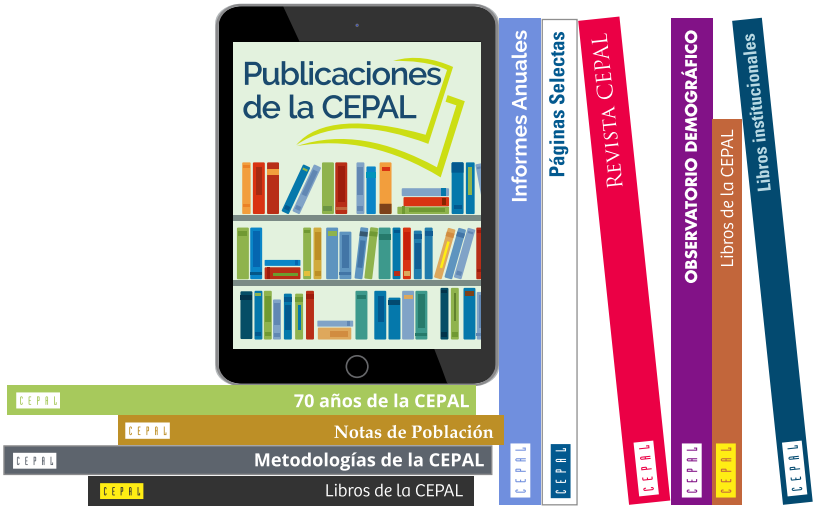
Panorama del uso de tecnologías digitales para la recolección de datos sobre el comercio de servicios

Íñigo Herguera

Naciones Unidas
CEPAL
75 años
Trabajando por un futuro productivo, inclusivo y sostenible
25 SINCE 1999
FEALAC
FORUM FOR EAST ASIA-LATIN AMERICA COOPERATION

Gracias por su interés en esta publicación de la CEPAL



Si desea recibir información oportuna sobre nuestros productos editoriales y actividades, le invitamos a registrarse. Podrá definir sus áreas de interés y acceder a nuestros productos en otros formatos.

Deseo registrarme

www.cepal.org/es/publications

www.instagram.com/publicacionesdelacepal

www.facebook.com/publicacionesdelacepal

www.issuu.com/publicacionescepal/stacks

www.cepal.org/es/publicaciones/apps

NACIONES UNIDAS

CEPAL

Documentos de Proyectos

Panorama del uso de tecnologías digitales para la recolección de datos sobre el comercio de servicios

Íñigo Herguera

NACIONES UNIDAS

CEPAL

SINCE 1999

FEALAC

FORUM FOR EAST ASIA-LATIN AMERICA COOPERATION

La redacción de este documento estuvo a cargo de Íñigo Herguera, Consultor de la División de Comercio Internacional e Integración de la Comisión Económica para América Latina y el Caribe (CEPAL). El documento fue preparado en el marco de una asistencia técnica de la CEPAL a la Alianza del Pacífico y forma parte del proyecto "Apoyar el fortalecimiento de políticas para monitorear y promover la participación de las MIPYMES en el comercio electrónico transfronterizo de bienes y servicios para una respuesta y recuperación inclusiva y sostenible ante la pandemia en Asia Oriental y América Latina" del Foro de Cooperación América Latina - Asia del Este (FOCALAE). Se agradece a las instituciones que respondieron a la encuesta realizada en el contexto de la preparación de este documento y que hicieron comentarios. Se agradece asimismo a Javiera Arteaga, Sebastián Herreros y Nanno Mulder, Oficiales de Asuntos Económicos de la División de Comercio Internacional e Integración de la CEPAL, por sus sugerencias.

Las Naciones Unidas y los países que representan no son responsables por el contenido de vínculos a sitios web externos incluidos en esta publicación.

No deberá entenderse que existe adhesión de las Naciones Unidas o los países que representan a empresas, productos o servicios comerciales mencionados en esta publicación.

Las opiniones expresadas en este documento, que no ha sido sometido a revisión editorial, son de exclusiva responsabilidad del autor y pueden no coincidir con las de la Organización o las de los países que representa.

Publicación de las Naciones Unidas
LC/TS.2023/170
Distribución: L
Copyright © Naciones Unidas, 2023
Todos los derechos reservados
Impreso en Naciones Unidas, Santiago
S.23-00989

Esta publicación debe citarse como: Í. Herguera, "Panorama del uso de tecnologías digitales para la recolección de datos sobre el comercio de servicios", *Documentos de Proyectos* (LC/TS.2023/170), Santiago, Comisión Económica para América Latina y el Caribe (CEPAL), 2023.

La autorización para reproducir total o parcialmente esta obra debe solicitarse a la Comisión Económica para América Latina y el Caribe (CEPAL), División de Documentos y Publicaciones, publicaciones.cepal@un.org. Los Estados Miembros de las Naciones Unidas y sus instituciones gubernamentales pueden reproducir esta obra sin autorización previa. Solo se les solicita que mencionen la fuente e informen a la CEPAL de tal reproducción.

Índice

Introducción 7

I. Tecnologías digitales para la recolección y procesamiento de datos: una reseña 9
- A. Datos de Posicionamiento Móvil (MPD) 10
- B. El Sistema de Identificación Automática (AIS) como fuente de datos 14
  - 1. Aspectos metodológicos 14
  - 2. Aplicaciones 16
  - 3. Combinación de bases de datos 16
  - 4. Manejo de datos 17
  - 5. Validación de resultados con otras fuentes 17
- C. 5G e Internet de las Cosas (IoT) 18
  - 1. El IoT y la industria del transporte internacional y logística 20
  - 2. La logística conectada 22
- D. El comercio electrónico transfronterizo 22
  - 1. Comercio electrónico y comercio internacional 23
  - 2. Distinción entre bienes y servicios en el comercio internacional y el comercio electrónico 24
  - 3. Clasificaciones del comercio electrónico y las estadísticas oficiales de comercio 24
  - 4. Comercio electrónico, sus fuentes y retos de medición 25
- E. Uso de datos de plataformas 26

II. Mejores prácticas en el uso de nuevas tecnologías en la recopilación de datos relacionados con el comercio internacional: experiencias internacionales 29
- A. Experiencias de Big Data de ONE 29
  - 1. Acuerdo de la autoridad reguladora y de competencia de España (CNMC) con entidades de medios de pago 31
  - 2. Acuerdo BBVA-OCDE 31
  - 3. La medición del comercio digital en Argentina 32

