EL PAPEL DEL INE COMO GESTOR DE DATOS EN LA ADMINISTRACIÓN PÚBLICA

ALFREDO CRISTÓBAL CRISTÓBAL

Instituto Nacional de Estadística (INE)

Desde su creación, el INE ha gestionado eficientemente la información que la sociedad le ha confiado para realizar estadísticas oficiales, fundamentalmente a través de censos y encuestas a hogares y a empresas. Entre los diferentes elementos que garantizan la eficiencia de dicha gestión hay dos que pueden considerarse como fundamentales. Por un lado, la consideración de la carga a los informantes, que debe ser la mínima posible para cumplir

con los objetivos y necesidades marcados por las estadísticas y, por otro lado, la protección de la información suministrada por ciudadanos y empresas, eliminando cualquier posibilidad de que aparezcan brechas de confidencialidad.

La sociedad va evolucionando en el tiempo y, a día de hoy, vivimos en lo que podríamos llamar una sociedad altamente datificada. Los trámites administrativos que realizan personas y empresas alimentan continuamente bases de datos gestionadas por organismos públicos. También, ciudadanos y empresas dejamos rastro de nuestros datos en operaciones que alimentan grandes bases de datos, principalmente gestionadas por manos privadas, como consecuencia de la continua revolución digital (Big Data), en la que gran cantidad de transacciones se realizan a través de aplicaciones digitales residentes en dispositivos móviles.

En este nuevo ecosistema de datos convive información que, en la mayoría de los casos, no está

estructurada, por lo que es difícil su uso directo en la función estadística. Pero también debemos ser conscientes de que hoy no tenemos una completa necesidad de interactuar con la sociedad para proporcionar una información estadística de calidad, pues gran parte de los datos necesarios se encuentran en registros públicos o en bases de datos privadas. El acceso a dichas fuentes, G2G (1) en el caso público y B2G (2) en el caso privado permite la elaboración de estadísticas reduciendo considerablemente la carga a los informantes.

Adicionalmente, también deberíamos ser conscientes de que la gestión eficiente de este ecosistema de datos implica maximizar el valor añadido que generan. En esta línea, se debería proporcionar un marco de reutilización de los datos, un marco en el que los usuarios tengan acceso a la información necesaria para realizar sus trabajos de investigación con fines científicos de interés público, para que puedan ayudar a la sociedad a una mejor y más eficiente toma de decisiones basada en informa-

ción, acceso que deberá realizarse de forma segura y manteniendo en todo momento la privacidad de la información compartida.

Esta evolución de la sociedad, que supone un cambio y un nuevo paradigma en la estadística oficial se ha visto acelerada en los últimos dos años como consecuencia de la pandemia COVID-19. En este período se ha avanzado lo que, en situación de normalidad hubiese costado posiblemente más de diez años. Las oficinas de estadística se han mostrado resilientes y la ausencia de información estadística tradicional, especialmente en los meses iniciales de la pandemia provocó una necesidad de innovación en fuentes estadísticas para elaborar, desde una perspectiva experimental, información con gran puntualidad, con mayor frecuencia y grado de detalle, basada tanto en información pública como privada que sirviera a los gestores de políticas para su toma de decisiones. Quedaba así demostrado el uso que, desde la perspectiva estadística, podía hacerse de estas nuevas fuentes. Esta demanda se mantiene en la actualidad, por lo que las oficinas de estadística debemos interiorizar esta innovación realizada durante la pandemia para ofrecer a la sociedad unas estadísticas más modernas, rápidas y granulares, manteniendo la solidez metodológica.

En este nuevo ecosistema, el dato es el actor central. Su correcta administración (acceso, metodología de uso, gestión de calidad, protección de la privacidad, reutilización) es clave para conseguir el mayor beneficio que éstos puedan proporcionar a la sociedad. Es importante, por lo tanto, decidir acerca de quién podría ser la institución mejor posicionada para hacerlo dentro de la Administración española. De entre las diferentes opciones, en este artículo se opta por describir la posición de privilegio que tiene el INE frente a otras instituciones para ejercer esa labor.

El Grupo de trabajo sobre data stewardship, creado por el Consejo Superior de Estadística, en el que participaron numerosos expertos y que fue presidido por el Doctor David Ríos, emitió un informe en marzo de 2021 en el que analizaba minuciosamente el papel que debería desempeñar la estadística oficial en las actividades de administración de datos. Buena parte de los aspectos comprendidos en ese informe han inspirado ideas para esta redacción.

La estructura de este artículo es la siguiente. En el primer apartado se describe el cambio de paradigma en la estadística oficial, evaluando la potencialidad de las nuevas fuentes de información para la elaboración de estadística oficial. A continuación, destaca los elementos fundamentales que justifican y avalan la asunción del nuevo rol del INE como administrador público de datos, haciendo hincapié en la experiencia de muchos años en la gestión eficiente de la información estadística y registral así como en el éxito de las incipientes estadísticas experimentales basadas en datos privados. Posteriormente, se ana-

lizan los desafíos que deberá afrontar la institución para desarrollar este papel, entre otros, el refuerzo de la formación de su personal, el diseño de una arquitectura de datos adaptada a esta nueva situación y, finalmente, las implicaciones que tendrá sobre la legislación estadística existente, tanto nacional como europea (mandato, acceso a datos privados...). Finalmente, se concluye con un apartado 5 de conclusiones.

LA ADMINISTRACIÓN DE DATOS ↓

Cambio de paradigma en la estadística oficial ↓

Los sistemas estadísticos elaboran estadísticas oficiales y las difunden a todos los ciudadanos a la vez y sin coste, por lo que, efectivamente, representan un bien público que beneficia a todos. Con dichas estadísticas, todos los actores en la sociedad (hogares, empresas, administraciones...) pueden tomar decisiones basadas en evidencias para determinar su comportamiento en el ámbito socioeconómico.

