESTIGACIÓN

El uso de datos administrativos presenta varias ventajas respecto a las encuestas para la evaluación de las políticas públicas. Al cubrir a toda la población, tienen un tamaño muestral mucho mayor, corrigiendo posibles sesgos de selección que a menudo afectan a las encuestas. Los datos administrativos tienen una tasa de no respuesta mucho menor, lo cual permite disponer de bases longitudinales (panel) más extensas en el tiempo y con menor desgaste (*attrition*). Además, tienen menos errores de medición, aunque estos nunca se eliminan por completo.

Por último, **los registros administrativos se pueden cruzar con otras fuentes de datos utilizando identificadores individuales o de hogar. Esto es esencial para poder realizar análisis más profundos sobre causalidad y evaluación de políticas públicas.** Por ejemplo, una base de datos con los historiales laborales de todos los trabajadores que a su vez incluya información sobre las características de las empresas permitiría estudiar el impacto de cambios en el salario mínimo sobre distintos tipos de empresas según su tamaño, sector, ubicación geográfica, etc. Es imposible realizar este tipo de estudio con los datos de encuestas disponibles actualmente.



Cruzar los registros administrativos con otras fuentes de datos es esencial para poder realizar análisis más profundos sobre causalidad y evaluación de políticas públicas

Los datos administrativos también presentan algunas limitaciones que se deben tener en cuenta. En primer lugar, los datos se recogen atendiendo a objetivos de gestión, lo que afecta al formato, estructura y alcance de los datos. Por ello, el número de variables disponibles en cada registro suele ser limitado. Por ejemplo, los datos fiscales no suelen tener información socioeconómica de los contribuyentes, más allá de la edad y el sexo.

Esta es otra de las razones por las que la posibilidad de cruzar datos de distintos registros es esencial para explotar su verdadero potencial. Además, implica que la construcción de bases de datos para investigación supondrá un importante esfuerzo. Idealmente, este trabajo lo deberían realizar conjuntamente funcionarios públicos e investigadores, dado que los primeros tienen el mejor conocimiento de los datos y los segundos tienen mayor experiencia analizando datos para investigación. Por último, el volumen de las bases de datos administrativos es grande, a veces difícil de manejar. Sin embargo, esta limitación es superable dado el rápido avance en las tecnologías de almacenamiento en la nube y el desarrollo de nuevas herramientas de análisis como el machine learning o la inteligencia artificial.

LA IMPORTANCIA DE PROTEGER LA CONFIDENCIALIDAD

Un factor muy importante por considerar es que los datos administrativos contienen información sensible sobre los ciudadanos y es esencial proteger la confidencialidad e integridad de los datos. Para ello, deben utilizarse métodos de anonimización de los datos, eliminando identificadores individuales y garantizando que no se pueda trazar la identidad de ningún individuo concreto usando otras variables. Igualmente, deben crearse protocolos que garanticen que el acceso a los datos sea seguro desde este punto de vista, como han hecho otros países.

La mayoría de los países de nuestro entorno han creado instituciones y sistemas de acceso a datos administrativos que protegen la confidencialidad. Algunos de estos sistemas se describen en la siguiente sección de este informe. No tenemos que reinventar la rueda: podemos inspirarnos en los modelos que han creado estos países y adaptarlos a la estructura administrativa de España.

“La mayoría de los países de nuestro entorno han creado instituciones y sistemas de acceso a datos administrativos que protegen la confidencialidad”

La reciente [Directiva UE 2019/1024 de 20 de junio de 2019](#), relativa a los datos abiertos y la reutilización de la información del sector público, es bastante clara al respecto: prevé que “cualquier documento o conjunto de datos conservado por organismos públicos (incluidas empresas públicas y organismos de investigación financiados públicamente) sea puesto a disposición general del público, para fines comerciales o no comerciales, y en formatos abiertos, legibles por máquina, accesibles, fáciles de localizar y reutilizables, junto con sus metadatos”. Esta Directiva debe ser objeto de transposición a la legislación nacional de los países miembros antes del 17 de julio de 2021.

Afortunadamente, no partimos de cero. En España existen ya algunas bases de datos administrativos disponibles para la investigación, como la Muestra Continua de Vidas Laborales o el Panel de Declarantes del IRPF, que se describen más adelante. Además, existen casos aislados de investigadores que han conseguido obtener el acceso a ciertos datos, la mayor parte de las veces gracias a contactos personales y a eternas gestiones, no regidas con criterios objetivos sobre el interés público de las investigaciones que proponen. No existen, sin embargo, protocolos generalizados que regulen la colaboración y la cesión de datos entre las administraciones y la comunidad científica. Es obvio que en un país con un alto grado de descentralización y con tantos niveles de administración como el nuestro no es sencillo acordar y escoger entre las diversas formas de hacerlo. Por ello, precisamente, es necesario un debate sosegado, que se enriquezca con la experiencia ganada de otros países de nuestro entorno, con realidades políticas muy diversas.

La disponibilidad de mejores datos también permitiría diseñar medidas políticas bajo criterios que permitieran su evaluación, creando una colaboración activa, y no solo pasiva, entre investigadores y administraciones. En este sentido, destacan iniciativas como el [Behavioral Insights Team](#) o las agencias del tipo [What Works](#) (“¿Qué Funciona?”) del Reino Unido. La creación de agencias que colaboren con investigadores para la realización y evaluación de políticas basadas en la evidencia puede ser uno de los muchos pasos necesarios hacia la modernización de la administración pública en España.



La disponibilidad de mejores datos también permitiría diseñar medidas políticas bajo criterios que permitieran su evaluación

Afortunadamente, muchos de estos pasos (disponibilidad de datos públicos, creación de una agencia de evaluación, apertura a la colaboración científica para el desarrollo de intervenciones que permitan basar las políticas en la evidencia), están alineados, crean sinergias y pueden marcar un verdadero cambio en la forma de tomar decisiones políticas en nuestro país. Una señal esperanzadora en este sentido es que el [Real Decreto-Ley por el que se aprobó el Ingreso Mínimo Vital](#) incluye en su artículo 30 el compromiso de evaluar los efectos de esta política. Esperemos que esa sea la norma en un futuro, y no la excepción.

02. Copiar a los mejores: ¿Qué modelos de otros países deberían inspirarnos?

Existen diversos modelos para facilitar el acceso público a datos administrativos, cada uno con sus ventajas y sus inconvenientes. Obviamente, la estructura administrativa de cada país es determinante para la elección de un modelo u otro. Aquí describimos las características principales de tres modelos existentes y, a continuación, comentamos las ventajas e inconvenientes que existen para seguir uno u otro modelo en España.