B. Experiencias de Big Data en áreas concretas 33
1. Turismo 33
2. Experiencias de explotación de datos AIS 34

III. Los desafíos en la explotación de Big Data de movilidad 39
A. Retos administrativos en acceder a los datos de movilidad (MPD) 39
B. Desafíos técnicos del procesamiento de datos y retos metodológicos 41
C. Datos de referencia, evaluación y estimación: nuevos métodos para el tratamiento de Big Data 43
1. Sesgos en estudios de población a partir de Big Data 44
2. Ejemplo: obtención de indicadores a partir de Big Data y comparación con indicadores tradicionales 44

IV. Resultados del cuestionario sobre el uso de nuevas tecnologías digitales para la compilación y procesamiento de estadísticas de comercio de servicios en los países de la Alianza del Pacífico 47
A. Uso de Big data 47
B. Uso de IoT o IA 48
C. Síntesis de resultados del cuestionario 49

V. Conclusiones y recomendaciones 51

Bibliografía 55

Anexos 55
Anexo 1 60
Anexo 2 65

Cuadros

Cuadro 1 Fuentes de datos, métodos de colecta y aplicaciones en el comercio internacional 7
Cuadro 2 Campos de información básicos en la colecta de datos de movilidad 13
Cuadro 3 Tratamiento del comercio electrónico transfronterizo en bienes y en servicios según tipo de transacción 24
Cuadro 4 Ejemplos de uso de Big data para el análisis del turismo 34
Cuadro 5 Ejemplos de uso de 5G en puertos, rutas transfronterizas y tráfico 37
Cuadro 6 Ventajas e inconvenientes del uso de MPD para la elaboración de estadísticas de turismo 41

Gráficos

Gráfico 1 Proporción de empresas en países de la OCDE que utilizan IoT según industria, 2021 20
Gráfico 2 Proporción de empresas que utilizan comercio electrónico y proporción de ventas en línea sobre ventas totales en la Unión Europea (27 miembros), 2011-2020 26

Diagramas

Diagrama 1 Colecta, tratamiento y procesamiento de Big Data procedente de operadores de red móvil 12
Diagrama 2 Fuentes de información complementarias utilizadas en la explotación de datos AIS para la estimación de flujos de comercio marítimo internacional 16
Diagrama 3 Redes utilizadas para Internet de las Cosas 22

Acrónimos

AEPD	:	Agencia Española de Protección de Datos
AIS	:	Sistema Automático de Identificación (*Automatic Identification System*)
ANR	:	Autoridades nacionales de regulación
CBDS	:	Centro para Estadísticas de Big Data (*Center for Big Data Statistics*, Países Bajos)
CBS	:	Oficina Central de Estadísticas (*Central Bureau of Statistics,* Países Bajos)
CDR	:	Registros de detalles de las llamadas (*Call Detail Record*)
CRM	:	Sistema de gestion de clientes (*Customer Relationship Management*)
CIIU	:	Clasificación Internacional Industrial Uniforme
CNMC	:	Comisión Nacional de los Mercados y la Competencia
CROS	:	Colaboración en Investigación y Metodología de la Estadística Oficial (*Collaboration in Research and Methodology for Official Statistics*)
EDI	:	Intercambio electrónico de datos (*Electronic Data Interchange)*
eMBB	:	Banda ancha movil Mejorada (*Enhanced Mobile Broadband*)
GPS	:	Sistema Global de Posición (*Global Positioning Sytem*)
GSMA	:	Asociación mundial de operadores móviles (*Global System for Mobile Communications Association*)
FMI	:	Fondo Monetario Internacional
IA	:	Inteligencia Artificial
IMEI	:	Registro único de terminal móvil (*International Mobile Equipment Identity*)
IoT	:	Internet de las Cosas (*Internet of Things*)
IPDR	:	Registro de datos del Protocolo de Internet (*Internet Protocol Data Record*)
mMTC	:	Comunicaciones masivas de máquina a máquina (*Massive Machine-to-Machine Communications*)
MNO	:	Operadores de red móvil (*Mobile Network Operators*)
MPD	:	Datos de Posicionamiento Móvil (*Mobile Positioning Data*)
NACE	:	Nomenclador Estadístico de las Actividades en la Comunidad Europea
NCM	:	Nomenclatura Común de Mercosur
NLP	:	Procesamiento de Lenguaje Natural (*Natural Language Processing*)
OCDE	:	Organización para la Cooperación y el Desarrollo Económico
OMA	:	Organización Mundial de Aduanas
OMC	:	Organización Mundial del Comercio
ONE	:	Oficina Nacional de Estadística
SA	:	Sistema Armonizado
UIT	:	Unión Internacional de Telecomunicaciones
UNCTAD	:	Conferencia de las Naciones Unidas para el Comercio y Desarrollo (*United Nations Conference on Trade and Development*)
ULLC	:	Comunicaciones ultra bajas, confiables y de baja latencia (*Ultra Low Reliable and Low Latency Communications*)
VHF	:	Frecuencia muy alta en la emisión de señales de radiofrecuencia (*Very High Frequency)*
VDES	:	Sistema de intercambio de datos de muy alta frecuencia (*Very High Frequency Data Exchange System*)

Introducción

En el presente estudio se examina el uso de nuevas tecnologías digitales para la recolección de datos sobre el comercio internacional, con énfasis en el comercio de servicios. Con ello se busca identificar oportunidades para complementar las estadísticas tradicionales, compiladas para efectos de la balanza de pagos, con las nuevas fuentes de datos que ofrecen diversas tecnologías digitales. En el cuadro 1 se sintetizan las fuentes y experiencias que se recogen en este informe.

Cuadro 1
Ejemplos de recolección de microdatos y sus aplicaciones en el comercio internacional

Modo de colecta	Fuente	Aplicaciones en el comercio
Datos de Posicionamiento Móvil (MPD)	Redes telecomunicaciones para movilidad	Turismo
Sistema de Identificación Automática (AIS)	Sensores y redes satelitales y móviles	Logística, transporte internacional, gestión de tráficos y puertos
E-commerce	Intermediarios financieros y plataformas	Comercio transfronterizo de bienes y servicios
Internet de las Cosas (IoT)	Sensores y redes telecomunicaciones móviles 4G y 5G	Logística conectada

Fuente: Elaboración propia del autor.

El uso de Big Data y de nuevos métodos de colecta, como las fuentes del Sistema Global de Posición (GPS, por sus siglas en inglés), Datos de Posicionamiento de Móvil (MPD, por sus siglas en inglés), web scraping, Sistema Automático de Identificación (AIS, por sus siglas en inglés) o el scanned data, ha multiplicado el volumen de datos disponible para la medición de la actividad económica, las interacciones sociales o fenómenos naturales. No solo el volumen de datos es de otra magnitud, también el modo de gestionarlo requiere de cambios importantes para cualquier agente que quiera explotar estas posibilidades. Para las Oficinas Nacionales de Estadística (ONE) ello supone un nuevo modo de gestionar sus tareas, lograr acuerdos de colaboración con los agentes que disponen de estos datos y nuevos métodos de análisis.