Tradicionalmente, los institutos de estadística recababan información de los ciudadanos para la elaboración de estadísticas, fundamentalmente a partir de censos y encuestas. La labor de los institutos en el ámbito de la gestión de datos era la de custodiar la información recabada de la sociedad, evitando cualquier brecha de confidencialidad mediante una correcta aplicación del secreto estadístico. Desde el punto de vista de la producción, se elaboraban estadísticas que reflejaban el comportamiento de la sociedad frente a determinados aspectos socioeconómicos.

Más adelante, con el desarrollo de las administraciones públicas, los ciudadanos y las empresas fuimos dejando rastro de nuestra interacción con las administraciones, en forma de datos que se recogían en registros administrativos de gestión y propiedad pública. Desde finales del siglo pasado, las oficinas de estadística incorporamos paulatinamente estas fuentes registrales a la producción de estadísticas oficiales.

Sin embargo, la producción estadística basada en encuestas y registros no es suficiente para cumplir la demanda creciente de estadísticas más puntuales, frecuentes y detalladas, por parte de la sociedad.

Así, la cada vez mayor digitalización de la sociedad implicó que los ciudadanos dejáramos rastro también de nuestras operaciones a través de dispositivos móviles o, genéricamente, a través de nuestra comunicación en internet. Estos datos que masivamente dejamos los ciudadanos directa o indirectamente a través de máquinas (internet de las cosas) se recogen en bases de datos masivas (Big Data) de propiedad privada en la casi totalidad de los casos.

La única posibilidad para que las oficinas de estadística puedan hacer frente a la creciente demanda

de información es incorporar como fuentes, las bases de datos privadas que han surgido de esa transformación digital.

Así, se completa el actual ecosistema de datos, con información estadística, registral y de bases de datos privadas.

El acceso a datos privados

Para el INE, los datos contenidos en registros públicos son fácilmente accesibles para elaborar estadísticas oficiales. Basta con un acuerdo/convenio entre la administración cedente y el INE para poder tener acceso a esa información. Así, el INE puede tener acceso a los ficheros de la Agencia Tributaria o de la Seguridad Social.

Sin embargo, este no es el caso de las bases de datos privadas, donde el acceso no es tan sencillo (B2G). No obstante, el INE ha podido acordar con empresas privadas, en algunos casos, la provisión de información de sus bases de datos para la elaboración de estadísticas, como es el caso de los acuerdos con los operadores de telefonía móvil para estudios de movilidad de la población.

El tema del acceso a fuentes de datos no tradicionales (Big Data, redes sociales, internet y otras fuentes privadas) ya se había destacado en el Plan Estadístico Nacional 2017-2020. No obstante y más recientemente, entre las líneas estratégicas del Plan Estadístico Nacional 2021-2024 se destacaba la recomendación de *realizar un análisis de la normativa vigente y estudiar una modificación, si fuera necesario, que garantice la cesión a los organismos estadísticos oficiales de los datos de las grandes bases de datos privadas para la realización de estadísticas de interés público incluidas en el Plan Estadístico Nacional sin que ello afectase a la explotación de esta información por las empresas poseedoras de la misma.*

Los aspectos jurídicos que afectan a la legislación nacional se tratan en el apartado 4 del artículo. Ellos, sin duda, ayudarán a mejorar el acceso a datos privados pero, no cabe duda que estas referencias a la legislación nacional deben complementarse con un acto a escala europea que proporcione un soporte más robusto.

La nueva Ley de Datos Europea (European Data Act) proporcionará apoyo legislativo sostenible a los institutos de estadística para acceder a los datos del sector privado, siempre que sean necesarios para el interés público y, concretamente, para el desarrollo, producción y difusión de estadísticas oficiales.

La realidad es que no todas las empresas privadas que poseen datos están dispuestas a colaborar para poner sus datos a disposición del interés público. Usualmente, las oficinas de estadística, especialmente durante la pandemia, han acordado con algunas de ellas ciertos niveles de colaboración que

no son muy sostenibles en el tiempo. Esta inseguridad en la disponibilidad continua de la información ha impedido o, al menos, ha retrasado la realización de inversiones en infraestructuras tecnológicas en las oficinas de estadística que permitiera la reutilización eficiente de esa información en pro del interés público.

No obstante, el INE, al igual que otras oficinas de estadística ha demostrado en los últimos dos años que es capaz de producir estadísticas experimentales con datos procedentes de fuentes privadas (algunos ejemplos se muestran en el apartado 3) que se han mostrado tremendamente útiles para los usuarios. El beneficio público ha superado con creces el coste incurrido y la carga para los tenedores de datos privados no ha sido desproporcionada.

La Unión Europea ha reaccionado a esta demanda de información privada, proponiendo una Ley de Datos Europea. Asimismo, el Sistema Estadístico Europeo (2021) se ha posicionado recientemente sobre dicho proyecto de Ley y sobre la necesidad de acceder a información de bases de datos privadas para producir nuevas estadísticas oficiales más rápidas y detalladas.

En concreto, para que el acceso a los datos privados sea sostenible en beneficio del interés público, la Ley de Datos Europea debería establecer los derechos, deberes, obligaciones y límites de todas las partes, tenedores de datos privados, oficinas de estadística, investigadores, elaboradores de políticas, otros agentes sociales y económicos y, finalmente, los ciudadanos. Naturalmente, en el ámbito de la estadística, el Reglamento de Estadísticas Europeas debe ser un marco de referencia, que equipare las características asociadas a los datos procedentes de fuentes privadas con las relativas a las fuentes tradicionales y registrales. Todo ello en aras de fortalecer la integridad de los sistemas estadísticos de los países.