02. 1. El modelo nórdico

DATOS CENTRALIZADOS EN EL INSTITUTO NACIONAL DE ESTADÍSTICA, ACCESO REMOTO, COBRO DE TASAS POR USO DE LOS DATOS

Los países nórdicos (Dinamarca, Finlandia, Noruega y Suecia) [1] han sido **pioneros en la creación de mecanismos para que la comunidad científica pueda realizar investigaciones utilizando sus datos administrativos de forma segura**. En estos cuatro países, el organismo encargado de recopilar datos de diferentes registros administrativos es el instituto nacional de estadística respectivo. **Estos institutos operan como un “tercero seguro”**: reciben datos de diversas entidades gubernamentales –que no necesariamente quieren compartir datos entre sí– y se encargan de combinar los datos (usando identificadores individuales) y modificar su formato para la investigación. Las páginas web de estos institutos de estadística tienen un diseño moderno y accesible, y todas ellas incluyen una sección en la que se explica cómo se puede acceder a sus microdatos para fines de investigación (se pueden consultar aquí: [Dinamarca](#), [Finlandia](#), [Noruega](#) y [Suecia](#)).

Estos repositorios incluyen microdatos anonimizados sobre población (censos), empleo (historiales de cotización a la seguridad social), empresas (registros mercantiles), impuestos (renta, sociedades, IVA y especiales), sanidad (historiales hospitalarios), educación (resultados académicos), vivienda (precios de venta y alquiler), justicia (procesos judiciales), medio ambiente (datos históricos de clima y de partículas contaminantes) y comercio internacional (transacciones de importación y exportación).

“ Los países nórdicos han sido pioneros en la creación de mecanismos para que la comunidad científica pueda realizar investigaciones utilizando sus datos administrativos de forma segura

Los criterios de acceso de investigadores a los datos varían entre países: en el caso de Dinamarca y Noruega es necesario tener una afiliación a una de las instituciones de investigación del país que esté aprobada por el instituto de estadística, mientras que los datos de Finlandia y Suecia se pueden obtener teniendo una afiliación con una institución dentro del Espacio Económico Europeo.

En todos los casos, los investigadores deben realizar una solicitud detallada en la que explican el diseño de su estudio y las bases de datos específicas a las que necesitan acceder para su investigación. En algunos casos, los datos están disponibles en el formato deseado, pero en otros se requiere un trabajo por parte del instituto de estadística para vincular múltiples bases de datos. Existe también la posibilidad de que los investigadores aporten datos externos y el instituto de estadística los cruce con sus propios datos utilizando identificadores individuales, para crear una nueva base de datos anonimizada.

El acceso a los datos se realiza en todos los casos de forma remota por internet. Es decir, los investigadores reciben unas claves de acceso para conectarse a un servidor remoto donde están almacenados los datos y está instalado el software necesario para su análisis. Cada investigador tiene acceso exclusivamente a los datos de su estudio y por tanto no comparte el espacio con otros investigadores. El acceso remoto no permite la transferencia directa de archivos al ordenador de los investigadores. Cuando estos quieren extraer sus resultados, los archivos son revisados por personal del instituto de estadística para garantizar que se cumplen los requisitos de confidencialidad.

Los institutos de estadística cobran tasas por el uso de sus datos a los investigadores. Estas tasas incluyen un cargo por cada hora de trabajo que suponga la preparación de las bases de datos (alrededor de 100-150€ por hora de trabajo) y también el coste de mantenimiento del servidor remoto (entre 150€ y 300€ por mes según el volumen de datos almacenado), entre otros. Estas tasas garantizan que el sistema es sostenible, dado que asegura que se pueden financiar los costes, tanto de recursos humanos como de sistemas informáticos, aunque aumente el número de proyectos de investigación.

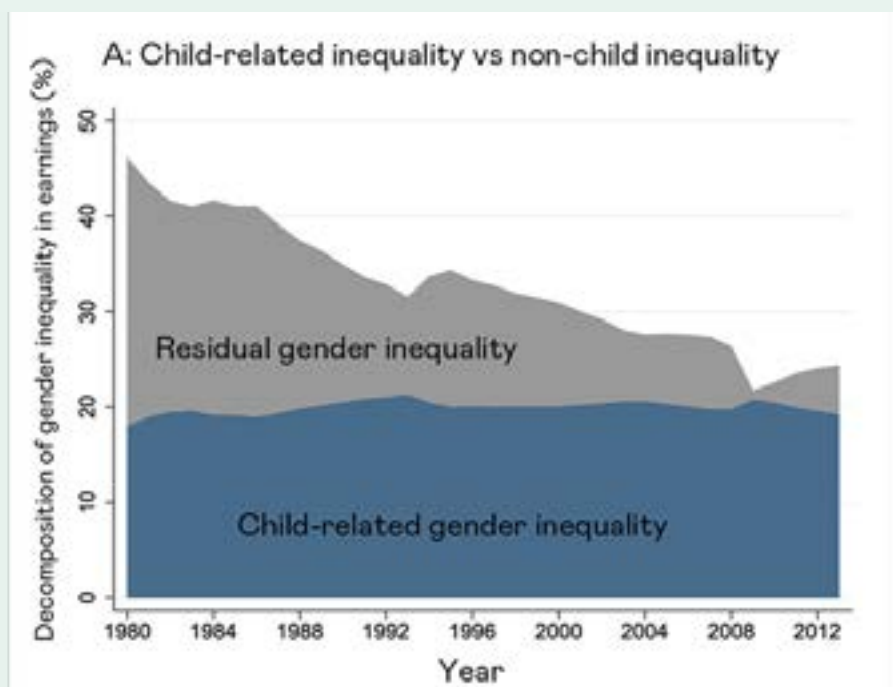


BOX 1:

los efectos de distintos incentivos sobre el ahorro y los efectos de los hijos sobre la brecha salarial entre hombres y mujeres

El acceso a datos ha dado como fruto un número considerable de estudios académicos sobre temas muy diversos. Un ejemplo interesante es el trabajo de [Chetty et al. \(2014a\)](#), que estudia las decisiones de ahorro de los trabajadores en Dinamarca. Los investigadores combinan datos individuales de múltiples registros administrativos: declaraciones del impuesto sobre la renta, censo de población, una base de datos integrada de relaciones laborales, saldos de cuentas bancarias y aportaciones anuales a planes de pensiones (individuales y de empresa). Esta impresionante combinación de datos les permite observar todas las decisiones de ahorro del conjunto de la población danesa en el periodo 1995-2009. Los autores encuentran que los incentivos fiscales al ahorro previsional tienen un efecto moderado sobre el ahorro total (incluyendo planes de pensiones y otros instrumentos financieros).