El resto de este documento se organiza como sigue. Luego de esta introducción, en la sección 2 se presenta una reseña de las principales tecnologías digitales para la recolección y procesamiento de datos. A continuación, en la sección 3 se analizan algunas experiencias internacionales exitosas de uso de nuevas

tecnologías en la recopilación de datos relacionados con el comercio internacional. En la sección 4 se examinan los desafíos que plantea la explotación del Big Data de movilidad. Por último, en la sección 5 se presentan los resultados de un cuestionario enviado a las ONE y bancos centrales de los países de la Alianza del Pacífico sobre el uso de Big Data y nuevas tecnologías digitales en su trabajo de compilación y producción de estadísticas. Las respuestas recibidas indican que el Big Data es aún poco utilizado por dichas instituciones, y que su uso es principalmente exploratorio o experimental.

I. Tecnologías digitales para la recolección y procesamiento de datos: una reseña

Las fuentes tradicionales e institucionalizadas de estadísticas conducidas por las ONE tienen ventajas claras en cuanto a que la metodología es robusta, transparente y cumple con la protección de la privacidad de datos personales o la confidencialidad de ciertos datos. Los datos recogidos periódicamente por las ONE sobre múltiples aspectos de la economía, sociedad o los hábitos de los ciudadanos son datos de referencia para la política pública, los agentes privados y de uso intenso por múltiples agencias internacionales.

Las fuentes tradicionales se basan en dos grandes métodos de recolección de datos: i) registros de agentes, que pueden ser por ejemplo empresas hogares o individuos y los cuales recopilan en ciertas intervenciones estadísticas a todos los agentes existentes; o ii) muestreos, hechos sobre una población concreta objeto de estudios (por ejemplo, empresas) donde la recogida de datos es efectuada con criterios de representatividad y de calidad conocidos y diseñados ex ante. Las ONE suelen utilizar metodologías de recogida de datos, de tratamiento y de explotación bien conocidas, frecuentemente similares de un país a otro y donde la representatividad y la comparabilidad de las mediciones suele estar garantizada.

Las fuentes tradicionales de datos también tienen desventajas, en especial si las comparamos con las nuevas fuentes de datos digitales. Estas últimas utilizan una red, servicio digital o aplicaciones varias para la recogida de datos de alguna actividad, como puede ser la red desplegada por un operador de comunicaciones móviles o una plataforma de comercio electrónico con la cual se recaban muchos campos de información de cada individuo que utilice sus servicios.

Un ejemplo de estadística tradicional que muestra los pros y contras de un modo claro es la encuesta sobre pernoctaciones de turistas, acometida con periodicidad anual por muchas ONE. Esta encuesta consiste en un muestreo aleatorio estratificado a ciertas dimensiones para posterior inferencia, que se realiza sobre una muestra de establecimientos de hospedaje registrados. Este tipo de encuesta es muy costoso por varias razones: la contratación de encuestadores y su formación en el

objetivo y detalles del cuestionario; el diseño de la muestra, la aplicación en el campo, y el tratamiento. La parte encuestada, en este caso hoteles, debe también dedicar recursos para su cumplimiento. Muchos países hacen este muestreo solo anualmente por su alto costo. Otra desventaja es la existencia de individuos que no se encuentran disponibles y que deben ser reemplazados por otros similares, dados unos parámetros de estratificación definidos ex ante. Otro reto es que la muestra, una vez recogida, provee información ex post sobre realidades del pasado. En el caso de estadísticas estructurales esto no es muy grave, pero para otras más coyunturales representa una desventaja.

La recolección de grandes volúmenes de datos a través de herramientas digitales, llamada Big Data, tiene varias ventajas. Los Big Data son recogidos y procesados más rápido que mediante métodos tradicionales ya que utilizan la huella electrónica (ya sea de facturación o de uso de un terminal móvil) que dejan los usuarios. La recolección de datos digitales se puede repetir a un costo relativamente bajo, lo cual redunda en un mayor número y en mejores mediciones y la posibilidad de acometer análisis y estudios posteriores más variados. Dados los volúmenes de datos que pueden llegar a recogerse, el uso de técnicas de Big Data implica un costo de almacenamiento, tratamiento y procesamiento. Sin embargo, también posibilita aplicar métodos estadísticos o de Inteligencia Artificial (IA) con los cuales poder predecir ciertas características sociales, o incluso cambios en el comportamiento social, de un modo más rápido y con precisión.

La recolección tradicional de datos y la recolecta masiva por medios electrónicos pueden ser complementarias. El Big Data permite mediciones más desagregadas o granulares que las estadísticas tradicionales. Además, la medición puede hacerse a un costo más bajo y con una mayor precisión. El Big Data también puede generar indicadores de actividades o comportamientos sociales hasta ahora imposibles de medir. Una ventaja adicional es que puede reducir la carga que la colecta tradicional de datos implica para las empresas e individuos. Estas ventajas están siendo ya exploradas y utilizadas por las ONE de varios países en América Latina y otras partes del mundo. Algunas ONE incluso han creado unidades o institutos de explotación de Big Data, ya sea en colaboración con instituciones privadas o mediante acuerdos de colaboración con otras entidades generadoras de Big Data específicos en proyectos concretos.

El Big Data se refiere al volumen y propiedad de los datos recogidos. Un estadio anterior es la colecta de los datos. Han surgido técnicas de colecta de datos, como AIS para el tráfico marino o MPD para las comunicaciones hechas por redes móviles, scanner data o web crawling, Procesamiento de Lenguaje Natural (PLN) entre otras, que en esencia proveen al analista de volúmenes enormes de información para su tratamiento y análisis posterior. Se trata de nuevos métodos de colecta de datos, y a la vez, de nuevos modos de medir aspectos concretos de una actividad. Cada método de colecta y los datos específicos que recoge definen a su vez los métodos que se deben aplicar para el tratamiento de esa información y las propiedades que los datos recogidos y las mediciones que se elaboren a partir de ellos van a tener.

A. Datos de Posicionamiento Móvil (MPD)

Un tipo de Big Data frecuentemente utilizado son los Datos de Posicionamiento Móvil (MPD). Éstos se refieren a registros o trazas que dejan los usuarios de comunicaciones móviles en la red que son recogidos por los operadores de telecomunicaciones. Los MPD tienen múltiples aplicaciones, especialmente si se combinan con otras fuentes de datos.