Tres elementos muy importantes:

Primero. Las salvaguardas de confidencialidad de las fuentes tradicionales deben extenderse a las fuentes de datos privados, debiendo reconocerse la responsabilidad en la gestión de la privacidad, tanto de los tenedores de datos como de los elaboradores de estadísticas y posibles terceras partes que puedan utilizar esos datos. Todo ello en aras de fomentar la transparencia y la confianza en la producción estadística y en el uso de los datos privados en beneficio del interés público.

Segundo. Se debe reconocer que los datos privados podrían necesitar un tratamiento diferente a los datos de las fuentes tradicionales o registrales. Así, no será posible en muchos casos acceder al microdato como fuente, sino que el tenedor de los datos necesitará procesar esta información básica y elaborar una suerte de datos intermedios que serían proporcionados para la producción de estadísticas.

ticas. Además de perseguir una protección de la confidencialidad, este hecho se justifica porque la información disponible por el tenedor privado puede tener un formato difícil de gestionar y de reutilizar por parte de los estadísticos o terceras partes (imaginemos por ejemplo el caso de los operadores de telefonía móvil, cuyos microdatos corresponden a posicionamientos de terminales móviles recogidos por las diferentes antenas del operador). Se entiende entonces que el tenedor privado debe realizar algún proceso de transformación del dato dentro de sus oficinas para proporcionar información fácilmente reutilizable, dado que dispone del know-how y de los recursos apropiados para ello.

Tercero. Los datos privados que se proporcionen para el interés público no pueden tener un precio, en sintonía con las reglas aplicables a los datos de fuentes tradicionales o registrales. Las oficinas de estadística y terceras partes no van a entrar en ningún campo de competición empresarial con el tenedor de los datos ni revelar datos a la competencia. Sin embargo, deben reconocerse incentivos y mecanismos de compensación de las inversiones y costes marginales en los que el tenedor privado haya incurrido para el procesamiento de los microdatos que dé lugar a los datos suministrados a la oficina de estadística.

Es muy importante también que las empresas privadas que disponen de estos datos proporcionen toda la metainformación necesaria para que los datos disponibles puedan interpretarse correctamente por todos los usuarios.

En estas condiciones, la Ley europea y las posibles modificaciones que se hagan a la Ley española pueden establecer un marco de referencia normativo para garantizar el acceso a fuentes de datos privadas tanto para la elaboración de estadísticas como para su posible reutilización para el interés público. Será importante para ello establecer nodos de interoperabilidad que faciliten la disponibilidad de esta información a todas las partes. El INE está en una posición privilegiada para ser ese nodo de interoperabilidad.

Funciones del administrador de datos ↓

En la actualidad, los datos han pasado de ser un medio a ser una fuente de actividad económica, pudiendo decirse que son el nuevo oro del siglo XXI. Las empresas consideran como un activo valioso, los datos de que disponen, sabiendo que pueden obtener una buena rentabilidad a través de transacciones en modelos B2B (3). Por su parte, las administraciones valoran los datos que están incorporados en sus registros, no tanto por el fin administrativo con el que fueron recabados de los ciudadanos, sino por la posible reutilización que pueda hacerse de ellos.

Estos datos, tanto los recogidos en registros administrativos como en fuentes de datos privadas, tienen

una naturaleza diferente a los que tradicionalmente se recogían en los institutos de estadística. Así, la visión del gestor de datos ha cambiado y ha pasado de, simplemente, custodiarlos, a la de maximizar el valor añadido que pueden generar aquellos, compartiéndolos con aquellos agentes que puedan utilizarlos en beneficio de la sociedad, facilitando la toma de decisiones basada en evidencias, preservando la confidencialidad de la información.

Por tanto, en la actualidad, los datos deben verse como un bien público que pertenece a la sociedad democrática en su conjunto. No deben verse como una propiedad privada.

En esta nueva forma de gestión surge un nuevo rol que es el de administrador de datos o data steward. El administrador debe tener como principal objetivo garantizar que los datos estén accesibles a aquellos usuarios que puedan generar valor añadido para la sociedad y debe preocuparse asimismo de que esos datos puedan ser utilizados de todas las formas posibles, siempre en un contexto de mantenimiento de la confidencialidad y seguridad.

La accesibilidad de la información puede realizarse a través de la difusión de ficheros de datos abiertos y anonimizados o bien a través de facilitar el acceso a datos confidenciales (de información no directa) en un contexto seguro y que preserve su privacidad.

Las atribuciones de un administrador de datos van más allá de la pura gestión de calidad o gobernanza de los datos. Las aspiraciones de un administrador de datos deben centrarse en tres metas (Verhulst, 2018). Por un lado, debe colaborar activamente para poner a disposición de los usuarios, todos los datos que puedan servir al bien público. Además, debe proteger los datos, administrándolos de forma ética evitando daños a la privacidad. Finalmente, debe actuar proactivamente para identificar los agentes que estén en mejor posición de generar valor a través de la información contenida en los datos privados.

El mismo autor identifica cinco funciones esenciales de un administrador de datos para conseguir esas metas:

1. Desarrollar una actitud proactiva para cooperar con socios potenciales que puedan proporcionar datos y así como con potenciales beneficiarios de esa información
2. Perseguir la coordinación interna y la implicación de su personal
3. Realizar una auditoría ética de datos y evaluar su valor y riesgo
4. Difundir y comunicar (compartir) los resultados obtenidos de las colaboraciones
5. Promover las colaboraciones para garantizar la sostenibilidad

Si bien estas funciones parecen estar pensadas para que sean desarrolladas por una persona dentro de la estructura organizativa de una empresa, ello no impide que pueda ser una institución la encargada de llevarlas a cabo. Efectivamente, en el campo de la Administración Pública, como se ha descrito, son varios los tipos de información que se recogen de los informantes/registros y asimismo hay una gran variedad de usos que pueden tener en un marco de reutilización. Por ello, puede ser más eficiente que estas funciones puedan ser desarrolladas por una institución pública o más concretamente, por una división o subestructura de dicha institución.

