

Dt

DOCUMENTO DE TRABAJO

DEL DATO ADMINISTRATIVO DE FIRMAS AL DATO ESTADÍSTICO: DESAFÍOS A
VENCER EN AMÉRICA LATINA

Octubre 2018

N° 2018/17

DEL DATO ADMINISTRATIVO DE FIRMAS AL DATO ESTADÍSTICO: DESAFÍOS A VENCER EN AMÉRICA LATINA

Rojo Brizuela, A.

DEL DATO ADMINISTRATIVO DE FIRMAS AL DATO ESTADÍSTICO: DESAFÍOS A VENCER EN AMÉRICA LATINA

Rojo Brizuela, A.

CAF – Working paper N° 2018/17

Octubre 2018

RESUMEN

El diseño, la implementación y la evaluación de las políticas públicas para el fomento de la productividad requieren cada vez más información estadística oportuna y de calidad. En los últimos veinte años, y especialmente en países desarrollados, los registros administrativos han contribuido a satisfacer esta demanda creciente de datos. En América Latina, en cambio, esto es mucho más incipiente. Sin embargo, varios países de la región están actualmente empleando registros que capturan la dinámica de la producción, el empleo, el comercio internacional y la innovación como fuentes para la producción de estadísticas útiles para la formulación de políticas de productividad. Este documento analiza las barreras que enfrentan varios países en América Latina para expandir de manera adecuada el uso de estas fuentes administrativas, y lo hace desde una perspectiva práctica.

Pequeñas secciones del texto, menores a dos párrafos, pueden ser citadas sin autorización explícita siempre que se cite el presente documento. Los resultados, interpretaciones y conclusiones expresados en esta publicación son de exclusiva responsabilidad de su(s) autor(es), y de ninguna manera pueden ser atribuidos a CAF, a los miembros de su Directorio Ejecutivo o a los países que ellos representan. CAF no garantiza la exactitud de los datos incluidos en esta publicación y no se hace responsable en ningún aspecto de las consecuencias que resulten de su utilización.

© 2018 Corporación Andina de Fomento

Del dato administrativo de firmas al dato estadístico: Desafíos a vencer en América Latina

Ana Sofia Rojo Brizuela

Octubre 2018

Resumen

El diseño, la implementación y la evaluación de las políticas públicas para el fomento de la productividad requieren cada vez más información estadística oportuna y de calidad. En los últimos veinte años, y especialmente en países desarrollados, los registros administrativos han contribuido a satisfacer esta demanda creciente de datos. En América Latina, en cambio, esto es mucho más incipiente. Sin embargo, varios países de la región están actualmente empleando registros que capturan la dinámica de la producción, el empleo, el comercio internacional y la innovación como fuentes para la producción de estadísticas útiles para la formulación de políticas de productividad. Este documento analiza las barreras que enfrentan varios países en América Latina para expandir de manera adecuada el uso de estas fuentes administrativas, y lo hace desde una perspectiva práctica.

Contenido

Introducción	3
1. Bases de datos administrativas y bases de datos estadísticas.....	5
2. Desafíos y oportunidades del uso estadístico de los registros administrativos para la producción de estadísticas.....	7
3. Descripción de las etapas necesarias para el aprovechamiento de datos administrativos en la producción estadística.....	9
a) Identificación y selección de los registros de interés.....	11
b) Construcción de acuerdos institucionales.....	11
c) Conformación del equipo de trabajo	13
d) Adecuación de los datos administrativos para usos estadísticos	13
e) Diseminación de resultados: estadísticas agregadas y microdatos	15
4. Aspectos técnicos de la adecuación de los datos administrativos para usos estadísticos .	16
A. Evaluación de la calidad de los registros.....	16
a) Análisis de los procesos de gestión que dan origen al registro	17
b) Análisis de la consistencia de los datos administrativos.....	20
B. Corrección de inconsistencias y errores, construcción de sistemas de información integrados y dinámicos.	22
a) Corrección de inconsistencias en cada registro	23
b) Integración de fuentes en sistemas de información.....	26
c) Construcción de sistemas dinámicos: Series longitudinales	27
5. Otros aspectos para enriquecer los sistemas de información: georreferenciación y análisis de ocupaciones	29
A. Georreferenciación de registros	29
B. Clasificador de ocupaciones	30
a) Conceptos que utiliza el clasificador	31
b) ¿Cómo funciona el clasificador?.....	31
d) Uso de los clasificadores de ocupaciones en las instituciones consultadas	32
6. Reflexiones finales.....	34
7. Bibliografía	36
8. Anexo: Fichas de análisis de casos	39

Introducción

El diseño, la gestión y la evaluación de las políticas públicas demandan cada vez más información, la cual muchas veces se obtiene a partir de fuentes de datos complementarias a las estadísticas tradicionales. El estudio de los problemas económicos y sociales, así como la evaluación de las políticas públicas, requieren observar la diversidad y heterogeneidad de distintos fenómenos y de múltiples actores, para lo cual los microdatos¹ son de vital importancia. Asimismo, la necesidad de considerar los aspectos dinámicos de esos fenómenos pone en evidencia la importancia de contar con bases de datos longitudinales apropiadas al estudio del objeto de interés.

En los últimos veinte años, los registros administrativos han contribuido a satisfacer esta demanda creciente de información. El uso de este tipo de datos es una práctica generalizada en los países más desarrollados, a diferencia de América Latina, donde su uso es más reciente.

A pesar de este uso todavía limitado en los países de América Latina, hay algunas áreas en las cuales la región ha avanzado más, empleando registros administrativos como fuente para la producción de estadísticas útiles para la formulación de políticas públicas. Este es el caso de algunas problemáticas sociales, en especial en temas de educación y salud.

Por su parte, en áreas vinculadas con el desarrollo productivo y la dinámica del empleo, los registros son un insumo clave para la construcción de los directorios de empresas y las estadísticas de comercio exterior y empleo, pero su uso en la investigación está menos extendido, por lo que se trata de un ámbito con un enorme potencial para explotar. En la mayoría de los casos, estos datos son aprovechados solo por las agencias productoras de estadísticas, que diseminan sus resultados como estadísticas agregadas. La limitada disponibilidad de microdatos hacia el exterior de las agencias productoras, lleva a que no resulte aún frecuente su empleo en estudios más complejos que analicen problemáticas productivas y laborales, como el caso de evaluaciones de impacto de políticas pro-productividad.

Si bien es clara la oportunidad y el potencial que existe detrás del uso más intensivo de los datos administrativos, también es evidente la necesidad de evaluar su calidad y de corregir sus inconsistencias antes de ser utilizados como fuente de información estadística. Este desafío obedece a que esos grandes volúmenes de datos suelen ser transaccionales², y en su estado puro no necesariamente presentan los atributos de calidad (precisión y confiabilidad, entre otros) que se exige a las operaciones estadísticas tradicionales, como los censos y las encuestas (Eurostat, 2003; CEPAL, 2003; OECD, 2015).

En este contexto, el objetivo de este documento es ofrecer una guía que oriente el trabajo de las agencias estadísticas de América Latina que deseen transformar registros administrativos - de firmas y de empleo- en fuentes para informar a las políticas orientadas a aumentar la productividad. Este documento se ha nutrido tanto de trabajos metodológicos publicados sobre la temática, como de consultas a funcionarios de instituciones que usan estas fuentes para producir estadísticas, especialmente los participantes del taller “Herramientas prácticas para la

¹ Los microdatos son datos asociados a unidades decisorias, ya sean personas físicas o jurídicas.

² Los datos transaccionales son los que respaldan intercambios administrativos, en este caso, entre individuos y el Estado.

apertura y uso seguro de microdatos de firmas”, organizado por la Dirección de Investigaciones Socioeconómicas de CAF, que tuvo lugar en Buenos Aires el 8 de junio de 2017³.

El documento está organizado en dos grandes bloques. El primero está compuesto de tres secciones y aborda la diferencia conceptual que existe entre los datos administrativos y los estadísticos; mientras que el segundo bloque está compuesto de dos secciones, donde se profundiza en los aspectos metodológicos que deben ser considerados para la transformación de los datos administrativos en estadísticos, enriqueciendo la descripción con numerosos ejemplos extraídos de experiencias de la región.

De esta manera, en la primera sección se muestra la necesidad de realizar transformaciones sobre los registros antes de utilizarlos como fuente en las investigaciones sobre las políticas públicas; en la segunda sección, se discuten los desafíos y las oportunidades que ofrece este tipo de fuentes y; en la tercera sección, se presenta una descripción sencilla del proceso de gestión, así como también de los actores que deben intervenir en todas las etapas de transformación de los datos, destacándose la complejidad institucional que tal desarrollo requiere.

Por su parte, la cuarta sección muestra los principales aspectos técnicos necesarios para la adecuación de los datos, es decir: la evaluación de la calidad, las correcciones de errores que se realizan con mayor frecuencia y la construcción de sistemas integrados de información y de paneles longitudinales. La quinta sección presenta dos aspectos específicos que ofrecen grandes oportunidades para un mejor aprovechamiento de los registros: los clasificadores de ocupaciones y la geo-referenciación de los datos.

Luego, se presentan algunas consideraciones finales.

³ Fueron consultados: tres institutos de estadísticas (DANE de Colombia, INEI de Perú, INEC de Ecuador); dos áreas de estudios de organismos de recaudación que producen estadísticas a partir de los registros administrativos que ellas producen (el Banco de Previsión Social y el Servicio de Impuestos Internos de Uruguay); y, finalmente, dos áreas de estudios vinculadas a ministerios que desarrollan estadísticas y estudios a partir de registros de otras instituciones (IPEA de Brasil, y el Observatorio de Empleo y Dinámica Empresarial del Ministerio de Trabajo, Empleo y Seguridad Social de Argentina).

1. Bases de datos administrativas y bases de datos estadísticas

¿Qué diferencia existe entre los datos administrativos y los estadísticos? ¿Por qué se insiste en recomendaciones de evaluar y corregir los registros antes de sumarlos a la producción de estadísticas? ¿Qué riesgos se asumen cuando los registros administrativos se utilizan sin ser evaluados? ¿Por qué no podemos confiar en que los grandes volúmenes de datos compensan esos errores? Esta sección presenta la diferencia conceptual que existe entre los datos administrativos y estadísticos, con el objeto de dar respuesta a esos interrogantes.

Los registros administrativos son fuentes de datos que se utilizan para la producción de estadísticas, aun cuando no son construidas específicamente para esos fines. No obstante, los registros tienen una alta frecuencia de generación, además de contar con una cobertura considerable, por lo que tienen el potencial de brindar información de calidad a costos más bajos que las encuestas y los censos.

Los registros administrativos (RA), son la información recolectada y conservada por las instituciones públicas, para controlar o intervenir en asuntos referidos tanto a personas físicas como jurídicas (INEGI, 2012; Fortin, 2000). CEPAL (2003) utiliza una definición que resalta la naturaleza transaccional de los RA, al decir que son el volumen de información resultante de gestionar procesos fiscales, tributarios u otros que se crea con la finalidad de viabilizar la administración de los programas de gobierno o para fiscalizar el cumplimiento de obligaciones legales de la sociedad. El registro civil de las personas, la administración tributaria, de seguridad social y las aduanas son algunos ejemplos de instituciones que producen registros administrativos con potencial para utilizarse con fines estadísticos.

De esta manera, los datos administrativos se diferencian claramente del resto de las fuentes estadísticas porque su diseño y gestión obedecen a los criterios normativos establecidos por el servicio administrativo que los crea, que suelen ser diferentes de los criterios metodológicos que orientan a las encuestas y a los censos. Están diseñados para registrar un acontecimiento administrativo y utilizan normas e instrumentos específicos para la captación de información, por ejemplo, aplicativos informáticos o formularios, los expedientes que regulan el flujo de mercancías a través de las fronteras (aduanas), los documentos que almacenan información sobre acontecimientos demográficos como nacimientos y muertes (registro civil) o la documentación que se utiliza para administrar obligaciones fiscales (impuestos internos), por nombrar algunos.

Por su parte, los datos estadísticos son aquellas fuentes de información que resultan de la aplicación de criterios metodológicos (por ejemplo, criterios probabilísticos que aseguren representatividad) que son específicos para la producción de estadísticas y que, además, cumplen con determinados estándares de calidad, como puede observarse en el **Recuadro 1**. Así, los datos para fines estadísticos se construyen para ser precisos y comparables, bajo determinados criterios estandarizados que les permiten ser integrados a distintos sistemas de información, a pesar de provenir de distintos tipos de fuentes: ya sea encuestas por muestreo, censos o registros administrativos debidamente procesados.