Hay tres tipos de datos que se puede capturar a través de la explotación de Big Data procedente de redes móviles y que generan interés:

Análisis de la movilidad de las personas: los usuarios hacen comunicaciones de voz o de datos o uso de aplicaciones por red móvil de un modo muy frecuente, y cada uso tiene asociada una localización (estación base o torre utilizada). Uniendo estas localizaciones se obtiene evidencia de los movimientos en una ciudad, entorno, país o de tipo internacional.

Interacciones sociales: las conexiones que por redes móviles realiza una persona con otras conforman un mapa de sus consumos y patrones sociales. Que quedan registradas en cuanto a consumos, localización y duraciones en las redes.

Actividad económica: la actividad que un individuo, o grupo, realiza con el terminal móvil es una buena proxy de otras variables socioeconómicas en ocasiones no observadas. Se ha explotado, por ejemplo, el consumo personal del teléfono móvil —en cuanto a minutos de uso o Mb utilizados en datos—y obtenido en base a éste aproximaciones al nivel de renta de personas con alta precisión. Así, cambios drásticos en los niveles de consumo pueden ser una interesante aproximación al efecto de una crisis, o expectativa de la misma, en la actividad económica más general. Los patrones y concentraciones de consumo de servicios móviles pueden ser también de gran ayuda en la detección de focos de economía sumergida.

La Unión Internacional de Telecomunicaciones (UIT) desarrolló en 5 países una experiencia pionera a nivel mundial. El proyecto "Big Data for measuring the information society" aplica una metodología de Big Data para la colecta, procesamiento y explotación de datos para medir distintos aspectos de la sociedad de la información con el fin de complementar o incluso de sustituir indicadores existentes recogidos con métodos más tradicionales. La UIT ha realizado varios proyectos conjuntos de Big Data con cinco países (Colombia, Emiratos Árabes Unidos, Filipinas, Georgia y Suecia). En cada país la UIT y la Autoridad Nacional de Regulación suscribieron un acuerdo de colaboración con los operadores de comunicaciones móviles con el fin de explotar la información de uso de red móvil de sus suscriptores. Este proyecto provee suficientes datos para poder construir indicadores agregados de consumo de telefonía e internet móvil y de itinerancia internacional —servicio este que se presta desde un país a individuos procedentes de otros países.

Como resultado de esta exitosa experiencia, en términos de la colecta y procesamiento de datos, así como en lo relativo a aspectos legales y de privacidad, la UIT (2017) publicó unos principios de buena gestión de Big Data que se presentan a continuación. Se distinguen tres fases en el procesamiento de los datos de movilidad:

Fase 1: la extracción de los microdatos, que suelen ser confidenciales, privados y/o personales[1] al tratarse de datos referidos a consumos o localizaciones en red móvil de individuos. En un primer momento, el propio operador de red móvil realiza una agregación inicial para impedir la correspondencia entre los datos referidos a un individuo y el individuo en particular. Si el operador no hace esta agregación, éste anonimiza los datos de algún modo para poder transmitirlos a otro agente —como el regulador de las telecomunicaciones— para su posterior tratamiento. Este tratamiento inicial de los datos es la responsabilidad del operador móvil, ya que es el mismo quien debe garantizar la anonimidad de esos datos.

En esta etapa, los campos que el operador maneja son los procedentes del registro de detalles de las llamadas (CDR, por sus siglas en inglés): identificador de la antena, tecnología, coordenadas de localización de la antena (latitud/longitud), si la suscripción es local o extranjera, mensajes y llamadas de entrada/salida, tráfico IP, Registro Único de Terminal Móvil (IMEI, por sus siglas en inglés). Los CDR son registros digitales que se generan automáticamente en la red del operador móvil; recogen datos del momento del inicio y final de la comunicación, las células o estaciones base utilizadas en la comunicación- y a partir de esta información el operador sabe las localizaciones de las torres involucradas[2].

La información de los CDR se puede fusionar con datos provenientes del sistema de gestión de clientes (CRM, por sus siglas en inglés), bases que incluyen información específica sobre el suscriptor del

---

[1] Dato personal (y por tanto sometido en general a la legislación sobre protección de datos y privacidad) es "toda información sobre persona física identificada o identificable", tal y como se recoge en el Reglamento General de Protección de Datos de la Unión Europea (UE).

[2] Las estaciones base de un mismo operador pueden estar a distancias distintas, dependiendo de si se trata de un entorno urbano (hasta 500 metros) o rural (varios kilómetros). La localización del individuo/usuario a partir de la torre utilizada en su comunicación siempre será una aproximación.

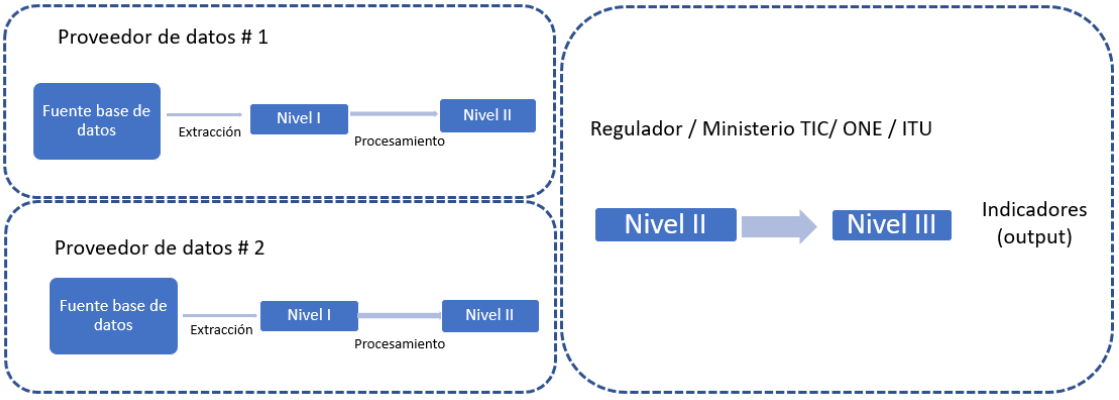
servicio, utilizada con fines de facturación por el operador: sexo, edad, idioma elegido, registro terminal móvil (International Mobile Equipment Identity, IMEI), tipo de contrato (prepago/contrato, personal/de empresa), tipo de evento (comunicación de voz, entrante/saliente, doméstica/en itinerancia, de mensaje, IP), duración, día, hora, volumen de datos consumidos, velocidad de conexión, área de localización del usuario (áreas administrativas LAU, nivel 1, 2 o 3).