En el ámbito de la Administración Pública española, existen varios modelos que podrían discutirse para designar su administrador de datos. La Oficina del Dato, creada dentro del Ministerio de Asuntos Económicos y Transformación Digital se crea con el objetivo de dirigir la Estrategia Nacional del Dato y, en ese papel, podría pensarse que es un candidato a ser el administrador de datos de la Administración Pública.

Pero probablemente, la oficina de estadística, el INE, es el organismo que podría estar mejor posicionado para asumir este rol. Es el punto central de recepción de datos estadísticos y registrales, debe ser el nodo central de recepción de datos de fuentes privadas, como se ha discutido arriba y podría ser, en un modelo de gestión compartida, el nodo de comunicación con la comunidad investigadora para proporcionar acceso con fines científicos de interés público.

La legislación europea en materia estadística otorga al INE, podríamos decir, el «monopolio» para ejercer la función estadística pública. Así, toda la información recabada para fines estadísticos únicamente puede recogerse en las bases de datos de dicha institución y administrarse desde aquella.

Sin embargo, no está claro que la función estadística alcance la administración de datos privados de empresas que puedan, potencialmente, beneficiar a otras empresas (B2B). Ese papel podría ser desarrollado por la Oficina del Dato.

El apartado siguiente de este artículo destaca las fortalezas que tiene el INE para desarrollar las tareas de administrador de datos, debido fundamentalmente a la larga experiencia en la gestión de datos y en la provisión a usuarios de información para uso científico. Es decir, el INE no partiría de cero para ejecutar las tareas de data steward. No obstante, la asunción de este rol no sería inmediata, dado que se deberían afrontar ciertos retos que se describen más adelante.

FORTALEZAS DEL INE PARA DESARROLLAR EL PAPEL DE DATA STEWARD DE LA ADMINISTRACIÓN ↓

La asignación al INE del rol de administrador de datos sería realmente eficiente dado que el INE

ya cuenta con una serie de competencias legales, instrumentos, herramientas y gran experiencia en su aplicación, que son necesarios para ejecutar debidamente este papel. No obstante, para su completa implementación, sería necesario llevar a cabo algunas modificaciones en los cometidos que le asigna la LFEP así como en su estructura organizativa, aspectos que se desarrollan en el siguiente apartado del artículo.

Debe tenerse en cuenta, por ejemplo, que el INE pone a disposición de todos los usuarios en su web los ficheros de microdatos de 45 estadísticas, que en 2019 generaron 54.000 accesos y más de 113.000 descargas. Estos ficheros se hacen públicos tras un proceso de anonimización que impide la identificación de las unidades individuales (ciudadanos o empresas).

Las principales garantías o avales que presenta el INE para poder desarrollar su papel de administrador de datos de la Administración Española, aparte de su independencia y su robustez en la elaboración y aplicación de metodologías están ligados con su experiencia en el uso de registros administrativos, tanto desde una perspectiva colaborativa con otros organismos, como metodológica o práctica, su experiencia en la propia elaboración de registros, en el uso de fuentes privadas para la elaboración de estadísticas o en el aseguramiento de la confidencialidad y privacidad de la información que maneja.

Experiencia en el uso de registros administrativos ↓

Desde hace más de veinte años, el INE viene gestionando convenios de colaboración con organismos de la Administración Pública española recibiendo registros administrativos para su uso en la elaboración de estadísticas oficiales.

De esta manera, a través de convenios de colaboración con la Agencia Estatal de Administración Tributaria, la Seguridad Social u otros organismos de la Administración General del Estado, el INE dispone de más de cien registros administrativos. Pero también cabe destacar los convenios realizados con las Haciendas Forales del País Vasco y Navarra, que permiten la recepción de registros de estas comunidades autónomas.

Así, en lo que se refiere el uso de registros administrativos, el INE goza de gran experiencia tanto para su explotación estadística directa, como para realizar fusiones entre ellos y con otras bases de datos de operaciones estadísticas. Dicha experiencia abarca tanto los desarrollos metodológicos necesarios para un buen aprovechamiento de los registros como la utilización y gestión prácticas de los mismos.

Entre los desarrollos metodológicos cabe destacar la catalogación o inventario de registros de la Administración, incluyendo su metadato, para permitir su reutilización eficiente y analizar su potencialidad

para el diseño de nuevas operaciones estadísticas o para la realización de estudios.

Por lo que a los aspectos prácticos se refiere, hay estadísticas de gran tradición en el INE, como las Estadísticas del Movimiento Natural de la Población (nacimientos, defunciones, matrimonios) que se elaboran desde hace tiempo a partir de la explotación de los Registros Civiles, esenciales para la estimación de las cifras de población en España. Asimismo, el Directorio Central de Empresas (DIRCE) que combina la información de doce registros administrativos y operaciones estadísticas diferentes y que es la columna vertebral de las operaciones estructurales y coyunturales de estadísticas a empresas. Estadística de Hipotecas, Sociedades Mercantiles, Ejecuciones Hipotecarias, de Condenados adultos y menores, de Violencia Doméstica y de Género o el Índice de Precios de la Vivienda en Propiedad son otros ejemplos de la producción estadística del INE basada en la gestión y explotación de registros administrativos.

No obstante, en los últimos años, el INE ha desarrollado dos proyectos que han supuesto un hito sin precedentes en la producción estadística basada en la gestión y uso de información registral.