Recuadro 1. Dimensiones claves en la calidad de las estadísticas

De acuerdo con el Marco y Directrices de Calidad para Actividades Estadísticas de la OCDE (2011), las dimensiones más relevantes para asegurar la calidad de las estadísticas son:

1. Pertinencia: las estadísticas cumplen con las necesidades de los usuarios.
2. Precisión: las estadísticas representan con precisión y fiabilidad a la realidad.
3. Credibilidad: confianza depositada por los usuarios en los productos estadísticos.
4. Oportunidad y puntualidad: las estadísticas son publicadas de manera puntual y oportuna.
5. Accesibilidad: los datos se encuentran almacenados en locaciones seguras y conocidas por los usuarios que los demandan.
6. Interpretabilidad: los datos cuentan con toda la información complementaria para ser entendidos por los usuarios.
7. Coherencia y comparabilidad: las estadísticas son coherentes internamente, en el tiempo y en el espacio, y es posible combinar y hacer uso conjunto de datos relacionados provenientes de diferentes fuentes.
8. Costo eficiencia: Si bien no es una dimensión de calidad propiamente dicha, es un factor que se debe considerar en los análisis de calidad.

Fuente: Elaboración propia con base en OCDE (2011).

En efecto, los datos administrativos constituyen un tipo de fuente que se utiliza para la construcción de datos estadísticos, pero en su estado puro no pueden considerarse bases estadísticas (Wallgren y Wallgren, 2007). Para que esto sea posible, los registros administrativos deberían presentar los mismos atributos de calidad que se exigen a las estadísticas descriptas en el **Recuadro 1**.

La transición que va desde la generación del dato administrativo a la construcción del dato estadístico requiere el desarrollo de distintas etapas de evaluación y procesamiento, orientadas a otorgar al dato administrativo los atributos de calidad que caracterizan a las estadísticas. Estos procesos deben darse en el marco de construcciones institucionales y de modelos de gestión complejos, sin los cuales el aprovechamiento de los datos administrativos resultaría imposible.

2. Desafíos y oportunidades del uso estadístico de los registros administrativos para la producción de estadísticas

Los registros administrativos, como fuente para la producción de estadísticas e insumo para la investigación social, presentan ciertas ventajas y desventajas, las cuales han sido discutidas en diferentes estudios (Eurostat, 2010; CEPAL, 2003; DANE, 2015; INEGI, 2012), señalando, en general, los mismos desafíos y oportunidades en relación con el uso de estas fuentes de datos.

Entre las principales ventajas de los RA, pueden mencionarse las siguientes:

- 1) Bajos costos de recolección de datos, tanto para la agencia productora de estadísticas como para los respondientes, ya que el uso de estas fuentes, por un lado, disminuye la necesidad de realizar operativos de campo y, por otro, evita que las personas dediquen tiempo y recursos a completar encuestas donde muchas veces se solicita la misma información que contienen los registros.
- 2) Cobertura amplia de distintos fenómenos económicos o sociales, dado que los RA no son muestras, sino que cubren a la totalidad de los individuos involucrados en intercambios con agencias públicas.
- 3) Menor frecuencia de errores de “no respuesta” respecto de otras fuentes, y no hay errores muestrales.
- 4) Posibilidad de desagregar la información a nivel territorial, así como también de sectores económicos y sociales de tamaños reducidos.
- 5) Posibilidad de construir series longitudinales con mayor facilidad, dado que se tratan de procesos continuos y con identificadores estables que, a diferencia de los relevamientos de datos puntuales, permiten seguir una misma unidad a lo largo del tiempo.

Por su parte, la mayoría de los desafíos que se presentan ante el uso de este tipo de fuentes derivan del hecho que los registros son creados y gestionados con objetivos distintos de la producción estadística, lo que constituye el eje central de sus principales desventajas:

- 1) Falta de correspondencia entre las unidades administrativas y las estadísticas.
- 2) Diferencias en las definiciones de las variables.
- 3) Inconsistencia de los datos y períodos de referencia con la finalidad estadística.
- 4) Posibles pérdidas de información debido a cambios de gestión, que pueden amenazar la continuidad de ciertos registros administrativos.
- 5) Ocasionalmente, ausencia de un identificador común en los registros que permita la conciliación de los datos.
- 6) Cobertura acotada al sector formal. Este problema implica que en la práctica la cobertura de los RA no es universal y que en su análisis hay que ser cuidadosos respecto a cuán generalizables o representativas son las conclusiones que pueden extraerse.
- 7) Bajo control del proceso generador de los datos y la calidad de la información almacenada, por parte de la agencia productora de estadísticas.

En estas dificultades que presentan los registros a nivel global, se deben enfatizar los que están relativamente más presentes en los países de América Latina, como puede ser la alta incidencia de la informalidad o el menor grado de informatización y fiscalización de los procesos administrativos que dan origen a los RA. Si bien la informalidad ha disminuido durante la última década, aún afecta a un porcentaje muy elevado del empleo y del producto de la mayoría de los

países de América Latina. A manera de ejemplo, se puede mencionar que, en el año 2015, más de la mitad de los trabajadores asalariados de Perú y México se encontraban excluidos del sistema previsional, mientras que en Argentina y en Colombia esta cifra alcanzaba los 30 puntos, como puede observarse en el **Cuadro 1**. Estas cifras reflejan una fracción de los asalariados que no se encuentran en los registros de la seguridad social de cada país, a la que debe sumarse la fracción de trabajadores cuentapropistas que tampoco se encuentran registrados. La elevada incidencia de la informalidad lleva a que los registros, que tienen cobertura solo sobre el sector formal, tengan menor cobertura sobre la actividad económica que en regiones más desarrolladas. Esta situación implica la necesidad de complementar estas fuentes con encuestas que den cuenta del tamaño y características del sector informal.

Por su parte, los indicadores de gobierno electrónico brindan una aproximación, aunque imperfecta, del grado de informatización de los procesos de gestión pública. Estudios recientes para la región muestran que estos procesos utilizan de manera incipiente las tecnologías de la información y comunicación que están a disposición (Naser y Concha, 2017). Asimismo, el menor nivel de informatización y fiscalización se refleja en bases de datos administrativos con mayores errores, omisiones e inconsistencias, poniendo en evidencia la elevada importancia que tiene la evaluación de la calidad de los RA y la corrección de sus errores, a fin de poder ser utilizados como fuentes para la producción de estadísticas.

Cuadro 1. Tasa de Informalidad laboral en algunos países de América Latina – Año 2015-

País	Tasa de empleo asalariado no registrado
Argentina	32,7
Bolivia	59,8
Brasil	22,4
Chile	17,5
Colombia	36,9
Costa Rica	30,1
Ecuador	41,5
El Salvador	52,0
Guatemala	64,9
México	62,1
Nicaragua	59,6
Paraguay	61,7
Perú	53,2
República Dominicana	22,7
Uruguay	11,9

Fuente: Elaboración propia con base CEDLAS, 2017

Nota: Porcentaje de trabajadores asalariados en empleos informales, considerando como tales a los que no tienen aportes al sistema de seguridad social.

3. Descripción de las etapas necesarias para el aprovechamiento de datos administrativos en la producción estadística

El aprovechamiento de datos administrativos para la producción de estadísticas requiere desarrollar un proceso de trabajo complejo de evaluación, transformación y uso de los datos en la producción de conocimientos útiles para el diseño, gestión y evaluación de la política pública. Este proceso involucra a una multiplicidad de actores, los cuales deben realizar actividades sobre la información de origen fiscal, no siempre de buena calidad y cuya confidencialidad suele estar fuertemente protegida en todos los países⁴. Es por ello, que la complejidad del proceso no es solamente técnica (grandes volúmenes de información que deben ser evaluados o el desarrollo de metodologías para mejorar su calidad), sino también institucional.

Entre los actores clave que participan en este proceso, deben mencionarse, en primer lugar, a las agencias que diseñan y gestionan los RA, las cuales tienen responsabilidad primaria sobre la información y un conocimiento más profundo sobre los procesos administrativos que dan origen a los datos. En segundo lugar, las agencias públicas que producen estadísticas (Institutos de Estadísticas u otras instituciones que forman parte de los Sistemas Estadísticos Nacionales), que acceden a los RA para convertirlos en fuentes de información consistente, contando con protocolos que garantizan su confidencialidad.

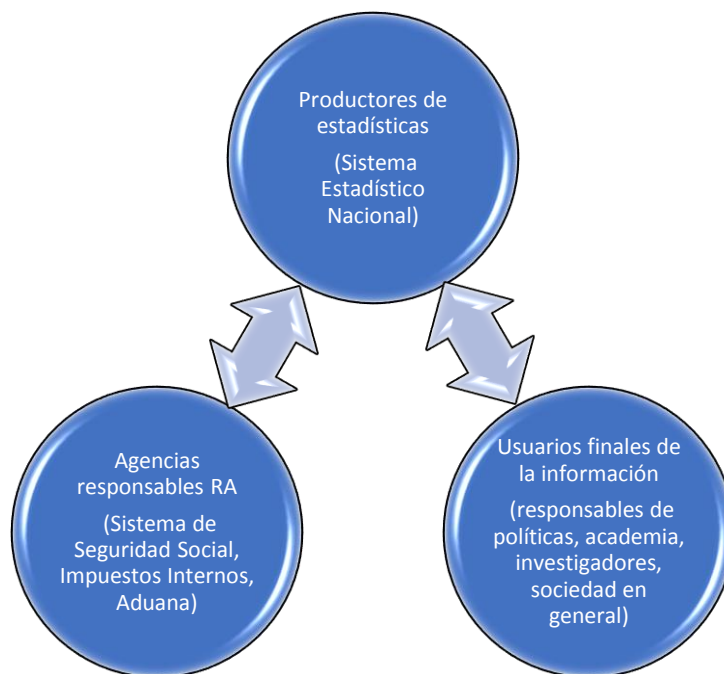
Una vez que los datos administrativos han sido transformados en fuentes estadísticas y se ha protegido la confidencialidad de los individuos, éstos pueden ser utilizados por un tercer grupo de actores (agencias responsables de las políticas públicas, universidades, centros de estudios, investigadores, periodistas, etc.), Estos pueden acceder a la información bajo la forma de datos agregados o como microdatos anonimizados, que serán utilizados en las investigaciones que alimentan los procesos de las políticas públicas, como puede observarse en la **Figura 1**.

Es importante destacar que, en el vínculo entre las agencias responsables de los RA y los investigadores, siempre debe intermediar una agencia productora de estadísticas que sea capaz de llevar adelante todo el proceso de adecuación de los datos, respetando los protocolos de confidencialidad mencionados⁵. También se debe destacar que la relación entre los tres grupos de actores no debe ser necesariamente lineal (proveedores de insumos a usuarios) ya que el flujo de información y la retroalimentación permiten mejorar el sistema de manera integral. Por ejemplo, es deseable que la agencia productora de estadísticas informe a los responsables de los RA las inconsistencias detectadas y que proponga mejoras. También puede ser muy beneficioso que el sector académico contribuya ofreciendo conocimientos o metodologías novedosas a la agencia productora de estadísticas.

⁴ Por ejemplo, puede citarse la protección de datos personales, así como también legislación de secreto fiscal, secreto bancario, entre otros.

⁵ En este sentido, parece poco factible que la función de adecuar RA pueda ser desarrollada por instituciones externas a los sistemas estadísticos nacionales, ya que esa función requiere condiciones legales y de infraestructura suficiente para resguardar la confidencialidad de los datos fiscales, los cuales no suelen existir en las instituciones externas. Por otro lado, es poco probable que esas instituciones, que tienen otros objetivos, estén dispuestas a invertir el tiempo y los recursos necesarios (dado que la adecuación de los RA es un proceso costoso) para producir datos a los que debería poder acceder toda la sociedad.

Figura 1. Actores clave que participan en el proceso de aprovechamiento de los registros administrativos para usos estadísticos



Fuente: Elaboración propia con base en los casos del OEDE (Argentina) e INEC (Ecuador).