Sin embargo, estiman que un 85% de los daneses son ahorradores “pasivos”, que aceptan las contribuciones ofrecidas por sus empleadores por defecto, y solo un 15% son ahorradores “activos”, que reajustan su uso de distintos instrumentos de ahorro para aprovechar al máximo los incentivos fiscales, sin cambiar su ahorro total. Estos resultados tienen unas implicaciones claras para el diseño de los incentivos al ahorro previsional: es más efectivo regular que los planes de pensiones de empresa tengan, por defecto, un porcentaje alto de contribución que ofrecer un incentivo fiscal a todas las contribuciones a planes de pensiones.



Fuente: Kleven et al. (2019a)

Otro estudio que ha tenido un gran impacto es el de [Kleven et al. \(2019a\)](#), que estudia la penalización laboral que sufren las mujeres cuando tienen su primer hijo. Los autores combinan datos de impuestos sobre la renta, censos de población, relaciones laborales y registros de nacimiento, que les permiten vincular a padres y madres con sus hijos. El principal hallazgo es que, con el nacimiento del primer hijo, la renta anual de la madre cae un 20%, mientras que la del padre se mantiene constante.

Esta brecha se mantiene inalterada durante los siguientes 10 años, y la misma tendencia se observa en las horas trabajadas, la tasa de empleo, y en el salario por hora. Como se ve en la figura, aunque la brecha salarial está decreciendo, el nacimiento del primer hijo explica cada vez un porcentaje mayor de la brecha en Dinamarca. Este estudio, junto con otros posteriores que se han hecho para otros países (incluyendo uno para España: De Quinto, Hospido y Sanz, 2020), indican que para cerrar la brecha salarial entre hombres y mujeres es fundamental diseñar políticas públicas que lleven a un reparto más equitativo de la crianza de los hijos entre padres y madres.

02. 2. El modelo continental

CONSORCIOS ENTRE VARIAS INSTITUCIONES PÚBLICAS Y DE INVESTIGACIÓN, ACCESO REMOTO, COBRO DE TASAS POR USO

Otros países de Europa continental han adoptado modelos diferentes al nórdico, pero que también están dando como resultado una fructífera colaboración entre la administración pública y la comunidad investigadora. La principal diferencia con el modelo nórdico es que no se construye alrededor del instituto nacional de estadística, sino que se han creado **instituciones dedicadas específicamente a la tarea de recopilar y preparar los datos para su uso en investigación**, además de la gestión de las solicitudes y los proyectos.

“Otros países de Europa continental han creado instituciones dedicadas específicamente a la tarea de recopilar y preparar los datos para su uso en investigación”

En el caso de **Francia**, en el año 2010 se creó un centro de acceso seguro a datos ([CASD](#)) que **actúa como tercero seguro** y dispone de registros administrativos de impuestos, trabajo, empresas, finanzas, justicia, medio ambiente, agricultura y sanidad, todos ellos provenientes de los diferentes ministerios y del instituto de estadística francés (INSEE). **El CASD está constituido como un**

“grupo de interés público” que reúne a varias instituciones: INSEE, GENES (las grandes *éco/es nationales*), CNRS (centro nacional de investigación) y dos universidades: École Polytechnique y HEC Paris. Según se indica en su [página web](#): “El objetivo principal de este grupo, de carácter industrial y comercial, es organizar e implementar servicios de acceso seguro a datos confidenciales para investigación, estudio, evaluación o innovación sin ánimo de lucro.”

Una característica particular del CASD es que ha creado su propia tecnología, la **SD-box**, para permitir el acceso remoto a los datos por parte de investigadores externos. Se trata de un aparato (similar a un descodificador) que, conectado a un monitor y a una red de internet, permite a los investigadores tener acceso remoto a los datos de su proyecto. Al igual que el modelo nórdico, los investigadores interesados tienen que enviar una solicitud detallando el objetivo de su estudio y las bases de datos requeridas, y el servicio tiene un coste que depende de varios factores. En este caso, los investigadores con proyectos aprobados deben viajar a París para recibir un breve curso sobre las normas para proteger la confidencialidad de los datos.

Para su funcionamiento, el CASD se apoya en tres comités ([ver aquí](#)): un comité científico, un comité de productores de datos y un comité de seguimiento de la política de seguridad de los sistemas de información. El primero, formado por 16 investigadores, asesora a la dirección del CASD en materias de prospectiva, innovación, ética y estrategia científica. El segundo asesora a la dirección en asuntos relacionados con las condiciones de acceso a los datos, la documentación, el archivo y la difusión de información. El tercero asesora sobre temas relacionados con la protección de la confidencialidad e integridad de los datos. Una lista completa (e impresionante) de todos los estudios que se están realizando con datos del CASD está disponible [aquí](#).

BOX 2:

Evaluación de políticas públicas en Francia gracias a la creación del CASD en 2010

Un ejemplo de estudio realizado con estos datos es el trabajo de [Garicano, Le Large y Van Reenen \(2016\)](#), que analizan los efectos de las regulaciones laborales sobre el tamaño empresarial. Para ello utilizan datos sobre la población de empresas manufactureras en Francia, mostrando que hay una notable acumulación de empresas justo por debajo del umbral de 50 trabajadores, a partir del cual se aplican una serie de regulaciones que hacen el coste de los empleados mayor para las empresas. Los autores estiman que estas regulaciones reducen la competitividad global de las empresas francesas, al incentivar la actividad de las más pequeñas, que son menos productivas en promedio.

Uno de los grandes beneficiarios de la creación del CASD ha sido el [Institut de Politiques Publiques](#), un think tank creado por la Paris School of Economics y las Grandes Écoles para promover la evaluación cuantitativa de las políticas públicas utilizando métodos de vanguardia. Este instituto ha tenido un notable impacto sobre el debate público en Francia, beneficiando al conjunto de la sociedad con sus evaluaciones independientes.

En el caso de **Portugal**, en los últimos años se han desarrollado varios mecanismos de acceso a datos administrativos. Por un lado, en 2014 [se formalizó](#) la creación de un consorcio entre el Instituto Nacional de Estatística (homólogo de nuestro INE), la Fundação para a Ciência e Tecnologia (similar a nuestro CSIC) y la Direção Geral das Estatísticas da Educação e Ciência. El objetivo de este consorcio es proveer de acceso a los microdatos del INE a científicos afiliados a centros de investigación portugueses, incluyendo a estudiantes de doctorado y máster (en cuyo caso la solicitud debe ser refrendada por sus directores académicos). Los datos disponibles incluyen registros administrativos de empresas, relaciones laborales y comercio internacional, así como una gran variedad de encuestas con información demográfica, de salud y educación, entre otros. El acceso a los microdatos se debe realizar en las instalaciones del INE portugués o en centros seguros aprobados por el mismo.