Fase 2: los datos en esta fase ya están agregados o procesados de tal modo que no muestran datos privados, es decir, que ya no es posible identificar individuos concretos a partir de los campos ofrecidos[3]. En esta fase, se utilizan datos disociados, aunque puede tratarse aún de datos sensibles o confidenciales que no se deben compartir. En este momento se agrupan los datos de diferentes operadores. Esta fase del tratamiento de datos se efectúa por un agente o en un lugar distinto del proveedor u operador inicial. Esta tarea puede ser acometida por una autoridad pública o independiente.

Fase 3: es la fase del macrodato. Estos son datos agregados como indicadores que ya no tienen carácter ni privado ni confidencial.

Las tres fases en la cadena de trabajo con Big Data se presentan en el diagrama 1. Al principio, varios operadores de red móvil son proveedores de datos iniciales que se encargan de la fase 1, desde la extracción inicial hasta la anonimización suficiente de los datos. Las fases 2 y 3 se refieren a la agregación de los datos de varios operadores y al cálculo de indicadores y análisis por una tercera institución, respectivamente.

Diagrama 1
Colecta, tratamiento y procesamiento de Big Data procedente de operadores de red móvil



Fuente: Unión Internacional de Telecomunicaciones (2017).

Una vez recogidos, los datos que los usuarios dejan en las redes móviles permiten innumerables aplicaciones de gran utilidad al estudio del comportamiento social en su dimensiones espaciales y temporales. Estos datos se refieren a la localización (MPD) o el consumo (CDR). Los agentes que recaban esta información, los operadores móviles de red (MNO, por sus siglas en inglés), pueden no tener incentivos a su provisión, más allá de su propio interés por los servicios que ellos venden. Desde el momento en que los datos de, por ejemplo, la localización de un individuo son datos privados, solo se podrán utilizar respetando la legislación existente sobre la protección de la privacidad.

[3] Un dato estará anonimizado o seudoanonimizado cuando la información que logra identificar a la persona física se encuentre separada y esté sometida a medidas técnicas que garanticen que los datos personales no se puedan asignar a una persona identificada o identificable.

Hay dos modos de obtener datos de localización (MPD):

i) *Posicionamiento activo*: cuando un agente reclama de la red la localización de un terminal o smartphone específico y la red se lo da en tiempo real. Puede ser realizado bien por el propio usuario o por el gestor de la red, el operador. En este caso, es necesario el permiso del usuario final. Puede ser utilizado en flotas de camiones o de transporte donde los vehículos llevan una tarjeta SIM y la red extrae su localización en tiempo real.

   Existen aplicaciones que pueden instalarse en el smartphone que posibilitan esta localización continua. Ellas son de descarga libre por el usuario, quien puede dar permiso o no a otros para que le localicen en tiempo real. El usuario final o propietario del terminal debe dar su expreso consentimiento ante una petición de posicionamiento activo, a través de SMS u otro medio, o bien debe instalar voluntariamente y aceptar las condiciones de uso y de rastreo de una aplicación. Esta aceptación puede darse también a través de la configuración de las opciones relevantes en el sistema operativo o en la aplicación de localización. Para esta recogida de datos cabe notar que el usuario debe dar un paso adelante expreso, y esto siempre conlleva un costo.

ii) *Posicionamiento pasivo*: el operador de red recaba información registrada de los consumos de todos los usuarios, con el fin de garantizar una calidad de servicio o de facturación posterior por los servicios prestados que queda recogida en sus sistemas de un modo periódico y automático y de ahí extrae los datos de interés. Los operadores de red móvil recaban mucha información en sus sistemas, como la calidad de servicio, volúmenes de datos consumidos, identidad de los usuarios, tiempo transcurrido y localización. De este conjunto de información extraen la de localización u otros datos de interés. Muchas veces la localización per se no es recogida, pero sí la identidad de la antena desde donde se inició la comunicación. Esta información combinada con una aplicación de mapas geográficos logra ofrecer la localización del usuario final con un buen grado de aproximación. Esta información se refiere a los CDR. Además, los operadores también recogen información sobre las redes utilizadas, así como el registro de datos del Protocolo de Internet (IPDR), entre otros[4].

   Los CDR recogidos son utilizados por los operadores para la facturación del cliente final, entre otras funciones, y proporcionan por tanto información personal o privada. Antes de ser compartidos estos datos, deben ser anonimizados, tarea que se acomete eliminando la información personal. Se puede aplicar una encriptación de las numeraciones implicadas, la del emisor y la del receptor de la comunicación, por ejemplo, o bien agregar varias numeraciones en una sola. A modo de ejemplo, y siguiendo a Naciones Unidas Global Pulse (2013), los CDR a compartir se presentan en el cuadro 2.

Cuadro 2
Campos de información básicos en la colecta de datos de movilidad

Identificación emisor	Localización torre emisor	Identificación receptor	Localización torre receptor	Duración
AB448HVC3319	Lat. 40o 24´13.731´´N Long. 2o 9´14.4252´E	B4RR7PKLL998	Lat. 36o 25´1.9166´´N Long. 23o 47´11.1924´E	4:06 min

Fuente: Elaboración propia.

[4] Para más detalles sobre la metodología de colecta, procesamiento y tratamiento de datos de movilidad, véase Ahas y otros (2014).

Nótese que ciertos campos están ya anonimizados: la identificación del emisor, por ejemplo, ha sido transformada por un algoritmo por el cual al nombre o código inicialmente asignado a este individuo ha sido transformado con una regla que no permite "vuelta atrás" en el proceso. Con ello, la identidad real del individuo queda eliminada, pero se mantiene una identificación única para el sujeto con el fin de mejor explotar la dimensión longitudinal de la información en la base de datos[5].

Los datos pueden ser recogidos con sistemas o clasificaciones diferentes, lo cual debe ser tenido en cuenta en el proceso de tratamiento de los datos. Por ejemplo, la coordenadas (latitud/longitud) en el cuadro 2 están recogidas en formato geodésico especificando grados, con detalle de minutos, con símbolos (grados, minutos) y con sufijos (N/S, W/E), pero podrían recogerse también en formato 412324730 o en otros diferentes.

El tratamiento y la extracción de los datos de interés en formato manejable por parte del operador en cambio sí conlleva costos especiales. Este es el costo principal que asumieron los operadores que asumieron el acuerdo con la autoridad reguladora nacional de cada país y con la UIT, dado que por motivo de tratamiento adecuado de datos privados de usuarios es el operador quien es responsable de los datos recogidos en su red móvil y quien por tanto se responsabiliza de la etapa de anonimización de los datos fuente. A la autoridad reguladora le llegaban los datos ya procesados inicialmente por el operador.