Por un lado, los Censos de Población y Viviendas 2021 que, por vez primera se van a realizar cruzando información de decenas de registros administrativos (tributarios, de la Seguridad Social, sanitarios, educativos...) utilizando como núcleo central el Padrón Continuo, para ofrecer la fotografía de la población española en ese momento, sin tener que recurrir a realizar ninguna operación de campo como se realizaba en los censos tradicionales. Esta forma de elaboración, además de reducir enormemente la carga de los ciudadanos y el coste de elaboración de la operación, pone al INE en la punta de lanza innovadora de la producción estadística. Además, un censo basado en registros podrá repetirse con una mayor frecuencia y no será necesario esperar diez años para replicar la foto social de la población española (de hecho se ha dispuesto de ficheros pre-censales en los años 2017 y posteriores). Finalmente, la información contenida en esta infraestructura podrá reutilizarse para otras operaciones y para análisis y estudios que puedan realizarse desde la óptica de la investigación.

De hecho, una explotación que se hizo de esta información censal, conjuntamente con la información de los ficheros de la Agencia Tributaria fue la construcción experimental de un atlas de distribución de renta de la población española, en la que se pudo extraer información de renta personal y de los hogares con un detalle de más de 3.000 áreas o territorios.

Por otro lado, el Censo Agrario 2020 que, también por vez primera, se va a basar de una forma mayoritaria en el aprovechamiento de registros administrativos del Ministerio de Agricultura, Pesca y Alimentación, con la consiguiente reducción de carga a los

ciudadanos y coste de la operación y que lo convierten en el primer censo agrario de un país de la Unión Europea elaborado con registros. Igualmente, esta infraestructura podrá ser utilizada para facilitar el análisis a los investigadores.

Experiencia en la formación de registros administrativos ↓

El INE no solamente reutiliza la información de registros administrativos de otros organismos de la Administración para su producción estadística oficial, sino que también elabora, amparado en la legislación vigente, sus propios registros administrativos.

Así, por un lado, el INE tiene asignadas las funciones de formación del Censo Electoral, de acuerdo con la Ley Orgánica de Régimen Electoral General (LO-REG), registro en el que se incluye a todos los ciudadanos que tienen derecho al voto en alguno de los procesos electorales que se convoquen.

Por otro lado, la Ley de Bases de Régimen Local asigna al INE la función de coordinación de los Padrones Municipales de más de ocho mil municipios, en virtud de la cual dispone de una base nacional de gestión denominada Padrón Continuo, infraestructura de gran valor añadido para la elaboración de estadísticas, por ejemplo, como se ha citado anteriormente, para los Censos de Población y Viviendas.

Finalmente, el INE dispone de otro registro administrativo denominado Padrón de Españoles Residentes en el Extranjero (PERE) que contiene las inscripciones de las personas que, gozando de la nacionalidad española, viven habitualmente fuera de España, sea o no ésta su única nacionalidad.

Experiencia en el uso de datos procedentes de fuentes privadas ↓

Como se ha descrito en el apartado 2 de este artículo, el acceso a fuentes de datos privadas es indispensable para desarrollar la Estadística Oficial del este siglo.

No obstante, hay información de fuentes privadas que ya están implantadas en algunas operaciones del INE y otras que se han utilizado para elaborar estadísticas experimentales que han tenido una especial importancia durante la pandemia y que en un futuro podrían acabar convirtiéndose en estadísticas oficiales.

Por ejemplo, a través de acuerdos entre el INE y algunas cadenas comerciales, se ha podido iniciar la inclusión de los denominados *scanner data* en la elaboración del IPC. Así, las empresas facilitan esta información al INE de forma regular y digitalizada y, a cambio, el INE no tiene que enviar a los agentes a los establecimientos para realizar la toma de precios.

Durante la pandemia, el INE impulsó la realización de estadísticas experimentales utilizando fuentes privadas de datos.

Así, cuando se inició el confinamiento, era crucial disponer de información relevante a la movilidad de los ciudadanos. Ello permitió recuperar una operación estadística que se había desarrollado de forma puntual en 2019 anexa a los Censos de Población y Viviendas en la que se describía la movilidad de la población española basada en los datos registrados por las principales Operadoras de Telefonía Móvil en lo que se refería al posicionamiento de los terminales móviles en más de 3.000 áreas territoriales. De esta manera, a partir de finales de marzo de 2020 se pudo acordar con estas compañías la provisión de datos agregados no confidenciales para poder elaborar un indicador diario (posteriormente semanal) de movimiento de la población residente.

Posteriormente, esta misma tecnología se está utilizando en el ámbito de las estadísticas de turismo, para analizar los viajes de la población residente así como la de los no residentes.

Igualmente, datos de operaciones financieras de los ciudadanos basadas en el gasto con tarjetas de crédito o débito han propiciado asimismo operaciones experimentales desarrolladas en el INE.

Aseguramiento de la confidencialidad en la cesión de datos ↓

Además de la producción de estadística oficial, el INE desarrolla una actividad importante en lo que se refiere a la cesión de información a usuarios. Esta cesión está amparada en la LFEP y, asimismo, en el Reglamento europeo de Estadística y presta especial importancia a la seguridad y confidencialidad que establece la Ley de Protección de Datos Personales.

El órgano interno que se encarga de la vigilancia de la confidencialidad de la información que se cede a usuarios es el denominado *Comité del Secreto Estadístico*.