BOX 3:

La reforma en el acceso a datos de 2014 en Portugal ha permitido conocer mejor las causas de la brecha salarial

Una de las bases de datos administrativos que más atención ha recibido de los investigadores son los [Quadros de Pessoal](#), que recogen información detallada de todas las relaciones laborales en el sector privado (incluyendo duración de los contratos y remuneraciones) en Portugal durante varias décadas. Utilizando esta base de datos, [Card, Cardoso y Kline \(2016\)](#) analizan la contribución de las empresas a la brecha salarial entre hombres y mujeres en Portugal. Gracias a la riqueza de estos datos, pueden distinguir dos efectos: por un lado, las mujeres tienen un menor poder de negociación salarial; por otro lado, las mujeres tienden a trabajar en empresas que pagan menores salarios. Estos dos efectos explican un 20% de la brecha salarial total entre hombres y mujeres en Portugal.

Además de los sistemas de acceso a datos en cada país, en 2017 se creó INEXDA (International Network for Exchanging Experience on Statistical Handling of Granular Data), una **red de bancos centrales** ([principalmente europeos](#), entre ellos el Banco de España) e institutos estadísticos. Los objetivos de este grupo son compartir experiencias en el manejo estadístico de datos granulares (también llamados microdatos) y homogeneizar los datos que gestionan los bancos centrales, de forma que las estadísticas producidas en distintos países sean comparables entre sí. El objetivo a largo plazo es que los datos estén disponibles para el análisis, el diseño de políticas públicas y la investigación. Aunque todavía no existe una lista completa de los datos que estarán disponibles, se pueden consultar las [conclusiones del grupo de trabajo](#) que ha definido los protocolos de acceso a los datos, las normas para combinar distintos registros y la información que se recolectará de los equipos encargados de cada proyecto de investigación.

02. 3. El modelo anglosajón

INICIATIVAS GUBERNAMENTALES DESCENTRALIZADAS, LIDERADAS POR EL EMPUJE DE INVESTIGADORES DE PRESTIGIO

En los países anglosajones también ha habido un enorme interés recientemente por el uso de datos administrativos para la investigación. **El modelo de acceso es generalmente más descentralizado que en los países nórdicos o de Europa continental, pues distintos organismos gubernamentales diseñan sus propios sistemas y protocolos de acceso a datos.**

“ En los países anglosajones el modelo de acceso a datos administrativos es generalmente más descentralizado que en los países nórdicos o de Europa continental

En **Estados Unidos**, un grupo de economistas de las universidades de Harvard y Berkeley escribió en septiembre de 2010 una [carta abierta a la National Science Foundation \(NSF\)](#) para pedir que se establecieran sistemas de acceso a datos administrativos (Card, Chetty, Feldstein y Saez 2010). En esta carta, argumentaban que se estaba realizando una gran cantidad de evaluaciones de políticas públicas en países europeos por la mayor disponibilidad de datos administrativos, y que **Estados Unidos se estaba quedando atrás** en este aspecto, perdiendo la posición de país líder en investigación en ciencias sociales. [A raíz de esa iniciativa](#), instituciones como el Internal Revenue Service (la agencia tributaria estadounidense) pusieron en marcha sistemas de acceso a microdatos fiscales para investigación, dando lugar a una explosión de estudios sobre los efectos de las políticas públicas en Estados Unidos.

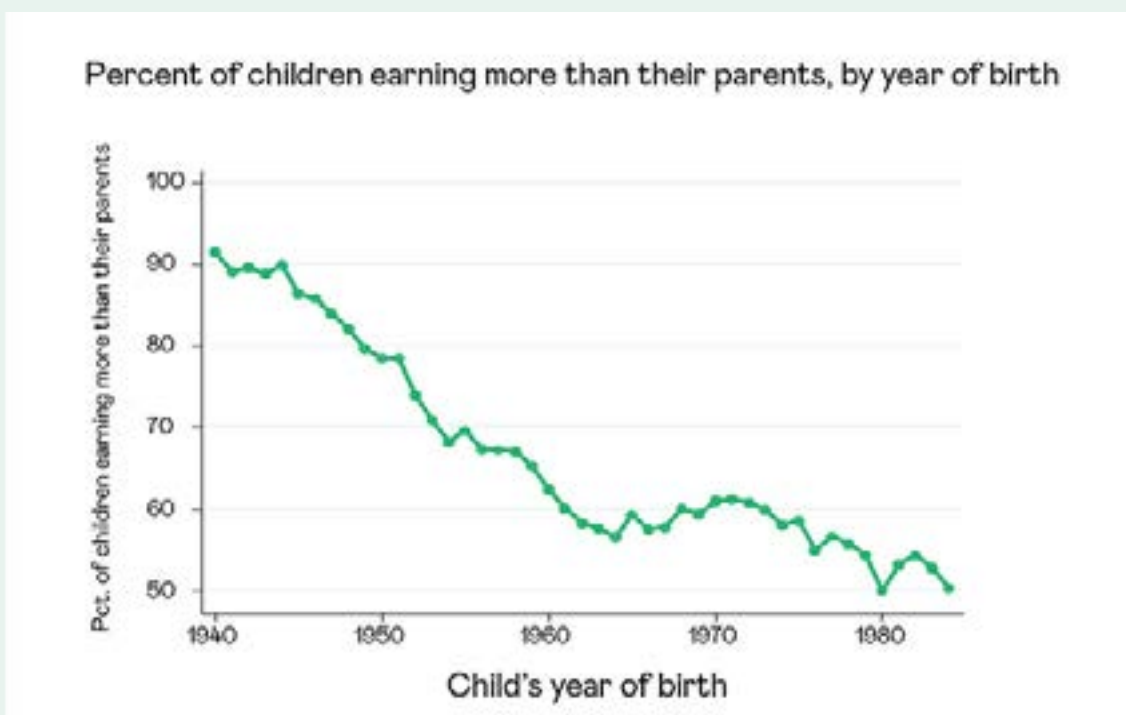
BOX 4:

Nueva evidencia sobre movilidad intergeneracional usando los microdatos tributarios

Uno de los trabajos resultantes analiza los cambios en la movilidad intergeneracional en Estados Unidos ([Chetty et al., 2014](#)). Utilizando datos de declaraciones del impuesto sobre la renta, observan la renta de todos los estadounidenses nacidos en 1980-82 cuando tienen aproximadamente 30 años (en 2011-12) y vinculan esta información con la renta declarada por sus padres en el periodo 1996-2000, cuando los hijos eran adolescentes. Esto les permite estimar la relación entre la posición relativa de la renta de los padres y la de los hijos cuando son adultos. Los autores encuentran una gran variación en la movilidad intergeneracional entre distintas ciudades: la probabilidad de que un niño nacido en una familia que está en el 20% más pobre de EEUU pase a estar en el 20% más rico del país en su edad adulta es del 12,9% en San José (en el Silicon Valley californiano) y solo del 4,4% en Charlotte (en el estado sureño de Carolina del Norte).