B. El Sistema de Identificación Automática (AIS) como fuente de datos

1. Aspectos metodológicos

El Sistema de Identificación Automática (AIS, por sus siglas en inglés) provee datos, ya sea vía la muy alta frecuencia de radio (VHF, por sus siglas en inglés) o vía satélite, de la localización, identidad y movimiento de un barco en el mar. Es un sistema que envía información digitalizada de los parámetros de un barco, sus características más básicas (longitud, calado, tipo) y su movimiento. Se introdujo a mediados de los años 90 como medida de seguridad para los barcos y evitar colisiones. La información es empaquetada en un formato definido por el Sector de Radiocomunicaciones de la Unión Internacional de Telecomunicaciones (UIT-R)[6]. El conjunto de información transmitida se puede englobar en: i) información estática: localización por GPS del barco en un momento concreto; eslora y calado del barco; nombre e identificación del barco y puerto de destino; y ii) información dinámica: junto con la información de geolocalización aporta la velocidad sobre tierra y orientación.

El sistema AIS genera en cada barco la emisión —ya sea en broadcast o dirigida a otro barco o a una autoridad portuaria o de rescate- de un conjunto de 27 tipos de mensajes. Algunos mensajes pueden ser binarios con información muy definida sobre por ejemplo condiciones climáticas, útiles en cuanto que representan datos de sensores o de telemetría que pueden ser recogidos de modo automatizado por estaciones receptoras en tierra. Los mensajes basados en AIS tienen la dimensión espacial— informando de la localización del barco emisor-, con una precisión de aproximadamente 10 metros, y una temporal, cuya frecuencia puede variar en intervalos de segundos o bien de pocos minutos (Batty, 2018).

Aunque inicialmente solo los buques de cierto tamaño fueron obligados a instalar el sistema AIS, hoy es prácticamente generalizado su uso por todo tipo de barcos. Uno de sus principales beneficios reside en que evita choques entre barcos, al avisar y proporcionar la identidad, velocidad y nombre del barco que pueda entrar en ruta de colisión. Dado el gran número de barcos surcando los mares y la disponibilidad de una red densa de satélites de baja órbita, la cantidad de información que se recoge en tiempo real es alta y creciente. El uso masivo de AIS ha abierto aplicaciones inimaginables en los años 90 y en la actualidad es un sistema de comunicación y uso de datos con el fin de mejorar la seguridad, el impacto medioambiental y la eficiencia en la gestión del tráfico.

---

[5] Para una discusión sobre los procedimientos posibles para la anonimización, ver Sección 4.

[6] Recomendación ITU-R M.1371-5, de 2014 [en línea] http://www.itu.int/rec/R-REC-M.1371-5-201402-I.

La fuente de datos AIS se puede combinar y enriquecer con otras fuentes, como por ejemplo con el desarrollo del Internet de las Cosas (IoT) localizado en cadenas de logística o en puertos, y proporcionar así elementos útiles en la gestión de tráfico marino, en la medición de los flujos de comercio internacional o en la gestión de los puertos, entre otros. Las mejoras de eficiencia en la gestión de los tráficos o en la de los puertos, con actividades de carga/descarga o de reparación de barcos son muy significativas. Se utilizan datos de AIS procedentes de barcos y trayectorias realizadas en últimos años para la identificación de rutas seguras y eficientes entre puertos, minimizando el tiempo y consumo de gasóleo. Se utilizan también para evitar cuellos de botella en las entradas y ordenes de carga/descarga en puertos comerciales con mucha demanda, como en Livorno (Italia), Singapur o Shanghai[7].

A nivel medioambiental, las aplicaciones de datos de AIS son de interés, como por ejemplo en la pesca. Barcos pesqueros de pequeño tamaño suelen recorrer áreas cercanas a la costa y su seguridad y éxito en las capturas puede ser mejorado con información de otros barcos y rutas utilizadas. En gran parte del mundo existe legislación que limita el total de capturas para ciertas especies a lo largo del año. Los datos AIS pueden ayudar a las autoridades pesqueras de cada país a monitorizar el total de capturas por barco, las áreas marítimas visitadas y posibles descargas efectuadas en puertos con el fin de preservar los bancos pesqueros y evitar su sobreexplotación.

El número de barcos que llevan instalados sistemas AIS, el aumento en el número de satélites de baja órbita dedicados al AIS, la frecuencia y campos de información que se transmiten y las aplicaciones tan numerosas han hecho que el volumen de información a recoger y tratar haya explotado recientemente. Se ha hecho necesario redefinir el ancho de banda VHF utilizado en la trasmisión y enriquecerlo con tecnología de VHF Sistema de intercambio de datos de muy alta frecuencia (VDES, por sus siglas en inglés)[8] que proporciona mayor ancho de banda y mejor nivel de autentificación del emisor.

Varios investigadores han aproximado el volumen del comercio marítimo de mercancías gracias a la explotación de los datos procedentes de AIS, como por ejemplo el volumen del comercio de petróleo (Adland y otros, 2017) o de gas licuado (Shibasaki y otros, 2020) o para el análisis de tráfico en tiempo real (Arslanalp y otros, 2019), la estimación de carga en buques de gran tamaño (Jia y otros, 2019) o la planificación de tráfico en puertos (Arifin y otros, 2018).

A continuación se presenta un ejemplo del uso de datos de AIS para la aproximación de volúmenes de mercancías descargadas en distintos puertos de Japón. Este ejemplo tiene interés por la combinación de grandes bases de datos procedentes de fuentes diferentes, la estructuración de las tareas necesarias para el tratamiento de los datos (desde su extracción hasta su explotación), la necesidad de contrastar la bondad de las aproximaciones obtenidas con datos de otras fuentes (en ocasiones fuentes oficiales) y para detectar los retos encontrados[9].

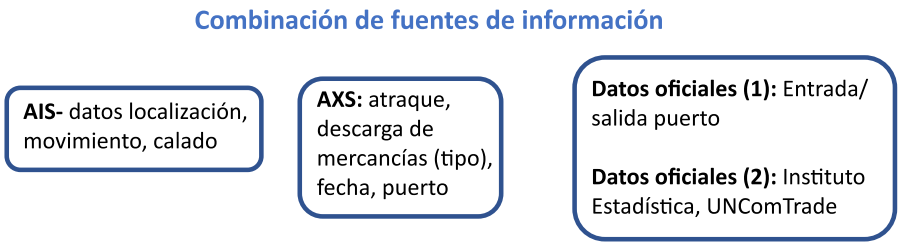
---

[7] Para aplicaciones de interés en el campo de la logística, ver European Union H2020 EfficienSea2 project [en línea] https://cordis.europa.eu/project/id/636329/es.