Así, este comité tiene por función informar acerca de las peticiones que se reciben de acceso a microdatos confidenciales y establecer el modo de acceso a los mismos, aplicando las reglas establecidas por el propio comité. Además, asesora sobre los métodos de anonimización que se deben utilizar, siempre bajo el criterio de que la pérdida de información sea mínima una vez conseguido el propósito de protección de los datos individuales. La aplicación de estos métodos a uno u otro tipo de datos, ya sean estadísticos o registrales, no presenta excesivas diferencias y su eficiencia está avallada por el desarrollo que ha adquirido a lo largo de estos años de existencia.

Desde una perspectiva práctica, el INE facilita información a medida a usuarios, garantizando la confidencialidad, a través de tres canales.

En primer lugar, a través de su propio personal, realiza las explotaciones necesarias para obtener la información demandada por el usuario, anonimizando antes de ser proporcionada a aquel.

En segundo lugar, a través de un centro seguro. En dicho centro, los usuarios pueden llevar a cabo las explotaciones que requieran, siendo auditados exhaustivamente los resultados obtenidos, para garantizar la protección de los datos personales y de aquellos otros sujetos al secreto estadístico. Así, los investigadores pueden acceder a información confidencial de identificación indirecta, pero con la garantía de que los resultados no comprometen la información individual.

Finalmente, a través de acceso remoto. Esta solución permite ampliar la capacidad de acceso a información sensible garantizando la seguridad, mediante el uso de certificados digitales que permiten asegurar la identidad de los investigadores que acceden al sistema.

RETOS QUE DEBE AFRONTAR EL INE PARA DESARROLLAR SU PAPEL COMO DATA STEWARD DE LA ADMINISTRACIÓN ↓

Como se ha destacado en el apartado anterior, son muchas las fortalezas que tiene el INE para desarrollar el papel de administrador de datos. Esas fortalezas le sitúan en una posición privilegiada para ese cometido. No obstante, en ese camino deberá afrontar diferentes desafíos, ligados tanto a aspectos legales, como de recursos, metodológicos y de infraestructura operativa.

Refuerzo del marco legal de la función estadística ↓

La actual LFEP data del año 1989. Las bases de datos masivas que, actualmente, disponen algunas empresas, no existían en esa época, por lo que era impensable que la Ley hiciera mención a ellas y a la posible disponibilidad de dicha información por parte del INE.

La ausencia de reglas y procedimientos claros para el acceso y uso de datos en la legislación nacional, se constituye como una barrera para aprovechar plenamente el potencial de los datos del sector privado con fines de interés público.

Aunque la LFEP obliga a las empresas a facilitar información para la elaboración de estadística oficial, sus provisiones están ligadas al marco de las encuestas a empresas y no tanto a proporcionar bases de datos masivas.

Por ello, el amparo que la actual Ley otorga a un intercambio de datos masivo B2G es bastante débil y debería reforzarse en una posible modificación del acto legal. Esta modificación tendría que tener en cuenta dos componentes.

En primer lugar, la disponibilidad por parte del INE de bases de datos masivas para la elaboración de estadística oficial, reconociendo la particularidad de este tipo de datos y otorgando una mayor seguridad jurídica a las empresas y al INE.

En efecto, para este cometido sería necesaria la modificación del artículo 10 de la LFEP para permitir al INE y, en general a los servicios estadísticos, la posibilidad de recabar de las empresas aquellos datos o informaciones que estén almacenados en cualquiera de sus bases de datos, para la realización de estadísticas oficiales. De esta manera quedaría más clara en el texto legal que la información que se desea disponer de las empresas comprende también sus bases de datos masivas.

De todos modos, como se ha citado en el apartado anterior, el INE ha desarrollado varias operaciones experimentales utilizando información procedente de bases de datos masivas privadas, acordando con aquellas los términos de referencia de la disponibilidad de la información. Esta modificación legal supondría la creación de un marco normativo para las relaciones de intercambio de información B2G para asegurar su disponibilidad futura.

En segundo lugar, pensando en el rol del INE como *data steward* de la Administración, deberían ampliarse las funciones del INE para permitir la posibilidad de cesión de los datos en poder del INE a los investigadores para usos científicos de interés público y no solo para usos estadísticos.

La LFEP solo contempla la función estadística del INE y, en ese sentido, no regula la cesión de datos para uso científico y, por lo tanto, no establece provisiones de cómo realizar dicha cesión.

Sería necesario, por un lado, revisar el artículo 26 de la LFEP, para incluir como función del INE, la puesta a disposición de toda la información recogida por el organismo y la procedente de cualquier entidad de la Administración, a instituciones de investigación, estudios o análisis, con fines científicos de interés público. Por otro, deberían incluirse en su artículo 15 las disposiciones reglamentarias generales que regulen esta puesta a disposición, es decir, los requisitos que debe cumplir la institución solicitante, el acceso que se permitirá (datos confidenciales de identificación no directa) o el respeto de la confidencialidad y el secreto estadístico en los mismos términos que el personal del INE, dejando los aspectos más de detalle para que puedan ser regulados por un acto legal de inferior rango.

En el caso de acceso a registros administrativos, es preciso tener en cuenta que la propia norma que crea o desarrolla cada registro puede establecer restricciones en lo que a su uso y cesión se refiere. Esta salvaguarda deberá aparecer también en el texto legal reforzado.

Como elemento que refuerza la seguridad jurídica de la cesión de datos para estos fines no estadís-

ticos, debe citarse que aunque el principio de especialidad que recoge el artículo 4 de la LFEP, dice que es exigible a los servicios estadísticos que los datos recogidos para la elaboración de estadísticas se destinen para fines estadísticos, este hecho debe producirse sin perjuicio de las cesiones ulteriores que puedan realizarse para fines científicos, de conformidad con lo previsto en el Reglamento Europeo de Protección de Datos. Es decir, el tratamiento para uso científico de los datos recogidos para fines estadísticos es compatible con éste.