Las etapas que componen el proceso de trabajo necesario para alcanzar los estándares de calidad de los datos estadísticos a partir de los RA contienen actividades que van desde la construcción de los acuerdos institucionales mencionados, pasando por etapas técnicas de análisis y corrección de la información, hasta el uso de los datos en la producción de conocimientos útiles para la sociedad. En base a distintas experiencias de la región⁶, es posible identificar, de manera estilizada, cuáles son las etapas que componen dicho proceso. En la **Figura 2** se presenta una breve descripción de las cinco etapas, identificando quienes son los actores clave involucrados y sus roles:

⁶ Se analizaron las experiencias del Observatorio de Empleo y Dinámica Empresarial (OEDE) del Ministerio de Trabajo, Empleo y Seguridad Social de Argentina y del Laboratorio de Dinámica Laboral y Empresarial del Instituto Nacional de Estadísticas y Censos (INEC) de Ecuador. También se revisaron experiencias documentadas para los casos de Colombia y México: el "Plan de Fortalecimiento de Registros Administrativos" del Departamento Administrativo Nacional de Estadísticas (DANE) de Colombia (DANE 2010), y el "Proceso estándar para el aprovechamiento de los registros administrativos" del Instituto Nacional de Estadística y Geografía (INEGI) de México (INEGI, 2012).

Figura 2. Etapas del proceso de adopción de registros administrativos para usos estadísticos



Fuente: Elaboración propia con base en los casos del OEDE (Argentina), DANE (Colombia, sobre el documento de DANE, 2015), e INEGI (México, sobre el documento de INEGI, 2012).

a) Identificación y selección de los registros de interés

El proceso inicia con una etapa de planificación, con el propósito de asegurar la pertinencia de los datos. Para ello, la selección de los registros administrativos con los que se va a trabajar debe realizarse siguiendo las demandas de información que recibe la agencia de estadísticas por parte de los usuarios (responsables de las políticas públicas y la academia). En este punto debe considerarse un factor sumamente importante y que puede ser determinante en la preferencia de un registro administrativo por sobre otro: la permeabilidad del organismo gestor de los registros a recibir sugerencias de las agencias productoras de estadísticas y a brindar espacios para discutir la definición de conceptos, variables, clasificaciones y demás. La participación del productor de estadísticas en el diseño del RA es vital para mejorar la calidad de la transición desde los RA a datos estadísticos.

b) Construcción de acuerdos institucionales

La segunda etapa consiste en la construcción de los acuerdos institucionales necesarios, con el objetivo de garantizar que la agencia productora de estadísticas pueda acceder a la información de manera segura y permanente, especificando la modalidad que se adoptará para diseminar

los datos una vez que hayan sido transformados. De esta manera, los usuarios finales siempre contarán con el acceso a la información respetando la legislación que corresponda.

En esta etapa, comunicar la utilidad que tienen los registros para la producción estadística resulta un factor importante. De esta manera, en las negociaciones que implica constituir formalmente un acuerdo de este tipo, suele esgrimirse como argumento el ahorro fiscal que significa reemplazar, aunque parcialmente, los censos y encuestas que se realizan periódicamente con datos de registros administrativos. Adicionalmente, los organismos encargados del registro también pueden beneficiarse de los controles y análisis estadísticos que se realicen, facilitando la incorporación de mejoras a sus propios procesos.

Los acuerdos institucionales entre las agencias de gobierno (responsables de los RA y productoras de estadísticas) deben formalizarse mediante convenios o regulaciones basadas en la elaboración de un proyecto o un plan de trabajo. Se considera que dejar estos compromisos entre agencias librados a relaciones interpersonales entre los funcionarios es una mala práctica, ya que fácilmente los acuerdos informales pueden desaparecer ante cambios de gestión, amenazando la continuidad del proyecto. Un avance importante en los marcos institucionales para el aprovechamiento estadístico de los registros se puede encontrar en la experiencia del Sistema Estadístico Nacional de Colombia, donde está garantizado por Ley el acceso del DANE a los registros administrativos necesarios para cumplir sus funciones. Esta experiencia se encuentra comentada en el **Recuadro 2**.

Recuadro 2. Acceso a los registros administrativos garantizado por Ley: La experiencia del SEN de Colombia

El artículo 160 de la Ley 1753 de 2015 (reglamentado por el Decreto 1743 del 2016) establece la creación del Sistema Estadístico Nacional (SEN) de ese país. Esta Ley dispone, en el segundo párrafo del artículo 160, que para la producción y difusión de estadísticas oficiales los integrantes del SEN deberán **poner a disposición del DANE, de manera inmediata y de forma gratuita, las bases de datos completas de los registros administrativos que sean solicitados por el departamento, para lo cual no será oponible ninguna reserva legal**. El DANE garantizará la reserva y confidencialidad de la información en concordancia con el artículo 5° de la Ley 79 de 1993.

Por otra parte, la definición de los integrantes del SEN es sumamente amplia e incluye entidades que produzcan y difundan estadísticas o sean responsables de registros administrativos, como puede ser: 1) las pertenecientes a las ramas del poder público, en todos los niveles de la estructura estatal, central o descentralizada por servicios o territorialmente; del orden nacional, departamental, municipal y distrital; 2) los órganos, organismos o entidades estatales independientes o autónomos de control; 3) las personas jurídicas, públicas o privadas, que presten servicios públicos; 4) cualquier persona jurídica o dependencia de persona jurídica que desempeñe función pública o de autoridad pública; 5) personas jurídicas que posean, produzcan o administren registros administrativos en el desarrollo de su objeto social, que sean insumos necesarios para la producción de estadísticas oficiales.

A finales de 2017, el DANE contaba con 100 registros administrativos para su uso, que cumplen una amplia variedad de temáticas, y había programado incorporar 20 más.

Fuente: Elaboración propia con base en la Ley 1753 de Colombia y consultas al DANE.

Además de garantizar el acceso de la agencia productora de estadísticas a los datos, de establecer cuál será el protocolo para garantizar la confidencialidad de los mismos en esta primera etapa y de especificar cuál será la política de disseminación de los resultados (hacia los usuarios institucionales o al público en general)⁷, los convenios deben incluir aspectos específicos de la operación estadística, como la cobertura temática y geográfica, los mecanismos para establecer los requisitos de la información y las adecuaciones conceptuales que puedan incorporarse. En este sentido, es conveniente que estos aspectos sean especificados ya en los acuerdos iniciales.

Cuando los organismos responsables de los registros administrativos participan en los procesos de trabajo de evaluación y adecuación, es preciso identificar las responsabilidades que cada institución asume. Por ejemplo, en cuanto a: (i) los aportes de recursos humanos, materiales y financieros; (ii) la forma y frecuencia de la transferencia de datos a la oficina de estadística; (iii) las funciones relacionadas con el control de la calidad de los datos; (iv) el intercambio de asistencia técnica para la generación de las estadísticas (INEGI, 2012).

Es importante recordar que una de las principales ventajas de utilizar registros como fuentes para la producción de estadísticas es que estos se actualizan de manera continua. Por ese motivo, el acuerdo institucional, lejos de garantizar el acceso a la información por única vez, debe considerar un cronograma de entregas periódicas que alimenten, de manera continua, la producción de estadísticas.

c) Conformación del equipo de trabajo

Una vez acordado el acceso al registro administrativo, se necesita conformar los equipos de trabajo, incluyendo tanto a los profesionales como a los recursos tecnológicos (hardware y software) que sean adecuados para manejar el volumen de información que representa el registro. En esta etapa es preciso contar con protocolos para el uso seguro de los datos, lo que implica contar con servidores adecuados y con accesos restringidos únicamente a personas que necesitan utilizar los datos.

Esta tarea requiere un equipo de profesionales dedicados de manera permanente a la transformación de los registros. Dado que completar el proceso es importante, no se recomiendan esquemas de alta rotación o con baja dedicación de tiempo, por resultar inadecuados a los fines propuestos, ya que la tarea implica trabajar con mucha información específica, que normalmente no está codificada y, por lo tanto, resulta costoso transferirla a nuevos miembros del equipo.

Por su parte, documentar los procesos de trabajo es una práctica que resulta central en estos casos. Los manuales, notas metodológicas, procesos informatizados no sólo son útiles para el usuario de los datos, sino que también permiten garantizar la continuidad de la manera de trabajar. La “codificación del conocimiento tácito” es una estrategia importante para transferir conocimientos a los nuevos miembros del equipo.

d) Adecuación de los datos administrativos para usos estadísticos

La cuarta etapa se centra, principalmente, en la evaluación de calidad y adecuación del registro. Si bien este tema se retoma con mayor profundidad en la próxima sección, aquí se describe

⁷ Más adelante se discute este tema en mayor detalle.

brevemente, y desde una perspectiva administrativa, las tres fases que se deben desarrollar para realizar la tarea con éxito.

Primero, la fase de evaluación de la calidad de los RA, que incluye: la recopilación de información; la evaluación y el diagnóstico de los registros administrativos; la detección de “fallas” desde la perspectiva estadística; la formulación de propuestas de mejora; y el plan de trabajo para la implementación de dichas mejoras, que se puede enviar a las agencias responsables de la gestión de los RA.

Segundo, la fase de corrección de fallas y adecuación de los datos, que incluye el desarrollo de metodologías para la corrección de los registros individuales que hayan mostrado errores en la fase anterior. En este sentido, resulta fundamental la elaboración de un plan de trabajo para abordar el problema, contemplando que la fase concluye con la integración de los registros a los sistemas de información que sean requeridos y con la construcción de series longitudinales.

Tercero, la fase de captación de los datos y su procesamiento, que se refiere a un proceso continuo de recepción de datos y la aplicación automática de las correcciones mencionadas. Según sea el caso, también puede incluir la producción de tabulados o indicadores resumen. Una vez implementado el proceso de adecuación del RA para uso estadístico, las acciones de captación y procesamiento se repiten permanentemente, siendo clave que periódicamente se evalúe como se está desarrollando el procesamiento de los datos, con el propósito de realizar los cambios y ajustes necesarios para lograr la mayor calidad posible de la información estadística.

Recuadro 3. Plan de Fortalecimiento de Registros Administrativos de Colombia

El Departamento Administrativo Nacional de Estadística (DANE) de Colombia, como coordinador del Sistema Estadístico Nacional (SEN), diseñó la Estrategia para el Fortalecimiento Estadístico Territorial (EFET). En ese marco, el DANE diseñó una metodología con el objetivo de aprovechar el gran número de registros que manejan las entidades públicas y privadas del país.

En ese documento, el DANE propone cinco etapas para la formulación de un plan de trabajo, con el que se busca contribuir a la mejora de los registros, a partir de lineamientos técnicos que permiten alcanzar una producción de datos de calidad. Esas etapas son:

Etapas 1: Selección de registros administrativos (RA). Incluye la elaboración de un inventario.

Etapas 2: Recolección de información sobre las características técnicas requeridas para las fases de diseño, producción y difusión de RA con fines estadísticos.

Etapas 3: Diagnóstico del estado actual de los registros administrativos que formarán parte del plan de fortalecimiento.

Etapas 4: Formulación de propuestas de mejora para el fortalecimiento de los registros administrativos.

Etapas 5: Implementación del plan de fortalecimiento de registros administrativos en el país.

Fuente: Elaboración propia con base en DANE 2015.

e) Diseminación de resultados: estadísticas agregadas y microdatos

La quinta etapa corresponde a la difusión de las estadísticas y está orientada a satisfacer las demandas de otros usuarios, ya sean personas que están en la gestión de las políticas públicas como miembros de la comunidad académica. La difusión no solo debe estar en línea con las condiciones jurídicas y éticas del tratamiento confidencial de los datos que sean pertinentes en cada contexto, sino que también presenta un desafío de naturaleza técnica, ya que implica procesos de anonimización y documentación que no son triviales.

En particular, la difusión de datos puede hacerse a través de indicadores, tabulados con datos agregados, pero también se puede dar acceso directo a microdatos que estén desidentificados o anonimizados, siempre que esto sea posible, es decir, cuando las condiciones legales, tecnológicas y administrativas así lo permitan⁸. Cabe mencionar, además, que la naturaleza cambiante del proceso legal, administrativo y tecnológico que genera los datos implicará que los acuerdos institucionales diseñados para compartir los datos también puedan cambiar con el tiempo y que sea necesario reconsiderar los aspectos contemplados inicialmente en dichos acuerdos.

La experiencia del DANE (Colombia) es un ejemplo de buena práctica en esta área. En 2015 se estableció un Programa Anonimización de todas las bases las operaciones estadísticas que sean susceptibles a anonimizar, cuya primera etapa consiste en investigar cuáles son las metodologías más adecuadas para aplicar a las encuestas económicas, con el objetivo de garantizar la reserva, causando la menor perturbación posible en los datos (Prada Lombo, 2017).