[8] Recomendación ITU-R M.2092-0 de 2016.

[9] Para una visión más general de las aplicaciones que se hacen en la actualidad explotando Big Data de AIS en el campo del transporte, ver Tu y otros (2017).

Diagrama 2
Fuentes de información complementarias utilizadas en la explotación de datos AIS para la estimación de flujos de comercio marítimo internacional



Fuente: Elaboración propia del autor.

2. Aplicaciones

Kanamoto y otros (2021) aproximaron el flujo global del comercio marítimo de hierro, fertilizantes, grano y carbón en los distintos puertos de Japón para 2016, explotando por un lado la información AIS de los buques, y por otro lado, la información sobre atraques en puerto proporcionada por distintas autoridades portuarias. La información utilizada procedía de tres fuentes, las dos primeras utilizadas para la aproximación de volúmenes de las mercancías mencionadas de puerto- a- puerto y la tercera como contraste para la bondad de la aproximación:

i) AIS: con datos de localización, permanencia y movimiento de barcos o buques. Se identifica el puerto de atraque de cada buque. Se extrae la información de: puerto de atraque, localización del atraque o muelle específico utilizado por el barco, hora/día de llegada y salida así como información del calado del buque cuando llega a puerto y cuando departe.

ii) Datos de atraque de barcos en puertos (fuente privada utilizada AXS Dry). Con estos datos se construye una lista de productos descargados en cada puerto, así como el registro de barcos entrantes y salientes en cada momento del tiempo.

iii) Datos agregados a nivel de país de otras fuentes oficiales (UN Comtrade y otras) para el contraste y validación posterior de las estimaciones realizadas.

3. Combinación de bases de datos

Con datos de geolocalización procedentes de AIS, los autores pudieron identificar un buque y el puerto y el muelle de estancia, su calado y tiempo de permanencia. El calado es muy relevante porque permite descubrir si el barco ha cargado mercancía o la ha descargado en muelle. Por otro lado, con la base de datos externa AXS[10] los autores obtuvieron un registro de los buques que habían entrado a muelle cada día, así como el agregado (pero no por buque individual) de mercancías por tipos descargadas en muelle.

Dado que el objetivo es obtener los volúmenes de mercancías individuales transportados entre puertos, ninguna de las dos bases de datos es suficiente y la combinación de ambas por sí sola tampoco lo es, ya que ninguna proporciona la mercancía individual descargada o cargada en puerto, aunque sí si hubo descarga o no. Dado que los buques, especialmente los grandes, transportan diferentes mercancías simultáneamente que pueden cargar/descargar en un mismo o en diferentes puertos, se hace necesario acometer aproximaciones que combinen ambas fuentes de datos.

[10] https://public.axsmarine.com/axsdry.

Una de las ventajas del estudio con estos datos es el nivel de agregación posible. Allí donde las estadísticas oficiales, como las procedentes de las ONE o la base de datos Comtrade de las Naciones Unidas, solo proporcionan información de flujos por mercancía agregados a nivel de país y año, la información que los autores obtuvieron gracias a la explotación de datos procedentes de AIS les permitió un análisis de flujos comerciales con mayor frecuencia y desagregados por rutas y puertos.

4. Manejo de datos

Para la identificación de la localización y puertos visitados, con datos AIS se estiman con alta frecuencia trayectos que unen varias localizaciones utilizando los datos de geolocalización del barco en momentos diferentes (trayectoria del barco). De este modo cuando la trayectoria era nula, identificaban si el barco estaba atracado. La estimación de trayectorias se realizaba cada un cierto número de horas, con lo que la información de llegada o salida del atraque era bastante precisa y podía complementarse con la información alternativa, procedente de AXS, sobre el registro de llegadas y de salidas de buques en cada puerto.

Para la extracción de datos de geolocalización de los barcos utilizaron Geohash, técnica de geo-codificación que permite utilizar cadenas alfanuméricas de distintas longitudes y proporciona distintos grados de precisión en la geolocalización[11]. Por ejemplo, para identificar el puerto de llegada se utilizaban una codificación de 4 dígitos, pero para identificar el muelle o atraque se utilizaba una de 7 dígitos (que permite localizar un área con precisión de 100 metros). En este caso se utilizaron Geohash en dos etapas (con dos niveles distintos de agregación).

Los autores escogieron las variables disponibles de AIS (localización, trayectorias, destinos, longitud barco, calado de alta frecuencia, atraque en muelle) y la información disponible del registro de atraques en puerto y descarga mercancías de AXS Dry Data (matriz de puerto de salida/llegada, cambios en calado) y con estas estimaron los volúmenes de descarga de cada barco y para cada mercancía cruzando los datos de AIS para el momento/día del evento y los datos de AXS para el tipo de mercancías registradas en puerto.

5. Validación de resultados con otras fuentes

Una vez los flujos de tráfico de las mercancías puerto a puerto han sido aproximados, se realiza una validación con datos más agregados. Se agregan en primer lugar los resultados de flujos de transporte a nivel de país y con periodicidad anual. Después se contrastan los resultados con los publicados en fuentes oficiales como UN Comtrade y otras privadas como AXS, Clarkson (2019) e IHS. Para el contraste sobre la bondad de las aproximaciones hechas con sus fuentes y método, los autores corrieron una regresión sencilla: estimación propia=f(fuentes alternativas)+ε. Los resultados obtenidos muestran una explicación mayor al 99% de los flujos observados (reales) país a país, para un conjunto de 16 países que intercambiaron con Japón, tanto para el volumen de mercancías importado como para el exportado, excepto para el transporte de hierro y acero (que los autores achacan a las diferentes clasificaciones de estos productos utilizadas en distintos puertos). Los autores contrastaron también la bondad de la aproximación con datos de frecuencia mensual de importaciones en Japón de mineral de hierro y carbón, consiguiendo aproximaciones superiores al 94% de fiabilidad en casi todos los puertos analizados.

[11] Geohash es un modo de expresar la localización utilizando una cadena corta alfanumérica para las coordenadas posicionales (latitud/longitud).