Comparando estos resultados con estudios de otros países, la movilidad intergeneracional en San José es similar a la de Dinamarca, pero la de Charlotte y otras ciudades del sureste es inferior a la de todos los demás países avanzados para los que existen datos. Los datos agregados de movilidad por ciudades se pueden [consultar en esta página web](#). Estos resultados sugieren que EEUU ya no es, en gran parte de su territorio, una “tierra de oportunidades” y que la idea del sueño americano cada vez choca más con la realidad de una sociedad desigual en la que el éxito de una persona depende más de quiénes son sus padres que de su esfuerzo y mérito personal.

Adicionalmente, el porcentaje de jóvenes que tiene mayor renta que sus padres al llegar a la edad adulta no ha parado de bajar (ver figura). Este estudio, y otros relacionados que también hacen uso de datos tributarios, ha tenido un enorme impacto en el debate público sobre la desigualdad económica y racial en EEUU.



Fuente: www.opportunityinsights.org/nationaltrends

En el caso del **Reino Unido**, la agencia tributaria (*Her Majesty's Revenue and Customs*, HMRC) también ha creado un mecanismo para permitir el acceso de investigadores a datos fiscales y aduaneros, llamado [HMRC DataLab](#). Al igual que en otros países, para obtener acceso a los datos se completa una solicitud en la que se explican los objetivos del estudio y las bases de datos necesarias. Para algunos proyectos se han combinado datos fiscales con otros datos externos, como por ejemplo los provenientes de registros mercantiles. El acceso a los datos solo se puede realizar presencialmente en una de las sedes de HMRC en Londres, lo cual limita parcialmente el acceso a investigadores residentes en otras ciudades del país. Además, el servicio no está sujeto al pago de ninguna tasa, una de las razones por las que no se ha realizado la inversión en un sistema de acceso remoto. (Se puede consultar [aquí](#) una comparación del sistema británico con el finlandés).

BOX 5:

Estudios sobre el efecto de las políticas fiscales en Reino Unido gracias al HMRC DataLab

Este sistema de acceso ha permitido que se realizaran investigaciones de alto impacto, como un estudio sobre los efectos de los impuestos a las transacciones inmobiliarias sobre el mercado de la vivienda ([Best y Kleven, 2018](#)), otro sobre los salarios obtenidos por estudiantes graduados en distintas universidades británicas ([Belfield et al., 2018](#)), una estimación del tipo impositivo efectivo que pagan las multinacionales extranjeras en Reino Unido en comparación con empresas locales del mismo tamaño ([Bilicka, 2019](#)) o el impacto de los incentivos fiscales a las donaciones a ONGs y fundaciones ([Almunia et al., 2020](#)).

Además, desde 2017 se puede acceder a microdatos del instituto de estadística británico ([Office of National Statistics](#), ONS). El acceso se realiza a través de una red de centros seguros distribuidos por todo el país, donde los investigadores deben acudir en persona. Dada la restricción al acceso remoto, la pandemia ha supuesto una limitación en el acceso a datos para diversos proyectos de investigación.

03. El acceso a datos administrativos en España

El acceso a datos administrativos para investigación en España es limitado en comparación con los países mencionados en la sección anterior, pero no partimos de cero. España comparte con otros países europeos una gran ventaja a la hora de combinar diferentes registros administrativos: la existencia de un identificador individual que coincide con el número del documento nacional de identidad (DNI).

En España, existen mecanismos para acceder a algunas bases de datos administrativos, como la Muestra continua de vidas laborales, el Panel de empresas-trabajadores y el Panel de declarantes del IRPF.

“El acceso a datos administrativos para investigación en España es limitado en comparación con otros países

La [Muestra Continua de Vidas Laborales \(MCVL\)](#), disponible desde 2004, es una muestra aleatoria que contiene información (anonimizada) sobre un 4% de los cotizantes y beneficiarios de la Seguridad Social. Se trata de una base de datos longitudinal que permite seguir el historial laboral de estos individuos a lo largo del tiempo. Estos datos se han utilizado en un gran número de estudios, por ejemplo para analizar la evolución de la desigualdad salarial a lo largo del ciclo económico ([Bonhomme y Hospido, 2017](#)) o evaluar la empleabilidad de los parados de larga duración ([Bentolila, García-Pérez, Jansen, 2017](#)).

El [Panel de Empresas-Trabajadores \(PET\)](#) se presentó en 2019 y supone un avance importante respecto a la MCVL. El muestreo se hace a nivel de empresa, en lugar de trabajador, de forma que los investigadores observan a todos los individuos que durante el periodo 2013-2016 han estado afiliados al Régimen General de la Seguridad Social a través de cada empresa seleccionada en la muestra. La muestra cubre un 3-5% de las empresas pequeñas (entre 1 y 9 trabajadores) y un 15% de las más grandes (más de 500 trabajadores). La manera de construir la muestra permitirá a los investigadores analizar cuestiones como los flujos de creación y destrucción de empleo, la contribución de los salarios a la desigualdad de la renta, los efectos del uso (y abuso) de los contratos temporales y la segregación horizontal y vertical de los trabajadores.

El [Instituto de Estudios Fiscales \(IEF\)](#) ofrece acceso a [microdatos del Impuesto sobre la Renta de las Personas Físicas \(IRPF\)](#) para el periodo 1999-2015 en dos formatos: muestras aleatorias anuales (que incluyen un 10-15% del total de declaraciones anuales) y un panel de declarantes (que incluye un 3% de las declaraciones anuales, en media). Las muestras son especialmente útiles para hacer [ejercicios de microsimulación](#), mientras que el panel es útil para analizar los efectos de reformas fiscales sobre el comportamiento de los contribuyentes, como se hace en un [estudio reciente](#) que estima la elasticidad de la renta respecto a cambios en el tipo impositivo marginal.

Estas tres bases de datos administrativos se distinguen de las descritas anteriormente para otros países en que se trata de muestras representativas. Es decir, incluyen solo un porcentaje de la población total. Desde el punto de vista del análisis estadístico, asumiendo que el muestreo esté hecho correctamente, es posible inferir características de la población con una muestra. Sin embargo, muchas preguntas de investigación solo se pueden abordar con datos poblacionales, porque las muestras no son representativas para grupos muy específicos. Por ejemplo, un estudio reciente de [Llaneras et al. \(2020\)](#) ha tratado de replicar el análisis de movilidad intergeneracional realizado por [Chetty et al. \(2014\)](#) para Estados Unidos, pero las limitaciones de los datos disponibles hacen que este análisis no sea directamente comparable (como ha argumentado [Polavieja, 2020](#)). Esta limitación también afecta a estudios sobre la desigualdad de la renta y la riqueza, en los que es crucial tener información detallada sobre las personas que están dentro del top 1% de la distribución.