⁸ A pesar de no ser el foco de este trabajo, vale la pena destacar que una práctica usual para anonimizar información que requiere ser protegida por la legislación pertinente en cada país es eliminar los identificadores más obvios de las bases de datos que se comparten hacia fuera la agencia generadora de los RA o de la agencia productora de estadísticas. Estos identificadores son usualmente nombres, números de identificación tributaria o de la seguridad social, direcciones, etcétera. Sin embargo, esta práctica no siempre garantiza una anonimización segura, por una multiplicidad de motivos que van desde unicidad de características que determinan identidad de firmas o individuos (por ejemplo, una única firma o contribuyente grande en un sector o zona determinada) a la posibilidad de complementar la información de los RA con otras fuentes externas que permiten reconstruir la identificación original. Por su parte, existen otras formas de proteger la confidencialidad de los RA, como puede ser la autorización de acceder a los mismos únicamente desde “enclaves” dispuestos a tal efecto por las propias agencias productoras de datos; o algunas formas de acceso remoto pensadas para resguardar la información de manera apropiada. Vilhuber (2017) resume los aspectos centrales de las distintas alternativas para difundir o compartir microdatos que puedan devenir de RA.

4. Aspectos técnicos de la adecuación de los datos administrativos para usos estadísticos

La evaluación de la calidad de las fuentes originales y el desarrollo de las correcciones necesarias para alcanzar niveles de calidad apropiados es la parte del proceso que efectivamente permite el uso de los datos administrativos para fines estadísticos. En esa etapa, se analizan de manera exhaustiva, tanto los procesos administrativos que originan los registros como los datos almacenados, se detectan problemas y se identifican las mejores estrategias para corregirlos. Es importante destacar que la adecuación para usos estadísticos va más allá de la simple corrección de inconsistencias e incluye procesos orientados a construir sistemas de información más complejos y versátiles, como la integración de distintos RA y la construcción de series longitudinales.

Esta sección está focalizada en esas actividades. Con un enfoque eminentemente práctico, el texto está destinado a los equipos técnicos de las agencias estadísticas que utilizan registros administrativos de empresas y empleo. Por ese motivo, en su desarrollo, presenta numerosos detalles metodológicos y ejemplos tomados de experiencias de la región, que pueden ser de gran utilidad para los equipos que se encuentren iniciando esta tarea. Según el caso, las metodologías para realizar esos ajustes pueden ser conocidas, o más innovadoras cuando los problemas requieren soluciones creativas.

A. Evaluación de la calidad de los registros

Cuando los registros administrativos constituyen los principales insumos para la producción de las estadísticas, la calidad del resultado final dependerá directamente del tipo de información que se utilice desde el comienzo. Es importante aclarar que la “calidad” de la información se define a partir de su utilidad para fines estadísticos⁹. Es decir, no sólo es importante la precisión que tienen los datos para reflejar un fenómeno económico o social, sino también importa un conjunto de dimensiones/atributos que ha sido identificadas por los organismos que coordinan la producción de estadísticas. Por ejemplo, podría ocurrir que los datos fueran precisos, pero que no tuvieran una cobertura adecuada o que no se pudiera acceder fácilmente a ellos, limitando de esa manera su utilidad (y su calidad). El **Recuadro 1** destaca las dimensiones priorizadas por el “Marco y Directrices de Calidad para Actividades Estadísticas” de la OECD.

Para evaluar los RA, es preciso verificar que los atributos de calidad mencionados estén presentes en los datos. En este sentido, el DANE de Colombia (DANE, 2015), propone organizar la evaluación de calidad de los RA en dos niveles. Por un lado, estudiando los **procesos** que generan los registros y, por otro, analizando la estructura de **los datos** almacenados. El análisis de la calidad del proceso debe contemplar aspectos legales, conceptuales, de cobertura de eventos, de cobertura y desglose territorial y temporal, así como también de procedimientos y tecnologías para recabar la información necesaria. Para el caso de las bases de datos propone

⁹ Como se ha mencionado anteriormente en este documento, es diferente la calidad de un RA según se considere su uso para la producción de estadísticas o su uso para la gestión administrativa. Un RA puede presentar un alto nivel de calidad desde una perspectiva administrativa, cumpliendo con todos los atributos requeridos por el proceso de gestión, pero presentar dificultades para la producción estadística. Las diferencias entre los conceptos administrativos y estadísticos (por ejemplo, “contribuyentes” versus “unidad productora de bienes o servicios”), los clasificadores de actividad y de localización, las unidades de tiempo, las coberturas relevantes, son algunos aspectos que ayudan a ilustrar esas diferencias.

evaluar la calidad con criterios de coherencia, precisión y exhaustividad, entre otros (DANE, 2015).

a) Análisis de los procesos de gestión que dan origen al registro

Entre los principales aspectos que se analizan en los procesos de gestión que originan los registros se pueden mencionar: la cobertura, las definiciones de las variables consignadas, los períodos de referencia y los criterios de codificación aplicados. Este tipo de análisis también se denomina “análisis conceptual”, ya que compara los conceptos y definiciones que contienen los datos administrativos con los que requieren las operaciones estadísticas.

Para analizar los procesos generadores de los registros administrativos se debe recurrir a la información (documentada o no) que esté disponible sobre el diseño y funcionamiento. Cuando la documentación de los registros existe, se consulta el marco legal, los procedimientos, los aplicativos informáticos u otros instrumentos utilizados para la carga de la información y los procesos de validación interna de los datos. Cuando la documentación no se encuentra disponible, es preciso recurrir a entrevistas y consultas con los técnicos de los organismos responsables, ya que, ante la ausencia de procesos documentados, la información necesaria suele estar en manos de estas personas.

Entre los organismos productores de estadísticas que fueron consultados para este estudio se reconoce la importancia de realizar el análisis conceptual de los registros. Los institutos de estadísticas consultados y el OEDE del Ministerio de Trabajo de Argentina realizan la actividad de manera rutinaria. En el DANE de Colombia el análisis conceptual de los registros está pautado en un manual. En el IPEA de Brasil, indicaron que algunos de los registros que utilizan están poco documentados y, por lo tanto, en esos casos, la evaluación de calidad se realiza directamente sobre las bases de datos, consultando, de manera puntual, a los organismos responsables sólo en el caso de detectar inconsistencias (Corseuil y Mation, 2017).

Si bien, la evaluación conceptual debe diseñarse considerando las especificidades de cada registro, en base a las experiencias estudiadas, es posible señalar un conjunto de aspectos que con frecuencia han presentado problemas en los registros de empresas, comercio exterior y trabajadores. Es importante recordar que, como se ha mencionado, muchas de estas situaciones no deterioran la calidad de los RA desde la perspectiva del uso administrativo de los mismos, sino desde su uso estadístico, que es el problema que buscamos abordar.

i. Definiciones administrativas y definiciones estadísticas

Resulta central conocer en profundidad cuales son las definiciones utilizadas en los registros administrativos y compararlas con las que requiere la producción estadística. Cuando se detectan discrepancias conceptuales, es necesario encontrar la manera de homologarlas. Cuando esas correcciones no se pueden realizar, las discrepancias en las definiciones deben explicitarse en los manuales o notas metodológicas, en lo posible identificando el sentido de los sesgos que podrían introducir al análisis, dado que la peor estrategia en este punto es ignorar las diferencias y los sesgos, pues conllevarán problemas en el futuro. A continuación, se presentan algunos ejemplos que se han observado en las experiencias estudiadas:

Contribuyentes, empresas, establecimientos productivos y locales: Para la gestión tributaria, la unidad de análisis es el contribuyente, definido como una unidad legal independiente e identificado por su código único de identificación (CUIT, RUT, RUN, por ejemplo). La definición

de contribuyente no siempre coincide con los conceptos de “local”, “establecimiento”, “firma” o “grupos económicos” que suelen utilizarse en la producción de estadísticas y en el análisis económico en general. Los contribuyentes más grandes suelen agrupar varios establecimientos productivos con distintas localizaciones que, incluso, pueden desarrollar diversas actividades económicas. Considerando que el tipo de actividad productiva y su localización son dimensiones importantes para los estudios regionales, los registros deben complementarse con otras fuentes que permitan desagregar la información al nivel relevante. Otro ejemplo son los grupos de contribuyentes que mantienen vínculos patrimoniales (grupos o conglomerados económicos), en esos casos, los estudios requieren información adicional al registro que permita agruparlos.

Un ejemplo interesante es el “multirut” de Chile. El marco jurídico de ese país permite que las empresas, incluso en un mismo domicilio de explotación y desarrollando la misma actividad económica, declaren su empleo segmentado en códigos de contribuyentes (RUT) diferentes. Ejemplos de esta situación son: la firma D&S (actual Walmart Chile) y el holding Quiñenco (Grupo Luksic) (Miranda, 2017; Krell, 2017). En los ejemplos mencionados, para el análisis económico se debe recurrir a fuentes externas que permitan agrupar la información a nivel de empresa, establecimiento o grupo económico; o marcar claramente, en una nota metodológica, la discrepancia conceptual detectada. En este ejemplo, el error introduciría un sesgo importante al análisis: cada contribuyente podría confundirse con una microempresa, cuando en realidad son fragmentos que componen grandes empresas.

Remuneraciones pagadas, percibidas o bases imponibles: Los registros administrativos de la seguridad social están diseñados para estimar los montos de aportes y contribuciones que se deben pagar a los distintos componentes de los sistemas de seguridad social (sistema previsional, salud, seguro de desempleo, por ejemplo). Es decir, la variable principal de remuneraciones del registro es una “base imponible” y no coincide con los conceptos de “salarios netos” o de “salarios brutos” de impuestos. En cada caso, habrá que realizar ajustes aplicando las tasas correspondientes para pasar de una definición a otra, o aclarar en una nota técnica.

Variabes truncadas: Esta situación se puede producir cuando los registros consignan el valor de una variable, pero sólo cuando estos se encuentran en un intervalo determinado, perdiéndose la información para los valores fuera del mismo. Por ejemplo, en países donde las tasas de aportes personales y contribuciones patronales se aplican al salario bruto hasta un valor máximo, el salario consignado como base imponible en los registros está “truncado” en ese valor máximo. Esta situación se observó en los registros de la seguridad social de Argentina y de Uruguay y es posible que se repita en otros sistemas previsionales similares (Ferreira, 2017; BPS, 2017; OEDE,2017).

Campos que caen en desuso y se destinan a usos internos: En ocasiones, la gestión de las políticas requiere cambios imprevistos en los procesos y en los registros, que no alcanzan a documentarse de manera adecuada. Estos cambios pueden ser temporales (el rediseño de los sistemas suele ser más lento que los rediseños en las políticas), corrigiéndose y documentándose posteriormente. Por ejemplo, en el año 2004, en el SIPA de Argentina se utilizó transitoriamente un campo llamado “hijos” -que había caído en desuso al cambiar el régimen de asignaciones familiares- para almacenar una variable de uso interno. En esa oportunidad, el cambio no fue informado a los usuarios y generó confusiones. Por este motivo, en los registros puede existir información sumamente relevante que no está codificada y, por ello, es central establecer vías de consultas y comunicación con los organismos productores de los registros, de manera de recibir alertas incluso informales sobre este tipo de situaciones.

ii. Cobertura del registro

Como se ha mencionado, los registros administrativos de actividades económicas y de empleo reflejan el sector formal de la economía, lo que constituye una limitación importante para los países de América Latina que presentan economías informales significativas. Este aspecto no se puede modificar, pero es importante aclararlo, para evitar generalizar análisis realizados sobre una parcialidad de la población. Más allá de la informalidad, los registros pueden tener coberturas parciales: por tipo de contrato laboral, región, edad de las personas, actividades económicas, tamaños de las empresas, entre otros aspectos relevantes. En esos casos, al no conocer con precisión la cobertura del registro, éste puede convertirse en una fuente estadística no exhaustiva. Esto se corrige solicitando registros complementarios, si estos existen, o explicitando que la cobertura de la fuente está acotada a determinados casos.

Información fiscal segmentada en distintas bases administrativas: En los países suelen existir diferentes regímenes impositivos que dan origen a distintos RA. Puede haber segmentaciones según el tamaño de la unidad productiva: grandes contribuyentes, sociedades, régimen general y regímenes de simplificación registral para microempresas (como los monotributistas), por ejemplo. También según sectores económicos: los sectores petroleros, financieros o lecheros, según la importancia de estas actividades en cada país, pueden ameritar regímenes especiales. Así, la información contenida en cada registro puede diferir, por ejemplo: para los grandes contribuyentes, los RA consignan datos del balance de las empresas, mientras que para los medianos sólo las ventas, y para los pequeños contribuyentes sólo se registra un rango presunto de facturación. Es muy importante que el productor de estadísticas conozca si accede al universo completo (o hay registros faltantes) y exactamente a qué información accede para cada colectivo de contribuyentes. También hay que desarrollar criterios para unificar información proveniente de distintos registros en bases de datos estadísticas integradas como, por ejemplo, los directorios de empresas.