Refuerzo de recursos humanos y técnicos ▾

La nueva actividad del INE como administrador de datos implica un incremento de tareas en diferentes campos. Por un lado, en el campo estadístico, centrado en la creación de metodologías, estándares, identificadores, normas de anonimización, cruce de fuentes de datos... y, por otro, el campo tecnológico e informático, en el que será necesario el desarrollo de plataformas de comunicación robusta y seguras o la gestión de bases de datos y, finalmente, los retos metodológicos asociados al intercambio B2G de datos masivos.

El INE dispone de recursos humanos que realizan eficientemente las tareas asociadas al campo estadístico y tecnológico. Por ello, podría decirse que la inversión inicial que supone asumir las tareas de un administrador de datos en el INE está ya realizada. Aun así, es previsible un aumento significativo de este tipo de tareas (por ejemplo, el metadotado de todos los ficheros que se vayan a poner a disposición de los investigadores) por lo que parece necesario que la plantilla de recursos humanos del INE debería reforzarse.

No obstante, el reto que supondría tomar el rol de administrador de datos debería estar apoyado en dos aspectos de vital importancia.

En primer lugar, de cara a una gestión eficiente de los intercambios B2G y al tratamiento de bases de datos masivas, el INE debería disponer de recursos capacitados en la gestión de Big Data. Este tipo de datos, de alta frecuencia y granularidad tiene un tratamiento metodológico y práctico diferente al de los datos utilizados en la elaboración de las estadísticas tradicionales. Así, aspectos como la agregación, la construcción de indicadores, la regresión entre variables, el ajuste estacional de series de alta frecuencia, el desarrollo de tecnologías de difusión en mapas... son elementos próximos a lo que se denomina ciencia de datos. La disponibilidad de recursos capacitados en estos temas es necesaria para un buen aprovechamiento de la información disponible.

En segundo lugar, es indispensable una arquitectura de IT suficientemente robusta que permita almacenar los datos en un *data lake* y gestionarlos de la manera más eficiente, en diferentes niveles, garan-

tizando la seguridad y la confidencialidad de la información y que muestre una capacidad de operar con otros nodos.

Esta infraestructura debe automatizar la interoperabilidad con los proveedores de información internos y externos al INE y debe mostrarse, asimismo, como un nodo que articule el intercambio de datos.

En efecto, esta arquitectura es crucial para el éxito de un sistema de acceso a investigadores a información en poder de la Administración, dado que permitiría interactuar directamente al usuario con el sistema para realizar los accesos a los que hubiese sido autorizado previamente, de forma autónoma, en un entorno seguro.

Además de la inversión inicial en la construcción de la infraestructura tecnológica, la gestión de aquella requerirá asimismo de un reforzamiento del personal técnico destinado a estas tareas en el INE.

Proporcionar datos a la comunidad investigadora con fines científicos de interés público. Un modelo de consorcio

Como se ha discutido a lo largo del artículo, el data steward de la Administración, no solo tiene que ser el nodo de recepción de información estadística, registral o de fuentes privadas. Debe maximizar el valor añadido de los datos que administra, poniéndolos a disposición de la investigación con fines científicos de interés público.

Esta función, reclamada por ejemplo en AIREF (2020) para datos registrales, con el fin de realizar la evaluación de políticas públicas basada en datos, podría ser llevada a cabo por diferentes organismos. Almunia y Rey-Biel (2020) describen diferentes modelos que podrían aplicarse.

Todos los modelos confluyen en que el INE debería tener una participación central para desarrollar esa función.

En este sentido, AIREF reconoce que *«el Gobierno debería designar una institución especialmente dedicada a facilitar el acceso a datos registrales personales producidos por todas las administraciones públicas, que persiga la estandarización en los protocolos de acceso, se responsabilice de la difusión de cada conjunto de datos protegiendo la privacidad personal y tenga en cuenta el interés público del uso de los datos para tomar las decisiones sobre su acceso. El INE podría ser esta institución al estar legalmente amparada en el ámbito español y europeo y disponer de infraestructuras y capacidad técnica para desarrollar esta misión o alternativamente se podría designar otra institución»*.

Es cierto que el INE es la institución que está mejor preparada para desarrollar esa nueva función y ser el data steward de la Administración. No obstante, la gobernanza para ejecutarla debería depender no

sólo del INE sino de otros organismos responsables de información registral pública.

En este sentido, el pasado mes de abril de 2021, el INE, la Agencia Tributaria, el Ministerio de Seguridad Social, el Banco de España y, posteriormente, el Servicio Público de Empleo Estatal acordaron comenzar a trabajar para diseñar un Sistema para proporcionar datos a investigadores con fines científicos de interés público.

Son muchos los aspectos que hay que delinear para que este sistema funcione correctamente.

En primer lugar, resolver los retos jurídicos que ya se han discutido en este mismo apartado del artículo. La legislación en materia estadística podría ser el principal paraguas que diera sentido a esta estructura.

Segundo; es necesario establecer los elementos de la gobernanza del Sistema, los órganos donde se tomaran todas las decisiones sobre la viabilidad o no de las solicitudes que los investigadores realicen. El INE sería el coordinador pero además deberían constituirse órganos (¿colegiados?) en los que participaran todas las instituciones implicadas, acordando un mandato y unas reglas de procedimiento.

Tercero; un elemento muy importante es determinar el catálogo de información que se pondría a disposición de los investigadores, los ficheros y bases de datos que podrían utilizar en sus trabajos. En este punto, es crucial el metadatado de toda la información, no solamente de las variables que contienen las bases de datos que se ponen a disposición, sino también las características de aquellas.