“Una desventaja de las bases de datos es que, en la actualidad, no se pueden cruzar con otros registros administrativos”

Otra desventaja de estas bases de datos es que, en la actualidad, no se pueden cruzar con otros registros administrativos. Por lo tanto, la MCVL no permite realizar un análisis completo de los efectos de cambios en el salario mínimo sobre distintos tipos de empresas, según su tamaño, sector o localización geográficas. Tampoco podemos estudiar con datos administrativos el efecto de distintos niveles de educación sobre la renta, o los efectos de tener hijos sobre la desigualdad salarial entre los cónyuges. En cuanto a las bases de datos tributarios, no incluyen a las Comunidades Autónomas de Régimen Foral (Navarra y País Vasco), por lo que no se puede obtener una visión del conjunto del país ni se pueden analizar los flujos migratorios (de personas y empresas) desde y hacia estas dos Comunidades Autónomas como respuesta a diferencias en la regulación.

Por otro lado, **el Banco de España ha lanzado recientemente una iniciativa, el BELab, para dar acceso a investigadores externos a varias de sus bases de datos.** Desde julio de 2019, se puede acceder a la Central de Balances, que compila la documentación aportada anualmente por las empresas no financieras al Registro Mercantil. Un aspecto positivo de esta iniciativa es que se ha creado la infraestructura para el acceso seguro a datos (en las dependencias del Banco de España) y no se descarta que en el futuro sea posible el acceso remoto a los mismos, aunque en estos momentos no existe esa posibilidad.

“ El Banco de España ha lanzado una iniciativa para dar acceso a investigadores externos a varias de sus bases de datos

Por su parte, **el INE proporciona acceso a través de su página web a microdatos de las encuestas que realiza** ([ver aquí](#)). Estas encuestas tienen información valiosa, pero tienen las desventajas respecto a los datos administrativos que hemos mencionado anteriormente. Los únicos microdatos administrativos que se ofrecen en esta página son los de nacimientos y defunciones. Estos últimos son útiles para estadística descriptiva, pero, al no poder cruzarse con ninguna otra fuente de datos, no permiten realizar un análisis más profundo de relaciones entre distintas variables.

Finalmente, **a nivel de las Comunidades Autónomas hay algunas iniciativas prometedoras, como el DataResSS (Data for Research in Social Sciences), un consorcio de la [Barcelona Graduate School of Economics](#) y el Instituto de Estadística de Cataluña ([Idescat](#))**. El objetivo de este organismo es poner datos administrativos de Cataluña que posee Idescat a disposición de investigadores de vanguardia, tanto españoles como extranjeros. El DataResSS tiene un comité académico que evalúa las propuestas de investigación en función de su calidad académica, contribución al diseño de políticas públicas, cuestiones éticas relacionadas con la confidencialidad, y su factibilidad y extensión temporal esperada. Uno de los aspectos positivos del diseño de DataResSS es la cooperación entre un grupo de universidades de máximo nivel con el instituto de estadística de la Comunidad Autónoma.

04. Propuestas para modernizar la gestión de la información administrativa en España

Como hemos defendido a lo largo de este informe, **las administraciones públicas españolas necesitan urgentemente la modernización de sus sistemas de recopilación y procesamiento de datos**. En estos meses se abre una ventana de oportunidad histórica para acometer reformas, con el apoyo de una mayoría de la sociedad motivada por la deficiente gestión de la crisis sanitaria y el apoyo financiero que se nos ofrece con el [Plan de Recuperación para Europa](#). También hemos defendido que la experiencia de países de nuestro entorno muestra que es posible ofrecer un mejor acceso a datos administrativos de forma segura de lo que se hace actualmente en España. Para ello, hace falta un acuerdo entre las administraciones públicas y la comunidad científica para colaborar en un esfuerzo colectivo.

A continuación, presentamos propuestas sobre cómo creemos que se debe abordar este doble reto, enumerando una serie de aspectos clave y evaluando ventajas e inconvenientes de distintas opciones.

Propuestas para el Reto 1:

Modernizar la gestión de los datos en las administraciones públicas

- 01. Es necesario reforzar los mecanismos de coordinación entre distintos niveles de la administración, y también entre distintos organismos dentro del mismo nivel.** Para ello, la revisión de los sistemas de datos se debe incorporar de manera transversal a la coordinación entre administraciones en temas diversos como educación, sanidad, justicia, vivienda, medio ambiente, etc. Posiblemente se avanzará más rápido en unos ámbitos que otros, lo cual no es necesariamente malo. Esto permitirá que se exploren mecanismos y soluciones en algunas áreas donde los acuerdos son sencillos que luego se pueden trasladar a otras donde los fallos de coordinación sean mayores.
- 02. Para llevar a cabo esta modernización será necesario invertir de forma ambiciosa en capital humano.** No es suficiente con adquirir sistemas modernos de manejo de datos. Es crucial contratar a un número sustancial de trabajadores con la cualificación necesaria –informáticos, ingenieros de sistemas, expertos en ciencia de datos– que sean capaces de construir y gestionar esta nueva infraestructura en los diferentes organismos públicos. Esta inversión tendrá un elevado retorno a medio y largo plazo, con lo cual tendría sentido haberla realizado ya, pero ahora es especialmente atractiva al disponer de los fondos del Plan de Recuperación europeo.
- 03. Es muy importante que esta iniciativa se mantenga al margen de la disputa política por competencias entre administraciones, y que no se identifique con una tendencia política concreta.** Se trata de un proyecto de país del que todos nos beneficiaremos a través de una mejor gestión de las políticas públicas. Uno de los retos a los que se enfrenta este esfuerzo colectivo es la descentralización de buena parte de los servicios públicos. No consideramos que sean necesarios cambios legislativos para obligar a las administraciones autonómicas a ceder sus datos al gobierno central. No sería deseable afrontar este reto como un problema legal, ya que la imposición probablemente no solucionaría el problema y conduciría a una disputa política sobre competencias. Consideramos que sería mucho más productivo resaltar los beneficios mutuos que tendría la coordinación de estas tareas en términos de mejorar la calidad en la implementación de políticas públicas. En lugar de obligar a las comunidades autónomas a integrarse en un nuevo sistema para compartir los datos, podrían utilizarse las conferencias sectoriales para desarrollar mecanismos comunes. El Observatorio del Sistema Nacional de Salud puede suponer un ejemplo, aunque con algunos problemas, de lo que sería deseable.

04. **Aprovechar que una parte sustancial de los fondos del [Plan de Recuperación para Europa](#) está precisamente destinado a “reformas para mejorar la eficiencia de las administraciones públicas a través de la digitalización”.** Concretamente, el Plan Nacional para coordinar la modernización en la gestión de los datos administrativos podría obtener apoyo financiero del [Mecanismo de Recuperación y Resiliencia](#), cuyo presupuesto asciende a 560.000 millones de euros, de los cuales 310.000 millones se destinarán a subvenciones y 250.000 millones se destinarán a préstamos. La Comisión Europea ha indicado que estos fondos deben estar destinados a inversiones y reformas en materia ecológica y digital. La obtención de estos fondos se gestionará por procesos competitivos, por lo que es fundamental que España presente un proyecto ambicioso y completo.