Por ejemplo, en Ecuador la información de impuestos internos se encuentra segmentada en distintos regímenes impositivos. Para completar el directorio de empresas, fue preciso solicitar las diferentes bases de datos administrativas que cubren ese universo formal. Otro ejemplo, es Uruguay, donde existen otras instituciones de seguridad social, además del BPS, cuyos registros no están consolidados. Otro caso es el SIPA (Sistema Integrado Previsional Argentino), que en el período 1994- 2011 no contaba con información completa sobre el empleo público, ya que un conjunto de provincias registraba su empleo público en cajas previsionales propias.

iii. Sistemas de codificación utilizados

Los sistemas estadísticos utilizan códigos armonizados para diferentes variables, por ejemplo, actividad económica, regiones, ocupaciones, productos, entre otros. El uso de estos códigos tiene una enorme importancia para producir información comparable, no sólo entre distintas operaciones de un mismo sistema estadístico a lo largo del tiempo, sino también con otros países. El uso de códigos armonizados permite también aprovechar diferentes taxonomías ya desarrolladas, que agregan complejidad analítica a la información. Como ejemplo se pueden mencionar categorías que agrupan actividades económicas según el uso de factores e intensidad tecnológica, o a las ocupaciones según sus requerimientos de calificación.

Sin embargo, para facilitar procesos de gestión, o por desconocimiento, algunos sistemas administrativos eligen adoptar códigos y nomenclaturas ad-hoc que no pueden ser fácilmente reconvertidos a los códigos estándar y que presentan un problema de difícil solución. Muchas

veces, la mejor manera de sanear esta situación es a través de la vinculación con otros registros que tengan la información codificada de manera correcta, a nivel del identificador del agente. También, es frecuente que a lo largo del tiempo cambien los sistemas de codificación utilizados en los registros, adoptando revisiones más modernas o estándares y nomenclaturas internacionales.

Si bien estos cambios persiguen el objetivo de mejorar la calidad de los registros, su implementación desordenada suele causar nuevos errores. Por ejemplo, puede darse casos donde conviven en los directorios nomencladores antiguos y nuevos, los cambios se realizan sin informar a los usuarios o no se prevén tablas de equivalencia biunívocas entre nomencladores. Entre los organismos productores de estadísticas consultados, todos mencionaron la importancia de verificar los nomencladores y relataron haber tenido dificultades para conseguir información, en particular sobre cambios ocurridos en el pasado.

Para sanear estos problemas lo habitual es utilizar los nomencladores antiguos durante más tiempo, hasta que la mayoría de los agentes se encuentren clasificados con el nuevo nomenclador. También se pueden elaborar tablas de equivalencias entre nomencladores, aunque esto muchas veces implica perder desagregación en las variables.

iv. Existencia de criterios internos de validación en la carga de los datos

Muchas veces, los RA cuentan con criterios de validación que están automatizados y no permiten el ingreso de datos inconsistentes al momento de la carga de la información. Por ejemplo: las variables categóricas solo pueden adoptar ciertos valores, las edades pueden tener un rango o pueden existir relaciones con otras variables de la misma base.

La consulta realizada a productores de estadísticas indica que, actualmente, los registros administrativos económicos más importantes (impuestos internos, aduana y seguridad social) tienen reglas internas de validación en los formularios electrónicos, lo que garantiza un mínimo de coherencia de la información, no obstante, hay inconsistencias y/o errores que son permeables a las reglas actualmente vigentes. Por este motivo, se revisan los procesos de captura de la información, así como el procesamiento de las bases de datos.

En otros casos, los sistemas de registros pueden carecer de procesos de validación automatizados, situaciones que son más frecuentes en registros antiguos o pequeños, particularmente en la gestión de programas o políticas públicas con baja cobertura, con poco presupuesto o que tienen desarrollos de sistemas muy básicos. En estos casos, el análisis de la consistencia de los datos, utilizando herramientas estadísticas, resulta central.

b) Análisis de la consistencia de los datos administrativos

Una vez analizado el proceso generador de los datos -y obtenido un panorama preciso- se procede a analizar con herramientas estadísticas la información contenida en las bases. Es muy importante destacar que, debido a la magnitud de información que contienen los registros administrativos, muchas veces el análisis estadístico es una tarea compleja -por ejemplo, los promedios de grandes números ocultan los errores que puedan persistir- es por ello que resulta particularmente importante detectar los posibles problemas de consistencia analizando el proceso de gestión que los origina.

El análisis de consistencia de los datos puede ser: interno (relaciones entre las variables de una base); externo (en relación con otras fuentes); o inter-temporal (análisis de tendencias y relación con comportamientos esperados).

i. Análisis de consistencia interna

En el análisis de consistencia interna de los datos, un primer aspecto a considerar, son las descripciones estadísticas tradicionales de las variables: distribuciones, máximos y mínimos, cantidad de observaciones perdidas, promedios, entre otras. Esta mirada suele generar las primeras preguntas: ¿por qué hay tantos casos perdidos? ¿a qué se deben algunos valores extremos sin sentido económico? ¿Por qué motivo las distribuciones tienen determinada forma?

Esta etapa de evaluación es sumamente importante. En particular, cuando los registros no están bien documentados, el análisis estadístico puede mostrar aspectos del diseño e implementación del registro que no fueron informados: variables truncadas, cambios en los sistemas de codificación, ausencia de determinados perfiles de agentes económicos, entre otros. Todos los organismos consultados realizan este tipo de análisis de los datos que, además, resulta accesible al no requerir información adicional.

ii. Comparación con datos externos sobre estructura, cantidades y composiciones

El análisis de consistencia externa implica la comparación del registro con otras fuentes estadísticas disponibles: censos, encuestas de hogares, distintos tipos de encuestas económicas, información proveniente de cámaras o de registros de comercio, etcétera. Este tipo de validación es el principal mecanismo para la detección de inconsistencias, aunque es importante destacar que las fuentes externas también pueden presentar problemas similares.

Para abordar el análisis, se puede comparar estructuras, promedios, brechas entre variables, distribución territorial, antigüedad, tamaños, sectores y cualquier otro tipo de medida que resulte relevante en nuestros datos. En caso de encontrar diferencias, la primera medida debe ser analizar el proceso generador del registro para encontrar una explicación razonable:

¿Falta una parte del registro?

¿Existen duplicaciones de datos?

¿La codificación tiene errores?

Una segunda medida, de ser posible, es comparar las fuentes a nivel de los microdatos. Es que, al vincular los RA (a nivel de agentes económicos) con otras fuentes, suelen encontrarse diferencias en determinadas variables, como los códigos de actividad, lo que permite evaluar la calidad de la información en las dos fuentes que se está comparando. Esto permite evaluar distintos aspectos de ambas fuentes:

¿De dónde provienen las variables?

¿Quiénes responden?

¿Cómo se fiscaliza?

¿Existen incentivos a responder de manera correcta?

Por ejemplo, el Laboratorio de Dinámica Laboral y Empresarial de Ecuador, realizó un estudio sobre la evolución del sector manufacturero, tomando como fuente la información de los registros de impuestos internos y de la seguridad social, adoptando la clasificación de rama de actividad consignada en el Directorio de Empresas, por considerarla de mejor calidad que las clasificaciones de rama que tienen los registros. Sin embargo, la misma institución, en un análisis del sector petrolero para el año 2009, encontró un número inesperado de firmas pequeñas. El análisis de los casos permitió detectar que esas supuestas pequeñas firmas en realidad eran distintas unidades de negocio de dos empresas (Petroecuador y Petroamazonas). De esta manera, para realizar el análisis económico del sector, optaron por consolidar la información a nivel de firmas (Garzón et al., 2016).

El INEI de Perú construye su Directorio de Empresas nutriéndose de los registros de contribuyentes y realizando comparaciones periódicas con otras fuentes de información como las encuestas y los censos desarrollados por el mismo instituto. Particularmente, en el caso de la actividad económica el Instituto prioriza la información captada por el censo económico o por la encuesta económica anual, por considerar que en esas fuentes la captan con mayor precisión que la autoridad fiscal. El DANE de Colombia incluye, entre los ejercicios de diagnóstico de los registros, siempre que resulte posible, un análisis de consistencia que se realiza mediante comparaciones, a nivel de microdatos, con operaciones estadísticas realizadas por esa institución. Otro ejemplo, es de Impuestos internos de Uruguay, donde a nivel de personas físicas se han hecho comparaciones con encuestas de hogares.

Es preciso destacar que es muy importante que la institución productora de estadísticas que se encuentre evaluando y corrigiendo los RA pueda acceder a datos no anonimizados: con códigos de identificación para poder cruzar las bases con otras fuentes a nivel de microdatos, e incluso con nombres para poder realizar controles rápidos. En el ejemplo de Ecuador, las firmas petroleras son conocidas para los analistas, a simple vista notaron que no se trataba de firmas pequeñas sino de unidades de negocios de empresas grandes. Al analizar los sectores con participación menor a la esperada en la estructura, el método más sencillo consiste en revisar si las firmas más grandes y conocidas, han sido omitidas.

B. Corrección de inconsistencias y errores, construcción de sistemas de información integrados y dinámicos.

A partir del ejercicio de diagnóstico y evaluación del registro, surgen propuestas de mejoras. En el corto plazo, esas propuestas están orientadas a corregir errores e inconsistencias mientras que, en el largo plazo, las propuestas deberían orientarse a alimentar un plan de mejoras que permita, a las agencias que gestionan los RA, potenciar su uso como insumo para la producción estadística. El proceso de implementación del plan de mejoras requiere de actividades adicionales por parte de la agencia de estadísticas, como el seguimiento y la evaluación, con el objeto de verificar la efectividad y el alcance de las recomendaciones y acciones de mejora propuestas, así como el cumplimiento por parte de las entidades de los compromisos adquiridos frente al plan de fortalecimiento para cada registro (DANE, 2015).

Recuadro 4. La importancia de corregir inconsistencias en los registros

Los equipos del INEI (Perú) consultados sobre el uso de RA para la producción de estadísticas, consideran imprescindible la etapa de corrección de los registros previo a su uso para fines estadísticos.

En el DANE (Colombia) se ha establecido un protocolo de evaluación y diagnóstico, seguido por la elaboración de un plan de fortalecimiento para cada RA que se incorpora a la producción estadística. El plan de fortalecimiento incluye la corrección de inconsistencias y debería ser desarrollado por la entidad responsable del RA. No obstante, cuando eso no ocurre, son los funcionarios que utilizan el RA y que participaron del diagnóstico quienes establecen, implementan y mantienen las correcciones necesarias que garantizan un uso estadístico adecuado.

En el OEDE de Argentina, el proceso de corrección de inconsistencias se realiza de manera continua y es una condición previa al uso de los registros con fines estadísticos.

Fuente: Elaboración propia con base en consultas realizadas a informantes clave.

a) Corrección de inconsistencias en cada registro

Para corregir las inconsistencias detectadas en los registros debe aplicarse aquella metodología que resulte más adecuada para cada situación. Algunas son simples y conocidas, mientras que otras requieren del uso de técnicas estadísticas más complejas o de cierta capacidad innovadora por parte de los equipos técnicos que se encuentren desarrollando la tarea.

Sin embargo, las transformaciones deben estar claramente documentadas en todos los casos. Por un lado, porque la documentación es crucial para poder replicar los procesos a lo largo del tiempo, ya que la gestión administrativa recolecta información de manera continua y los criterios de corrección deben automatizarse, constituyendo también un proceso continuo en sí mismo. Por otro lado, porque la documentación debe poder comunicarse a los usuarios. En este sentido, puede optarse por la provisión de metadatos o manuales metodológicos que deben estar orientados a la descripción de los procesos administrativos y estadísticos llevados a cabo para la construcción del registro y las bases de datos estadísticas.

A continuación, a manera de ejemplo, se presentan algunas correcciones frecuentes que realizan las agencias productoras de estadísticas consultadas:

i. Aplicar normas de validación ex – post, para eliminar valores inconsistentes

Para cada registro y conjunto de campos se pueden crear normas de validación específicas a la lógica de cada caso. Los criterios pueden ser diversos, como establecer rangos con los valores que puede tomar una variable o establecer relaciones entre variables de una misma base. La identificación de criterios que permitan distinguir entre valores perdidos y valores nulos es importante. Una vez determinados los criterios de validación, estos se pueden aplicar ex – post, para identificar valores inconsistentes, que generalmente se eliminan (*missing*). En algunos casos, es posible imputar la información para reemplazar los datos inconsistentes.