Finalmente, la operativa, que comprendería varios aspectos. Por un lado, la gestión de las solicitudes (ventanilla de acceso). También la forma en la que el investigador va a interactuar con el Sistema. Si bien es deseable que a medio plazo, el acceso sea remoto y que el investigador solo tenga acceso a aquella información a la que ha sido autorizado a acceder, debemos ser conscientes que a corto plazo, esto puede no ser posible y que el funcionamiento podría tener una componente más manual. Se utilizará para ello el *data lake* del INE, siendo importante en este punto los elementos de Seguridad o la gestión de permisos de acceso.

Este Sistema debería estar gestionando las primeras solicitudes de investigadores en la primera mitad de 2022.

CONCLUSIONES

El INE está situado en una posición de privilegio para adquirir el rol de data steward de la Administración Española, al menos en lo que a la administración de datos estadísticos se refiere. Dispone de gran experiencia en la gestión de datos estadísticos y en el uso

de registros administrativos, está ganando cada día más experiencia en la elaboración de estadísticas con datos de fuentes privadas y es plenamente capaz de administrar la confidencialidad y privacidad de los datos que detenta, tanto para elaborar estadísticas como en el ámbito de la cesión de datos a terceros para fines estadísticos en entornos seguros. Es decir, el INE no parte de cero en el desarrollo de esa nueva función.

Tanto la LFEP como el Reglamento de Estadísticas europeas le otorgan capacidad para solicitar al resto de la Administración, toda aquella información contenida en registros con el fin de elaborar estadísticas oficiales, permitiéndole incluso actuar sobre dichos registros con el fin de adaptarlos para reducir la carga al informante y perseguir asimismo la sostenibilidad de la metodología de las operaciones estadísticas. De esta manera, el INE tiene un papel central en la gestión de datos G2G.

No obstante, debe afrontar ciertos desafíos en el presente y futuro próximo. Uno de ellos es el ligado al intercambio de datos B2G. El INE debe fomentar más si cabe la utilización de datos de fuentes privadas para realizar operaciones estadísticas. Los retos son, por un lado, conseguir acceso a las fuentes privadas de datos masivos, persiguiendo la sostenibilidad de dicho acceso (reto ligado a la entrada en vigor de la Ley Europea de Datos y a la modificación de la LFEP). Por otro lado, invertir en tecnología y en recursos humanos de carácter técnico (científicos de datos) para gestionar eficientemente todo el volumen de información privada que reciba. De todos modos, debe destacarse que el INE ha desarrollado operaciones estadísticas experimentales que tuvieron su punto de partida durante la pandemia, por lo que ya dispone de experiencia en este ámbito.

En su rol de administrador de datos, el INE debe colaborar con las empresas que detenten datos cuyo valor sea evidente que puede servir al bien público, si se utilizan de forma responsable, a través de acuerdos de colaboración.

Finalmente, se deben perseguir las actuaciones jurídicas necesarias para que los datos que detente el INE y demás organismos de la Administración puedan ser utilizados, además de para realizar estadística oficial, para la investigación científica de interés público, dando a los datos una nueva vida o varias nuevas vidas, maximizando el valor añadido que tienen, y actuando en beneficio de la sociedad. El consorcio de cinco instituciones públicas para suministrar datos a investigadores con fines científicos de interés público es un gran paso adelante en esta dirección.

Los datos son un bien público que nos pertenece a todos. La verdadera sociedad libre y democrática es aquella que puede disponer de toda la información para que pueda realizar su toma de decisiones basada en datos.

NOTAS ↓

- [1] G2G: Government to Government. Flujo de datos de la Administración para la propia Administración.
- [2] B2G: Business to Government: Flujo de datos de empresas privadas a la Administración.
- [3] B2B: Business to business: Flujo de datos de empresas privadas a otras empresas.

REFERENCIAS ↓

- AIREF (2020). Opinión para una estrategia de acceso a datos administrativos. Opinión 1/20. Disponible en <https://www.airef.es/es/opiniones/estrategia-acceso-datos-administrativos/>
- Almunia, M. y Rey-Biel, P. (2020) «La Gestión de los Datos Administrativos en España: Diagnóstico y Retos de Futuro» Esade
- Consejo Superior de Estadística (2021) «El papel de la estadística oficial en la administración y gestión de datos» Informe del grupo de trabajo sobre data stewardship
- Cristóbal, A. (2021) «Un sistema para permitir a investigadores acceso a datos confidenciales de identificación indirecta con fines científicos de interés público» Revista Presupuesto y Gasto Público. Pendiente de publicación
- Cristóbal, A. (2021) «Lecciones aprendidas de la crisis provocada por la pandemia. Impacto en la estadística oficial». Revista Información Comercial Española, núm. 923.
- INE, AEAT, MISSM, BdE (2021) «Comunicado institucional: El INE, la Agencia Tributaria, la Seguridad Social y el Banco de España acuerdan comenzar a trabajar conjuntamente en el desarrollo de un sistema de acceso a sus bases de datos con fines científicos de interés público» Disponible en www.ine.es
- Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE relativo a la estadística europea. Disponible en <https://eur-lex.europa.eu>
- Real Decreto 1110/2020, de 15 de diciembre, por el que se aprueba el Plan Estadístico Nacional 2021-2024. Disponible en www.boe.es
- Sistema Estadístico Europeo (2021) «European Statistical System (ESS) position paper on the future Data Act proposal. Access to privately held data is urgently needed for producing new, faster, more detailed official statistics». Disponible en <https://ec.europa.eu/eurostat/web/european-statistical-system/governance-bodies/essc>
- Verthulst, S. (2018) «The three goals and five functions of Data Stewards. Data Stewards: a new Role and Responsibility for an AI and Data Age», Disponible en <https://medium.com/datastewardsnetwork/the-three-goals-and-five-functions-of-data-stewards-60242449f378>