Propuestas para el Reto 2:

Expandir nuestra capacidad de analizar políticas públicas

Nuestras propuestas sobre cómo poner en marcha un sistema de acceso a datos administrativos para la comunidad científica se centran en cuatro aspectos clave: ¿Qué institución, o instituciones, deben cumplir el papel de custodio de los datos y de “tercero seguro”? ¿Cuáles deberían ser los criterios de acceso a los datos? ¿Cómo se garantizará la confidencialidad de los datos? ¿Cómo se debería realizar el acceso a los datos?

(1) ¿Qué institución debería jugar el papel de depositaria de los datos y tercero seguro?

Una posible ventaja de adoptar el modelo nórdico, donde el instituto de estadística centraliza el acceso a todos los registros administrativos, es que nuestro INE ya dispone de un respaldo legal para recopilar y custodiar registros administrativos. Sin embargo, es importante ser conscientes de que el INE ya tiene asignadas una gran cantidad de funciones, como son la elaboración de la Contabilidad Nacional, el Censo de Población y numerosas encuestas. Por lo tanto, **si se quiere asignar esta nueva tarea al INE, será indispensable dotarlo de los recursos humanos y técnicos suficientes para llevarla a cabo.**

Una alternativa viable es seguir el modelo europeo continental, basado en la creación de consorcios de varias instituciones públicas con algunas entidades de investigación. En el caso de España, este tipo de consorcio podría incluir al propio INE, la Agencia Tributaria, la Seguridad Social, el Banco de España y la AIReF. Sería también deseable que las Comunidades Autónomas pudieran formar parte del consorcio, quizá a través de sus propios institutos de estadística. De hecho, no es necesario que se forme un solo consorcio. **Podría comenzarse a trabajar con acuerdos puntuales para compartir ciertas bases de datos** aprovechando

que algunas instituciones, como el Banco de España, el INE o la AIReF, ya tienen creada una infraestructura de centros seguros. Paulatinamente se podrían ir incorporando otras instituciones una vez se compruebe que el sistema funciona adecuadamente y se obtienen resultados de evaluación de políticas.

(2) Criterios de acceso a los datos

Para que el sistema funcione correctamente es importante que los criterios de acceso a los datos sean uniformes y transparentes, al contrario de lo que sucede actualmente. Cualquier investigador que desee acceder a los datos debería presentar una solicitud explicando el objeto de su estudio y especificando los registros a los que necesitaría acceder. Estas solicitudes deberán ser evaluadas por un **comité científico** en función de sus méritos científicos.

Además, deberá ser evaluada por un **comité de productores de datos** (representantes de las administraciones públicas) para evaluar la utilidad del estudio para el sector público y la viabilidad de ofrecer los datos requeridos en un formato apto para la investigación. Si estas evaluaciones tienen un resultado positivo, el investigador deberá firmar un contrato comprometiéndose a cumplir las reglas de uso y, especialmente, proteger la confidencialidad de los datos. Este contrato debe incluir la posibilidad de imponer sanciones al investigador si se incumplen las normas, con un objetivo disuasorio. Hay que tener en cuenta que cualquier tipo de infracción por parte de un investigador tendría también un efecto reputacional enorme, porque el resto de la comunidad científica podría ver en riesgo el acceso a los datos por la irresponsabilidad de una sola persona.

(3) Modo de acceso a los datos

El objetivo a medio plazo debe ser que exista la posibilidad de acceder a los datos de manera remota a través de internet, como se hace en varios países de nuestro entorno. Dicho esto, la secuencia habitual que han seguido otros países es desarrollar primero una infraestructura de centros seguros, en los que se maximiza el control de la actividad de los investigadores, que deben acudir presencialmente. Por ello, este suele ser el modo de acceso preferido de instituciones que ceden sus datos por primera vez.

Sin embargo, la experiencia internacional indica que, una vez los sistemas están en marcha y funcionan adecuadamente, se suele producir un aumento de la demanda de nuevos proyectos que en poco tiempo supera la capacidad de estos centros seguros presenciales. Además, **los centros seguros presenciales generan desigualdad territorial**, pues el acceso es mucho más sencillo para los investigadores que residen en la misma ciudad (habitualmente la capital del país) que para los que están en otros lugares.

Por todas estas razones, todos los modelos descritos anteriormente han terminado convergiendo hacia el acceso remoto a través de internet, haciendo uso de las tecnologías existentes. Es posible ofrecer acceso remoto y tener un control sobre la actividad de los investigadores muy similar al de los centros seguros presenciales. Por último, la opción del acceso remoto puede fomentar que accedan a los datos también **investigadores ubicados fuera de España**, expandiendo el grupo de investigadores potenciales que harán uso de los datos. En definitiva, la inversión a corto plazo en una tecnología segura de acceso remoto puede reducir enormemente los costes administrativos a largo plazo.

(4) Cómo garantizar la confidencialidad

Este aspecto, que en muchas ocasiones se cita como una barrera insalvable, se ha solventado en todos los países de nuestro entorno cuyos sistemas hemos descrito anteriormente. ¿Existe alguna diferencia radical entre el sistema legal español y el de otros países europeos que impida el uso de datos administrativos con fines de investigación? La respuesta es no, por la sencilla razón de que si los datos están correctamente anonimizados y existe un control sobre el uso que se hace de ellos, en ningún momento se produce una quiebra de la confidencialidad. Tanto la [Ley 19/2013, de 9 de diciembre](#), de transparencia, acceso a la información pública y buen gobierno, como la [Ley 37/2007, de 16 de noviembre](#), sobre reutilización de la información del sector público, son explícitas en este punto. La [Directiva UE 2019/1024 de 20 de junio de 2019](#) aplica por igual a todos los países miembros de la Unión Europea e indica claramente que cualquier documento conservado por organismos públicos sea puesto a disposición general del público, por lo que, con las debidas precauciones, el impulso legislativo desde Europa hacia la accesibilidad de los datos es tajante.

El paso más básico en el proceso de anonimización de los datos consiste en sustituir los identificadores individuales (por ejemplo, DNI de personas o NIF de empresas) por otros valores utilizando un algoritmo al que solo tenga acceso el depositario de los datos, pero nunca el investigador. En ocasiones, si las bases de datos administrativos son muy detalladas, puede ser necesario ir más allá y eliminar o distorsionar otra información (ubicación, edad, género, renta) que pueda permitir la identificación de un individuo o empresa concretos. Los detalles de estas técnicas van más allá del alcance de este informe, pero es importante resaltar que se ponen en práctica en todos los países mencionados anteriormente y por lo tanto no hace falta inventar nuevos métodos.