Uso de valores históricos del mismo agente para detectar errores de carga: El uso de información histórica del mismo agente económico puede permitir identificar “valores atípicos”. Un ejemplo es la corrección de errores en la base de remuneraciones de SIPA, realizada por el OEDE de

Argentina. A partir de la construcción de intervalos de confianza, basados en las historias laborales de cada trabajador (mediana del salario) se identifican valores inconsistentes con la trayectoria de remuneraciones de cada persona. En todos los casos, los valores observados fuera del intervalo fueron considerados valores perdidos¹⁰ (OEDE, Metodología del Boletín de Remuneraciones). En el caso de Impuestos Internos de Uruguay, un error frecuente es que los contribuyentes declaran el valor de sus ingresos incluyendo decimales (centavos), pero esta situación no está prevista en el sistema informático, por lo que el ingreso queda multiplicado por 100. Estos errores se detectan cruzando las bases con los datos de los empleadores y se corrigen fácilmente.

Eliminación de valores inconsistentes desde lo conceptual. Hay valores correctos, desde la perspectiva administrativa, pero que por su falta de sentido económico se vuelven inconsistentes. Por ejemplo, los registros administrativos de comercio exterior incluyen operaciones como mercaderías en tránsito o importaciones temporales.

ii. Eliminación de casos duplicados

En ocasiones, las bases de datos administrativas presentan casos duplicados, debido a errores o fallas en los controles originales. Esos casos se eliminan, cuidando que los campos de identificación confirmen que se trata de un caso duplicado y no de otro tipo de error. Estos errores resultan cada vez menos frecuentes, en la medida en que los sistemas evitan este tipo de situaciones. No obstante, en Colombia, Perú y Uruguay se realizan controles de manera sistemática con el objetivo de evitar duplicados.

iii. Corrección de clasificadores (ramas de actividad, localización, etc.)

Los errores de clasificación pueden ser múltiples: aplicación de clasificadores que no corresponden al estándar utilizado por el país; mala implementación del clasificador; entre otros. Generalmente, cuando se trata de características que no son “críticas” para la gestión administrativa específica—como la rama de actividad, o el código postal del domicilio de explotación en un registro orientado a recaudar impuestos— éstas no suelen ser fiscalizadas adecuadamente o, incluso, pasadas por alto. Esto lleva a que la calidad de los datos de los clasificadores resulte más apropiada en los casos de censos y encuestas por sobre los registros administrativos, en particular cuando se trata de firmas pequeñas o de personas de naturaleza jurídica, que reciben instancias de fiscalización menos frecuentes que las firmas grandes. Por este motivo, las distintas agencias estadísticas se esfuerzan en evaluar y corregir esta información de clasificación.

Los productores de estadísticas que fueron consultados corrigen las ramas de actividad de los registros utilizando principalmente fuentes externas (encuestas, censos y relevamientos específicos para limpiar las ramas). En el caso de Impuestos Internos de Uruguay (es el

¹⁰ En el período de precios estables 1996-2001 los límites definidos para los intervalos fueron: (i) límite inferior = 5% de la mediana; (ii) límite superior para el mes en el que cesa la relación laboral = 3 medianas; (iii) límite superior para el resto de los meses en los que la relación laboral es declarada = 5 medianas. En el contexto de salarios nominales crecientes (2002-2009) debieron ampliarse los límites superiores de los intervalos. La tendencia positiva de las remuneraciones ocasionó que, en los últimos meses de cada año, el límite calculado sobre la mediana anual resulte restrictivo. Asimismo, la Ley de Emergencia Económica estableció, que para ese período el monto de las indemnizaciones fuera doble. Por estos motivos, para ese período, los límites quedaron definidos de la siguiente manera: (i) límite inferior = 5% de la mediana; (ii) límite superior para el último mes de esa relación laboral = 4 medianas; (iii) límite superior para el resto de los meses en los que la relación laboral es declarada = 10 medianas.

organismo responsable de registros y a la vez productor de estadísticas) llama la atención porque manejan dos campos de ramas de actividad: uno para usos estadísticos y la original, de usos administrativos.

iv. Imputación de datos faltantes

En algunos casos, la información faltante en los registros se puede completar a partir de otros datos disponibles. A manera de ejemplo, se citan dos experiencias.

Sexo y año de nacimiento de las personas: En algunos países es posible imputar estas características obteniendo información contenida en los códigos de identificación. Por ejemplo, en Argentina se puede conocer el año de nacimiento de las personas a partir del número de Documento Nacional de Identidad (DNI) asignado de manera consecutiva a las personas en determinados eventos de su vida (al nacer, o al cumplir los 18 años). Por lo tanto, existe una correspondencia bastante ajustada entre los primeros dígitos del DNI y el año de nacimiento (salvo para los extranjeros residentes y nacionalizados, quienes reciben su DNI argentino en un evento independiente de la fecha de nacimiento). Otro ejemplo, también de Argentina, es que los primeros dígitos del Código Único de Identificación Laboral (CUIL) son diferentes para varones y mujeres, por lo tanto, el mencionado código brinda ese tipo de información.

Casos perdidos por incumplimiento (o demoras) en la presentación de declaraciones a la autoridad fiscal: Un proceso algo más complejo, es la imputación de datos perdidos en el caso de empresas que presentan sus plantillas de personal de manera intermitente. Esta situación suele ser frecuente en los registros de empleo de América Latina, en particular entre las firmas de menor tamaño. Desde una perspectiva económica, no es razonable pensar que estas firmas efectivamente despiden y recontratan todo su personal de manera intermitente (el marco legal no lo permite y los costos de ese tipo de gestión son elevados, sin mencionar los aprendizajes específicos de los trabajadores), por lo que es más razonable atribuir este tipo de comportamiento a fallas en los procesos de gestión, particularmente cuando los marcos regulatorios permiten corregir esas situaciones con sanciones leves. Desde la perspectiva estadística, estas fallas del registro incrementan de manera espuria los valores de los indicadores de rotación del empleo, por lo que, para corregir esta situación, el OEDE de Argentina publica series de empleo trimestrales donde los promedios se estiman utilizando los valores de las declaraciones válidas (no 0). Este procedimiento tiene un efecto idéntico a imputar los datos perdidos utilizando el promedio trimestral de los datos válidos para la misma firma en el mismo trimestre.

Imputación de ventas para las Microempresas. En el INEI de Perú, se tomó la decisión de imputar los ingresos de los contribuyentes cuando son personas físicas y no registran ventas, asignándoles un promedio mínimo. El equipo de funcionarios del DANE, que hace uso del registro administrativo, se encarga de realizar los procedimientos de imputación, tomando información auxiliar de registros que tienen relación con la temática y las mismas unidades de observación. Existen otras técnicas utilizadas como, por ejemplo, el uso de promedios móviles.

v. Uso de taxonomías

El poder analítico de los clasificadores se incrementa de manera notable a partir de la incorporación de taxonomías de distintos tipos. Para que las taxonomías puedan ser adoptadas, deben corresponder a realidades económicas similares a las del país del registro (para que sus

conclusiones puedan ser generalizadas o, caso contrario, adaptarse a lo local). También resulta central que la información de clasificación del registro esté en condiciones adecuadas.

b) Integración de fuentes en sistemas de información

Cada registro aislado cuenta con una cantidad muy limitada de variables. Si bien normalmente cuentan con un elevado número de observaciones, éstas se refieren a pocas dimensiones, por lo que cada registro aislado tiene una utilidad acotada para el análisis económico o social. En la mayoría de los países, las personas y las empresas se identifican con códigos tributarios únicos, utilizados por todos los niveles de gestión, que permiten la integración de los RA (a nivel de microdatos) en un verdadero sistema de información, que incluye también al resto de las operaciones estadísticas (Censos y Encuestas) disponibles. La dificultad para desarrollar este tipo de sistemas reside en la necesidad de un andamiaje institucional suficiente para garantizar la participación de distintos organismos públicos con responsabilidad en el sector y en las estadísticas, como se muestra en la **Figura 3**.

Figura 3. Construcción de sistemas integrados de registros, censos y encuestas



Fuente: Elaboración propia con base en OEDE del Ministerio de Trabajo, Empleo y Seguridad Social de Argentina (OEDE, 2017) e INEC de Ecuador (INEC, 2017)

Un sistema integrado de información puede ser muy potente, desde una perspectiva analítica, si consideramos que mientras más dimensiones se sumen al mismo, éste gana en poder explicativo y consistencia. Asimismo, la integración de fuentes otorga mayor versatilidad al sistema, ya que resulta posible seleccionar dimensiones provenientes de diferentes fuentes (vinculadas a nivel de microdatos) para responder a preguntas de investigación específicas. En este sentido, se puede mencionar brevemente algunas fuentes posibles de integrar dentro de

un sistema de información que pueda estar orientado al estudio de la política de productividad, desarrollo productivo y empleo, por ejemplo: padrones de contribuyentes, que contienen datos de identificación de las unidades; registros de la seguridad social, con nóminas de trabajadores y sus remuneraciones; registros de impuestos internos, con ventas y otras medidas de facturación; registros de comercio exterior; balances presentados por las sociedades a su organismo regulador; bases de deudores y créditos del sistema financiero; registros de proyectos de inversión, que incluyen nóminas de inversiones extranjeras directas; entre otros.

La combinación de dimensiones contenidas en esos registros permitiría responder a preguntas referidas a distintas problemáticas, por ejemplo, a los determinantes de la productividad, al vínculo entre el crédito y el desarrollo local, a la importancia del comercio exterior en el mercado de trabajo, entre muchas otras. Esta información se puede complementar con las operaciones estadísticas tradicionales, como son los censos económicos y las encuestas temáticas específicas (innovación, temas ambientales, incorporación de nuevas tecnologías, etcétera). Los registros administrativos de personas también pueden incorporarse, integrando sistemas “empleado – empleador”. Por otra parte, el registro civil de las personas, los registros administrativos del sistema educativo y de los sistemas de formación profesional permiten, por ejemplo, conocer qué perfiles de trabajadores se desarrollan en qué perfiles de empresas; mientras que los registros de beneficiarios de políticas públicas, anexados a este sistema, permiten desarrollar evaluaciones de impacto.

c) Construcción de sistemas dinámicos: Series longitudinales

El Estado ejecuta continuamente procesos administrativos, generando importantes volúmenes de datos. Esto permite que el sistema de información se pueda actualizar de manera permanente con un costo de gestión relativamente bajo. Los mismos códigos de identificación, que se usan para vincular distintas fuentes y conformar sistemas, se pueden utilizar para construir bases de datos longitudinales, agregando datos de períodos consecutivos para un mismo agente económico. No obstante, este proceso exige transformaciones importantes en la estructura de los registros y validaciones de la información en términos dinámicos.

i. Fechas de inicio y fin de los eventos

Para medir **el tiempo**, es necesario buscar la consistencia entre distintas fuentes de datos que indiquen el inicio y el fin de un evento. En ocasiones estas fechas no están cargadas de manera correcta en los registros, por lo que es preciso verificar si el valor registrado es consistente con el valor empírico. Por ejemplo, en el caso del inicio de la actividad de las empresas se debe verificar que no existan declaraciones de empleo o de ventas previas al inicio. De la misma manera, ante el cese de actividades, se debe verificar que no existan declaraciones posteriores a la fecha de cierre. En caso de encontrar inconsistencias, lo adecuado sería tomar los valores empíricos.

ii. Cambios en los ID de los agentes

Los cambios en la razón social de las firmas responden a necesidades administrativas, pero dificultan la construcción de bases longitudinales. De la misma manera, cambios en la propiedad de las empresas, fusiones y adquisiciones, frecuentemente generan cambios en los identificadores de los agentes económicos, que no reflejan eventos demográficos de nacimientos y cierres de firmas, sino cambios en su denominación. En general, cuando las

empresas involucradas son de cierta envergadura, puede encontrarse información en la prensa acerca de este tipo de eventos, pero cuando se trata de empresas más pequeñas este tipo de situaciones suelen quedar invisibilizadas. Por lo tanto, comparar los nombres de las empresas y sus direcciones puede resultar una información importante a la hora de construir datos relevantes.

Una buena práctica desarrollada por el OEDC de Argentina consiste en identificar estas situaciones a partir de las “migraciones colectivas de trabajadores” desde una empresa que cierra hacia otra que se constituye o que presenta un crecimiento extraordinario. En estos casos, se considera que las firmas se encuentran atravesando por algún proceso de cambio de sociedades (fusión, adquisiciones, escisiones, etc.) y que los puestos de trabajo destruidos y creados son los mismos puestos transferidos desde una hacia otra empresa.

5. Otros aspectos para enriquecer los sistemas de información: georreferenciación y análisis de ocupaciones

Basándose en las experiencias analizadas, pueden identificarse dos áreas que ofrecen oportunidades para desarrollar nueva información y conocimientos a partir de los registros administrativos: la georreferenciación y el uso de clasificadores ocupacionales. En ambos casos, los institutos de estadísticas muestran avances importantes en la aplicación de estas herramientas a otras operaciones estadísticas como censos y encuestas, pero resultan aún incipientes las experiencias sobre los RA. Es importante avanzar en la recolección de los domicilios en formatos adecuados durante los procesos administrativos, así como también de la información relevante acerca de las ocupaciones que desempeñan las personas. El potencial analítico de estos datos para el diseño y gestión de políticas públicas es sumamente importante.

A. Georreferenciación de registros

La georreferenciación de las estadísticas es una operación habitual dentro de los sistemas de información geográfica (SIG). Tradicionalmente utilizada para difundir resultados de censos, se está comenzando a aplicar también para difundir información de registros administrativos. No obstante, el principal obstáculo que suele enfrentarse a la hora de encarar un análisis estadístico espacial a partir de datos administrativos es que las direcciones (de las personas y de los establecimientos productivos) no siempre están cargadas de manera adecuada en los registros. Antes de empezar a georreferenciar la información hay que realizar una tarea de depuración y homogeneización en los domicilios. La transformación de direcciones postales a coordenadas no resulta trivial al gestionar grandes volúmenes de datos, en particular en países de gran tamaño.

La base cartográfica digital es otro elemento importante que suele estar disponible en los institutos de estadísticas, sin embargo, para las instituciones más pequeñas, no contar con una base cartográfica puede ser un obstáculo. Es preciso contar con un software que permita analizar y visualizar la información geográfica e integrar la base cartográfica. A pesar de las dificultades, las instituciones consultadas están realizando avances importantes en este sentido, pudiendo destacarse los siguientes casos:

- El INEI de Perú ha georreferenciado numerosos registros, aunque no la totalidad.
- En el DANE de Colombia, la Dirección de Geoestadística hace ejercicios de georreferenciación para localizar establecimientos, siempre y cuando se cuente con la información de dirección o coordenadas.¹¹
- El IPEA de Brasil se encuentra en el proceso de georreferenciar todos los registros administrativos a los que tienen acceso. Para ello, han comprado una plataforma que contiene mapas de todas las calles de Brasil.
- Impuestos internos de Uruguay tiene previsto implantar georreferencias en los RUT a fin del presente año. Esto comprende la georreferenciación de todo el registro de domicilios actuales y la georreferenciación en tiempo real de todas las modificaciones

¹¹ <https://geoportal.dane.gov.co/v2>

que se registren en el RUT (inscripciones y cambios de domicilios). Una vez dispuesta esta funcionalidad podría utilizarse en la producción de estadísticas.

Recuadro 5. Buena práctica de integración y de georreferenciación de registros: El Sistema Integrado de Estadísticas de la Criminalidad y Seguridad Ciudadana de INEI de Perú

El sistema integra al conjunto de instituciones que conforman el Comité Estadístico Interinstitucional de la Criminalidad (CEIC), compuesto por organismos como el Poder Judicial, el Ministerio Público del Perú, el Ministerio de Justicia y Derechos Humanos, la Policía Nacional del Perú, el Instituto Nacional Penitenciario, entre otras entidades vinculadas a esta temática, brindándoles a cada institución información relevante del conjunto de organismos.

Es una herramienta de apoyo al análisis, diseño y evaluación de políticas públicas en materia de criminalidad y seguridad ciudadana, tanto para las autoridades a nivel nacional, regional y local, como para otras instituciones vinculadas a esta problemática. Proporciona información en distintos niveles de agregación, desde cuadros estadísticos, gráficos y mapas temáticos (georreferenciados) sobre indicadores de victimización; hasta indicadores puntuales, como el número de denuncias registradas en las dependencias policiales y comisarías -hasta el nivel de manzanas-, el número de casos ingresados a las fiscalías, procesos judiciales, sentencias y detenciones emitidas por el Poder Judicial, así como también las características de la población penal del país. Esta información se complementa con datos sobre establecimientos por giro de negocios, lugares de interés, estrato de tamaño de población y por estratos de pobreza, entre los principales.

Es un programa con pocos precedentes que muestra un ejemplo de cooperación entre instituciones y de trabajo coordinado y responsable orientado a una mejor atención al ciudadano. Recientemente el proyecto ha sido distinguido con el Premio Especial de Datos Abiertos en la Gestión Pública por el Sistema Integrado de Estadísticas de la Criminalidad y Seguridad Ciudadana.

Fuente: Elaboración propia con base en INEI¹²

B. Clasificador de ocupaciones

La Clasificación Internacional Uniforme de Ocupaciones (CIUO), fue desarrollada por la Organización Internacional del Trabajo (OIT) para organizar la información de trabajo y empleo. Pertenece a la familia internacional de las clasificaciones económicas y sociales y, en sus diferentes revisiones, es la base de muchas clasificaciones de ocupación nacionales, dependiendo de la fecha en que fueron desarrolladas¹³.

La primera versión de la Clasificación Internacional Uniforme de Ocupaciones fue adoptada en 1957 por la Conferencia Internacional de Estadísticos del Trabajo (CIUO-58), versión que fue reemplazada en el año 1966 (CIUO-68). La tercera versión se desarrolló 20 años después, en 1987 (CIUO-88). Considerando los importantes cambios que tuvieron en lugar en el mundo del trabajo desde 1988, la CIUO fue nuevamente actualizada en 2007 (CIUO-08). En su versión actualizada, el clasificador conserva los conceptos y la estructura de las versiones anteriores.

¹² <http://www.inei.gob.pe/prensa/noticias/inei-gano-el-premio-de-datos-abiertos-en-la-gestion-publica-otorgado-en-la-xiii-edicion-del-concurso-buenas-practicas-en-gestion-publica-9895/> (Consultado el 8/10/2017)

¹³ Para más información, dirigirse a <http://unstats.un.org/unsd/class>.

a) Conceptos que utiliza el clasificador

En la revisión CIUO-08, se entiende por **empleo** al conjunto de tareas y cometidos desempeñados por una persona, o que se prevé que ésta desempeñe, para un empleador particular, incluido el empleo por cuenta propia. Se entiende por **ocupación**, un conjunto de empleos cuyas principales tareas y cometidos se caracterizan por un alto grado de similitud. Una persona puede estar asociada con una ocupación a través del empleo principal desempeñado en ese momento, un empleo secundario o un empleo desempeñado anteriormente.

Por otro lado, se entiende por **competencia** a la capacidad de llevar a cabo tareas correspondientes a un determinado empleo. A los efectos de la CIUO-88 se utilizan dos dimensiones de competencias para disponer las ocupaciones en grupos; se trata del nivel de competencias y de la especialización de las competencias.

El **nivel de competencias** se define en función de la complejidad y diversidad de tareas y cometidos cuyo desempeño corresponde a una ocupación. Se mide operacionalmente considerando uno o más de los siguientes elementos: (i) La naturaleza del trabajo desempeñado; (ii) El nivel de enseñanza formal necesaria; (iii) La cantidad de formación informal en el empleo y/o la experiencia previa.

La **especialización de las competencias** se considera desde el punto de vista de cuatro nociones conceptuales: (i) El campo de conocimiento necesario (ii) Las herramientas y la maquinaria utilizada (iii) Los materiales sobre los que se trabaja o con los que se trabaja; y (iv) Los tipos de bienes y servicios producidos.

b) ¿Cómo funciona el clasificador?

La CIUO-08 agrupa los empleos en ocupaciones y en grupos más generales, principalmente a partir de la similitud entre las competencias necesarias para desempeñar las tareas de los empleos. En la definición de los grupos de la CIUO-88 se utilizan dos dimensiones del concepto de competencia: el **nivel de competencias y la especialización**. Cabe destacar que se centra en las competencias necesarias para desempeñar las tareas de una ocupación determinada, y no en si este trabajador, con una ocupación particular, está más o menos capacitado que otro trabajador con la misma u otras ocupaciones.

Cuadro 2. Correspondencia de los grandes grupos de la CIUO-08 con los niveles de competencias

Grandes grupos de la CIUO-08	Nivel de competencias
1 – Miembros del poder ejecutivo y de los cuerpos legislativos y personal directivo de la administración pública y de empresas	3 y 4
2 – Profesionales científicos e intelectuales	4
3 – Técnicos y profesionales de nivel medio	3
4 – Empleados de oficina	2
5 – Trabajadores de los servicios y vendedores de comercios	
6 – Agricultores y trabajadores calificados agropecuarios y pesqueros	
7 – Oficiales, operarios y artesanos de artes mecánicas y de otros oficios	
8 – Operadores de instalaciones y máquinas y montadores	1
9 – Trabajadores no calificados	
0 – Ocupaciones militares	1 y 4

Fuente: Elaboración propia con base en ILO

Nivel de Competencias 1: típicamente requieren el desempeño de tareas simples y rutinarias, principalmente físicas o manuales. También pueden requerir la utilización de herramientas que se utilizan manualmente, como las palas o un equipo eléctrico simple, como la aspiradora. Abarcan tareas como la limpieza, la excavación, el levantamiento y acarreo manuales de material, la selección, almacenamiento o reunión de productos manualmente (a veces en el contexto de operaciones mecanizadas): operación de vehículos no motorizados y recolección de frutos y vegetales.

Nivel de competencias 2: suelen incluir el desempeño de tareas como la operación de maquinaria de tipo electrónico, la conducción de vehículos, el mantenimiento y reparación del equipo eléctrico y mecánico, y la manipulación, ordenamiento y almacenamiento de información. Es esencial poseer la aptitud de leer información, como instrucciones de seguridad, establecer por escrito registros del trabajo realizado y desempeñar con precisión cálculos aritméticos simples. Requieren una capacidad relativamente amplia de manejo de las letras y operaciones numéricas, y buenas competencias en materia de comunicación con las personas.

Nivel de Competencias 3: el desempeño de tareas técnicas y prácticas complejas que exigen un conjunto de conocimientos de hechos, técnicos y de procedimientos en un área especializada. Requieren un elevado nivel de conocimiento de las letras y las matemáticas, así como sólidas aptitudes en materia de comunicación personal. Estas competencias pueden incluir la capacidad de comprender un material escrito complejo, preparar informes basados en hechos. Los conocimientos y competencias necesarios suelen obtenerse cursando estudios en una institución de educación superior, una vez completada la educación secundaria, durante un período de uno a tres años. En algunos casos una experiencia laboral pertinente y una experiencia prolongada en el empleo pueden reemplazar la enseñanza formal.

Nivel de Competencias 4: suelen abarcar el desempeño de tareas que exigen la solución de problemas complejos y la adopción de decisiones, sobre la base de un amplio conocimiento teórico y fáctico en una especialización determinada. El análisis y la investigación para desarrollar los conocimientos humanos en un determinado ámbito, el diagnóstico y el tratamiento de una enfermedad, la transmisión de conocimientos a terceros, el diseño de estructuras o maquinarias o procesos para la construcción y la producción. Suelen exigir un sólido conocimiento de las letras y las matemáticas. Generalmente, los conocimientos y competencias requeridos se obtienen después de haber cursado estudios en una institución de enseñanza superior durante un período de tres a seis años conducente a la obtención de un primer diploma o título superior.

d) [Uso de los clasificadores de ocupaciones en las instituciones consultadas](#)

Habitualmente, los sistemas estadísticos relevan información a nivel de las ocupaciones de los trabajadores en los censos de población y en encuestas que estudian el mercado de trabajo, como las encuestas de hogares y las que se realizan a trabajadores en las empresas. Sin embargo, en la actualidad, pocos registros administrativos relevan este tipo de información. En efecto, entre las instituciones consultadas sólo el OEDE del Ministerio de Trabajo de Argentina (“Mi Simplificación”) y el IPEA de Brasil (la RAIS contiene información de ocupaciones). Los institutos de estadísticas consultados no cuentan con información de registros sobre las ocupaciones, brindando sólo información de censos de población y de encuestas de empleo.

- El DANE de Colombia cuenta con versiones de la CIUO-88 y 08 adaptadas a ese país,
- El INEI de Perú, está implementando una versión adaptada del CIUO-08.

- El INEC de Ecuador, en el año 2015, elaboró la nueva Clasificación Nacional de Ocupaciones CIUO-08 adaptada al país, la misma que reemplazó a la CIUO-88.