La práctica habitual en los centros de acceso a datos administrativos es que los investigadores no pueden extraer ningún resultado hasta que no sea revisado por personal del centro donde los haya consultado. Este personal se cerciorará de que no se viola la confidencialidad, por ejemplo, poniendo como requisito que haya un mínimo de unidades (personas, empresas) en cada celda de las tablas de resultados.

Conclusiones

La modernización de la administración pública para poder implementar y evaluar políticas basadas en la evidencia pasa necesariamente por **mejorar la recolección de datos públicos y la capacidad de análisis de los mismos**, mediante el acceso público a los datos a diversos agentes, como la comunidad investigadora.

España se está quedando atrás respecto a otros países de nuestro entorno, que han creado agencias públicas de acceso a datos administrativos siguiendo distintos modelos que se adaptan a las características propias de su administración. El éxito de estas agencias nos permite usar como ejemplo los mejores modelos de cada una de las importantes decisiones políticas que deben tomarse (centro seguro, confidencialidad, modo de acceso...) que mejor se adecúen a nuestra realidad territorial.

El cambio de cultura necesario para que nuestra muy fragmentada administración comparta y coordine sus datos supone un reto importante. Sin embargo, casos recientes muy notorios como la incorrecta gestión de la información durante la pandemia de covid-19 abren una ventana de oportunidad para conseguir un consenso de todos los sectores de la sociedad sobre que este cambio de cultura es imprescindible.

Existen casos aislados de éxito en nuestro país sobre el potencial que abre el acceso público a los datos administrativos. Estos casos muestran que en España ni partimos de cero ni es inviable conseguirlo. La articulación de un Plan Nacional para impulsar la creación de una agencia de datos puede ir dando pequeños pasos, utilizando aquellos datos y administraciones menos sensibles, que afiancen el gran cambio necesario.

NOTAS

- [1] Habitualmente se incluye a Islandia en este grupo de países, pero no lo hemos incluido en nuestro análisis dado su reducido peso en términos de población.



REFERENCIAS

Almunia, Miguel, Jarkko Harju, Kaisa Kotakorpi, Janne Tukiainen y Jouko Verho (2019): [Expanding access to administrative data: the case of tax authorities in Finland and the UK](#), International Tax and Public Finance, vol. 26(3), p. 661-676

Almunia, Miguel, Irem Guceri, Ben Lockwood and Kimberley Scharf (2020). [More giving or more givers?](#), Journal of Public Economics, vol. 183, 104110

Arellano, Manuel (2018): [El acceso a los microdatos administrativos públicos: la nueva frontera de la investigación económica y social](#), XVIII Aula de Verano Ortega y Gasset, Universidad Internacional Menéndez Pelayo (UIMP)

Belfield, Chris, Jack Britton, Franz Buscha, Lorraine Dearden, Matt Dickson, Laura Van Der Erve, Luke Sibieta, Anna Vignoles, Ian Walker, and Yu Zhu (2018). [The relative labour market returns to different degrees](#), Institute of Fiscal Studies

Bentolila, Samuel, José Ignacio García-Pérez, Marcel Jansen (2017). [Are the Spanish long-term unemployed unemployable?](#), SERIEs - Journal of the Spanish Economic Association, vol. 8(1): 1-41

Best, Michael, and Henrik Kleven (2018). [Housing market responses to transaction taxes: Evidence from notches and stimulus in the UK](#), Review of Economic Studies 85, 157–193

Bilicka, Katarzyna Anna. [Comparing UK tax returns of foreign multinationals to matched domestic firms](#). American Economic Review 109, no. 8 (2019): 2921-53

Bloom, Nicholas, and John Van Reenen. [Why do management practices differ across firms and countries?](#) Journal of Economic Perspectives 24, no. 1 (2010): 203-24

Card, David. Ana Rute Cardoso, Patrick Kline (2016). [Bargaining, sorting, and the gender wage gap: Quantifying the impact of firms on the relative pay of women](#), Quarterly Journal of Economics 131(2): 633-686

Card, David, Raj Chetty, Martin S. Feldstein, and Emmanuel Saez (2010). [Expanding access to administrative data for research in the United States](#). American Economic Association, Ten Years and Beyond: Economists Answer NSF's Call for Long-Term Research Agendas

Chetty, Raj, John N. Friedman, Søren Leth-Petersen, Torben Heien Nielsen, and Tore Olsen (2014a). [Active vs. passive decisions and crowd-out in retirement savings accounts: Evidence from Denmark](#). The Quarterly Journal of Economics 129, no. 3: 1141-1219

Chetty, Raj, Nathaniel Hendren, Patrick Kline, and Emmanuel Saez (2014b). [Where is the land of opportunity? The geography of intergenerational mobility in the United States](#). The Quarterly Journal of Economics 129, no. 4: 1553-1623

De Quinto, Alicia, Laura Hospido and Carlos Sanz (2020). [The child penalty in Spain](#), Documento Ocasional 2017, Banco de España

Einav, Liran, and Jonathan Levin. [Economics in the age of big data](#). Science 346, no. 6210 (2014)

Escrivá, José Luis (2018): [La cultura del análisis de datos](#), El País, 12 de julio de 2018

Garicano, Luis, Claire Le Large, and John Van Reenen. [Firm size distortions and the productivity distribution: Evidence from France](#). American Economic Review 106, no. 11 (2016): 3439-79

Kleven, Henrik, Camille Landais, and Jakob Egholt Sogaard. [Children and gender inequality: Evidence from Denmark](#). American Economic Journal: Applied Economics 11, no. 4 (2019): 181-209

Kleven, Henrik, Camille Landais, Johanna Posch, Andreas Steinhauer, and Josef Zweimüller. [Child penalties across countries: Evidence and explanations](#). In AEA Papers and Proceedings, vol. 109, pp. 122-26. 201

Llaneras, Kiko (2020). [Los problemas de usar datos del siglo pasado para una pandemia del siglo XXI](#), El País, 20 de junio de 2020

Llaneras, Kiko, Octavio Medina y Elena Costas (2020). [Atlas de oportunidades](#), proyecto conjunto de la Fundación COTEC y la Fundación Felipe González

Polavieja, Javier G. (2020). [Grandes datos, grandes sesgos, grandes errores: Sobre el Atlas de oportunidades](#), en prensa, Revista Internacional de Sociología, Vol 78, No 3

Rivera, Rafael y Pedro Garrido Chamorro (2020). [Lecciones del covid-19: España necesita una ley de datos](#), blog Hay Derecho, septiembre de 2020

Slemrod, Joel (2016): [Caveats to the research use of tax-return administrative data](#), National Tax Journal, vol. 69(4), p. 1003-1020