6. Reflexiones finales

Los registros administrativos son grandes volúmenes de datos que se producen y almacenan como resultado del desarrollo de la gestión que realizan distintas áreas de gobierno. La creciente incorporación de tecnología y la mayor disponibilidad de herramientas para la gestión de la información (hardware y software cada vez más potentes), convierten a estas fuentes en un insumo muy importante para la construcción de sistemas de información orientados al apoyo de la política pública. Utilizados adecuadamente, tienden a constituir el corazón de los sistemas estadísticos modernos. Sin embargo, no se debe olvidar que son datos recolectados con fines administrativos y que, si se utilizan en su estado “puro”, pueden brindar información incorrecta o sesgada. Para evitar este tipo de situaciones, es preciso realizar un trabajo previo de adecuación, controlando y corrigiendo ciertos aspectos.

El proceso de aprovechamiento de los registros para fines estadísticos, y como herramienta que apoya la gestión de las políticas públicas, requiere tres condiciones necesarias. Primero, contar con acuerdos institucionales adecuados en el Estado. Segundo, contar con capacidades técnicas en las instituciones de estadísticas que desarrollan el proceso de transformación de los datos administrativos en estadísticos. Y, por último, contar con un marco legal adecuado, que permita el acceso de los usuarios finales a la información estadística, para su transformación en conocimientos.

Los acuerdos institucionales que permitan la circulación de la información entre agencias del propio gobierno son condición indispensable. Esto es así, porque para realizar un proceso de este tipo las agencias responsables de la gestión de los registros administrativos deben estar en condiciones de brindar acceso a datos de origen fiscal (cuya confidencialidad suele estar fuertemente protegida en todos los países) a una institución productora de estadísticas que, a su vez, se comprometa a realizar un uso seguro de los datos hasta que se encuentren en condiciones de ser publicados. Es importante que la agencia responsable de los registros también debe estar dispuesta a informar sobre los procesos de gestión, cuando esto resulte necesario.

La institución productora de estadísticas debe contar con capacidades técnicas para desarrollar procesos de trabajo complejos sobre grandes volúmenes de datos, que incluyen la evaluación y transformación de información de origen fiscal, no siempre de buena calidad. Entre estas actividades, la “evaluación conceptual” de los registros implica analizar si los criterios utilizados para definir las variables, la cobertura, los sistemas de codificación utilizados, la periodicidad, son consistentes con los criterios estadísticos. Los aspectos del proceso de gestión pueden afectar la precisión de los datos, por lo que deben conocerse en detalle. El análisis estadístico de los datos implica analizar su consistencia interna (entre distintas variables y las mismas variables a lo largo del tiempo) y externa (comparándolos con otras fuentes, como encuestas, censos y otros registros). Esas actividades se orientan a detectar “fallas” que deben ser corregidas aplicando metodologías tradicionales o innovadoras, o documentarse e informarse, al menos, en aquellos casos que no sea posible corregirse el problema.

Finalmente, resulta fundamental contar con un marco legal y con herramientas tecnológicas que permitan brindar acceso a los microdatos, en condiciones seguras, a la comunidad (sector académico, investigadores, responsables de la gestión de las políticas, periodistas, entre otros),

ya que son los estudios los que transforman la información estadística en conocimientos útiles para la sociedad. Es decir, el aprovechamiento de los registros se completa recién en esta etapa.

7. Bibliografía

BPS Uruguay (2017) Análisis con bases de datos en la Asesoría Económica y Actuarial. Workshop CAF - Buenos Aires (ARG). 8 de junio de 2017.

Comisión Económica para América Latina y el Caribe (CEPAL) (2003). Registros administrativos, calidad de los datos y credibilidad pública: presentación y debate de los temas sustantivos de la segunda reunión de la Conferencia Estadística. Santiago de Chile.

Centro de Estudios Distributivos, Laborales y Sociales (CEDLAS) (2017) Estadísticas de informalidad.

Corseuil C.H., Mation L (2017). The use of administrative data in Brazil: La perspectiva desde IPEA. Workshop CAF -Buenos Aires (ARG). 8 de junio de 2017.

Departamento Administrativo Nacional de Estadísticas (DANE) (2010). Metodología para el fortalecimiento de registros administrativos. *Colección documentos – Actualización 2010 Núm. 99*. Bogotá. 2010.

Departamento Administrativo Nacional de Estadísticas (DANE) (2015). Plan de Fortalecimiento de Registros Administrativos. *Herramientas para la gestión territorial más efectiva N° 5*. Bogotá. 2015.

Eurostat (2003). Quality Assessment of Administrative Data for Statistical Purposes.

Eurostat (2010). Manual de recomendaciones de los Registros de Empresas. <https://unstats.un.org/unsd/EconStatKB/KnowledgebaseArticle10171.aspx> (consultado el 9/10/2018)

Ferreira M. (2017). Uso de datos administrativos para la Promoción de Exportaciones e Inversiones en Uruguay. Workshop CAF - Buenos Aires (ARG). 8 de junio de 2017.

Fortin, Marco. (2000). Linee guida metodologiche per rilevazioni statistiche. Nozioni metodologiche di base e pratiche consigliate per rilevazioni statistiche dirette o basate su fonti amministrative.

Garzón, N., Kulfas, M., Palacios, J.C., Tamayo, D. (2016). Evolución del sector manufacturero de Ecuador. *Cuaderno de Trabajo,1*, INEC. Quito.

Instituto Nacional de Estadística y Censos (INEC) (2017). Uso y accesibilidad de registros administrativos Ecuador. Workshop CAF - Buenos Aires (ARG). 8 de junio de 2017.

Instituto Nacional de Estadística y Geografía (INEGI) (2012). Proceso estándar de aprovechamiento de registros administrativos. México.

Krell R. (2017). Datos administrativos de firmas en Chile. Ministerio de Economía, Fomento y Turismo. Workshop CAF - Buenos Aires (ARG). 8 de junio de 2017.

Miranda Toledo, R. (2017). Experiencia de la Comisión Nacional de Productividad en el Uso de Datos Administrativos. Workshop CAF - Buenos Aires (ARG). 8 de junio de 2017.

Naser, A. y Concha G. (2017). El Gobierno Electrónico en la Gestión Pública. Serie Gestión Pública N° 73. ILPES. Santiago de Chile.

Organización para la Cooperación y el Desarrollo Económicos (OECD). (2015). Recomendación del Consejo de la OCDE sobre Buenas Prácticas Estadísticas.

Organización para la Cooperación y el Desarrollo Económicos (OECD). (2011). Quality framework and guideline for OECD statistical activities.
[http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=std/qfs\(2011\)1&doLanguage=en](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?cote=std/qfs(2011)1&doLanguage=en); (consultado el 9/10/2018)

Observatorio de Empleo y Dinámica Empresarial (OEDE). (sin datar). Metodologías de los boletines estadísticos de empleo, empresas, remuneraciones y estadísticas provinciales.

Observatorio de Empleo y Dinámica Empresarial (OEDE). (2017). El Observatorio de Empleo y Dinámica Empresarial (OEDE) de Argentina. Workshop CAF - Buenos Aires (ARG). 8 de junio de 2017.

Prada Lombo C.F. (2017). Acceso a la información estadística de Colombia: DANE para todos. Workshop CAF - Buenos Aires (ARG). 8 de junio de 2017.

Vilhuber, L. (2017). Confidentiality protection and physical safeguards. Presentación disponible en: <https://ecommons.cornell.edu/handle/1813/46207>

Wallgren, A., & Wallgren, B. (2007). Register-based statistics: administrative data for statistical purposes (Vol. 553). John Wiley & Sons.

Sitios consultados

<http://ec.europa.eu/eurostat>

<http://www.bps.gub.uy/>

<http://www.cedlas.econo.unlp.edu.ar/wp/estadisticas/sedlac/>

<http://www.dane.gov.co/>

<http://www.dgi.gub.uy/>

<http://www.ecuadorencifras.gob.ec/institucional/home/>

<http://www.inei.gob.pe/>

<http://www.trabajo.gob.ar/left/estadisticas/oede/index.asp>

<https://data.oecd.org/>

Funcionarios consultados:

Cra. Mariella Lazo. Gerente de Sector Valuaciones Actuariales. Asesoría Económica y Actuarial. Banco de Previsión Social (BPS) de Uruguay.

Gustavo González Amilivia. Adjunto - División Planificación, Estudios y Coordinación. Dirección General Impositiva. Uruguay.

Peter Abad. Director Técnico. Oficina Técnica de Planificación, Presupuesto y Cooperación Técnica. Instituto Nacional de Estadística e Informática. Perú.

Rodrigo Krell Loy. Jefe de División. División Política Comercial e Industrial. Ministerio de Economía, Fomento y Turismo. Chile.

Lucas Ferreira Mation. Chefe da Assessoria Técnica. Gabinete da Presidência. IPEA Instituto de Pesquisa Econômica Aplicada. Brasil.

Ana Paola Gómez A. Directora. Dirección de Regulación, Planificación, Normalización y Estandarización- DIRPEN. Colombia.

Victoria Castillo. Coordinadora del Observatorio de Empleo y Dinámica de Empresas. Ministerio de Trabajo, Empleo y Seguridad Social. Argentina.

8. Anexo: Fichas de análisis de casos

En su institución se encuentran utilizando registros administrativos como insumos para la producción de estadísticas.

- 1) Por favor, indique cuáles registros está utilizando actualmente y si es que tienen previsto incorporar otros registros que aún no está utilizando.

Registros	Utilizan	Está previsto utilizar
Padrones de contribuyentes (rama de actividad, domicilio, impuestos a los que contribuye, etc.)		
Registros de personas (educación, nacimiento, sexo, etc.)		
Nóminas de trabajadores declaradas al sistema de seguridad social (empleo y salarios)		
Liquidaciones de impuestos internos		
Balances de sociedades		
Registros del Banco Central (depósitos y/o deuda)		
Registros de las empresas de luz/agua, con los consumos de las empresas		
Registros de comercio exterior (exportaciones/ importaciones)		
Inversión extranjera directa		
Beneficiarios de políticas de fomento/ programas/ proyectos (especifique cuales)		
Registros empleados/ empleadores		
Otros		

- 2) Cuando tuvieron que incorporar los registros a la producción de estadísticas, ¿realizaron evaluaciones de calidad sobre los registros? ¿De qué tipo? (puede describirlas)
- ¿Identificaron diferencias entre las definiciones administrativas y definiciones estadísticas?
 - ¿Identificaron variables que están truncadas (máximos en las remuneraciones cargadas, por ejemplo)
 - ¿Compararon la cobertura del registro (falta de algún régimen especial de grandes contribuyentes, o sectores o regiones ausentes en el registro principal)?
 - ¿Verificaron si se produjeron cambios en los sistemas de codificación, a lo largo del tiempo?
 - ¿Verificaron si los registros tienen sistemas de validación internos para la carga de datos?
 - ¿Analizaron la consistencia de los datos de registro, comparándolos con fuentes externas (censos, encuestas, consultas a cámaras)?
 - ¿Pudieron comparar los registros con otras fuentes a nivel de microdato? (directorios de empresas para verificar ramas, por ejemplo, o con información de cámaras, llamadas telefónicas para verificar que la información esté bien, pequeñas encuestas para validar información, etc.)
 - Otras:
- 3) ¿Realizan correcciones de errores y de inconsistencias de los registros antes de utilizarlos como fuentes estadísticas?
- ¿Modifican códigos como los de las ramas de actividad, localización y otros?
 - ¿Aplican ex post normas de validación? ¿Para eliminar valores inconsistentes? ¿Cómo lo hacen?
 - ¿Eliminan casos duplicados?
 - Imputación de datos faltantes:
 - Casos perdidos, por demoras en la presentación de declaraciones a la autoridad fiscal
 - ¿Corrigen los códigos identificadores de las empresas (RUN, RUT, CUIT)?
 - Empresas que cambian de código de identificación, pero que es son la misma empresa,
 - Empresas que aparecen con muchos códigos y deben agregarlas (Multirut)
 - ¿Identifican fusiones y adquisiciones?
 - Agregan códigos de grupos económicos (muchas empresas de un mismo dueño)
 - Otros
- 4) ¿Aplican a los datos taxonomías de ramas, ocupaciones, tipos de productos?
- Tamaños:
 - Taxonomías de intensidad de uso de factores y de intensidad tecnológica: pueden incorporarse recodificando la rama de actividad
 - Taxonomías de productos (bases de comercio)
 - Taxonomías de ocupaciones:
 - ¿Georeferencian la información?
 - Otras