Uno de los desafíos encontrados se refiere a las divergencias entre las clasificaciones de mercancías utilizadas ya sea por la autoridad portuaria relevante, por las autoridades nacionales o por las fuentes adicionales de datos utilizadas (AXS, UN Comtrade, Clarkson). Si bien desde Naciones Unidas[12] se ha avanzado mucho en la homogeneización de clasificaciones de bienes comerciados a nivel mundial, las autoridades de cada puerto pueden seguir clasificaciones distintas. Esta divergencia ocasiona errores en las aproximaciones al dificultar el matching de productos cuando se utilizan fuentes distintas de datos.

Una experiencia de interés con respecto a la dificultad de corresponder distintas clasificaciones de bienes y servicios transaccionados a nivel internacional se dio en Brasil. Los países miembros del Mercosur (Argentina, Brasil, Paraguay y Uruguay) tienen una clasificación acordada para los bienes intercambiados en esta área comercial, la Nomenclatura Común del Mercosur (NCM). Esta clasificación utiliza 96 capítulos -categorías generales de bienes- y produce 10.000 registros únicos a asignar a cada mercancía comerciada en la región. El desafío reside en que Brasil utiliza en su sistema de aduanas una clasificación con fines arancelarios y fiscales distinta. Así, resulta necesario asignar a cada mercancía no solo el código fiscal brasileño correcto sino también el código de la NCM para poder atribuir el arancel común de un modo correcto a los flujos comerciales. Con este fin, Rodrigues de Lima et. al. (2022) utilizan técnicas de procesamiento de lenguaje natural —una rama de la Inteligencia Artificial— con el fin de encontrar una correspondencia rápida y con precisión adecuada a los artículos catalogados en base a la NCM con el sistema de clasificación fiscal brasileño.

C. 5G e Internet de las Cosas (IoT)

La quinta generación de comunicaciones móviles 5G (3GPP) o estándar IMT-2020 definido por la Unión Internacional de Telecomunicaciones, UIT, puede proporcionar velocidades hasta 100 veces mayores que el estándar utilizado más frecuentemente en la actualidad (4G/LTE), y con niveles de latencia mucho menores, lo cual hace que las comunicaciones sean muy fiables. El mayor ancho de banda que permite el estándar 5G hace posible la transmisión de grandes flujos de datos entre millones de objetos, desarrollo denominado Internet de las Cosas (IoT), con millones de objetos y elementos conectados de continuo e interaccionando entre sí.

El estándar 5G/IMT-2020 puede utilizar las bandas de frecuencia que actualmente utilizan los operadores móviles o bien nuevas bandas, como las de alta frecuencia (superiores a 6 GHz)[13].

El despliegue del estándar 5G en algunas zonas del mundo es muy rápido, como por ejemplo en Europa, Estados Unidos y Canadá, donde las coberturas de población con acceso a estas redes superan el 70%. A medida que los consumidores adquieren terminales capaces de utilizar 5G la penetración o uso efectivo de las redes 5G aumenta rápidamente. Ericsson (2021) estima que para 2027 el total de suscriptores de comunicaciones móviles en el mundo en 5G será superior al de suscriptores con cualquier otra tecnología móvil.

Los principales usos que se dan con conexiones 5G/IMT-2020 se pueden agrupar en tres grandes clases (IDATE,2021):

i) Banda ancha móvil mejorada (eMBB, por sus siglas en inglés)

ii) Comunicaciones masivas de máquina a máquina (mMTC, por sus siglas en inglés)

iii) Comunicaciones ultra bajas, confiables y de baja latencia (ULLC, por sus siglas en inglés)

---

[12] La clasificación de bienes utilizada por UN Comtrade viene recogida en Naciones Unidas (2010).

[13] Las bandas de frecuencias inferiores a 6 GHz son hoy las más utilizadas en el despliegue de 5G por sus propiedades en cuanto a cobertura y costos de despliegue. Las bandas más frecuentemente puestas a disposición de los operadores en los últimos años son las de 3,3 Ghz a 3,8 Ghz.

Una característica de las redes 5G/IMT-2020 es la densificación de las antenas. Dado que el escenario previsto es el de millones de aparatos o dispositivos conectados transmitiendo datos a alta velocidad, es necesario cubrir el territorio de pequeñas antenas que sitúen el punto del acceso a la red muy cerca del usuario, máquina u objeto conectado.

Nótese que de los principales conjuntos de usos mencionados, cada uno demanda propiedades de conectividad diferentes. Si la eMBB necesita de amplios anchos de banda, las conexiones masivas entre máquinas (mMTC) necesitan gestionar no tanto volumen de datos, pero sí millones de puntos de acceso. Las comunicaciones ultra-fiables (ULLC), como pueden ser ciertas aplicaciones de medicina a distancia o la gestión de vehículos o robots autónomos, necesitan tanto altos anchos de banda como latencias de órdenes de magnitud menores que las existentes[14].

Un avance que facilita proporcionar estos tres tipos de usos por una misma red 5G es el denominado network slicing. Este permite utilizar una misma red física desplegada para diferentes usos de un modo dedicado, garantizando diferentes calidades de servicio para cada uso previsto (GSMA, 2017). Para el transporte o logística inteligente, que requiere soportar muchos elementos conectados a la vez, esta posibilidad adquiere importancia ya que permite soportar estos servicios junto con otros a la vez, segmentando la red en diversos tramos, con el ahorro en costos incrementales del servicio que esto supone.

Relacionado con el desarrollo de las redes 5G está el IoT. Según lo define la Organización para la Cooperación y el Desarrollo Económico (OCDE) (2022), este "*incluye todo tipo de terminales, objetos cuyo estado puede ser alterado vía internet con o sin la ayuda de intervención humana*". Esto es, el IoT no se refiere solo a la parte de las redes de telecomunicaciones, ni solo a los objetos conectados, sino a toda la cadena de valor o el ecosistema que forman los agentes y servicios desde la red al usuario, incluyendo todo tipo de servicios que sobre las redes y los terminales se prestan haciendo que las "cosas" se comuniquen entre sí. Se trata por tanto de varias capas de valor al mismo tiempo: i) la infraestructura (no solo de redes, sino también de computación en la nube y los propios datos que se pueden colectar), ii) los terminales embebidos o integrados en "cosas", con el software correspondiente, así como la capacidad de comunicar, enviar/recibir datos, y iii) la capa de plataforma y de aplicaciones.

Un modo de entender IoT es como la agregación de puntos finales identificables y únicos que se comunican de modo bidireccional usando una o varias redes de un modo permanente y universal. La generación de valor más predominante en IoT es la colecta, almacenamiento, tratamiento y compartición de todos los datos o información que las distintas partes —elementos conectados— proporcionan y su explotación posterior, con el fin de mejorar un proceso concreto y poderlo monitorear y medir.

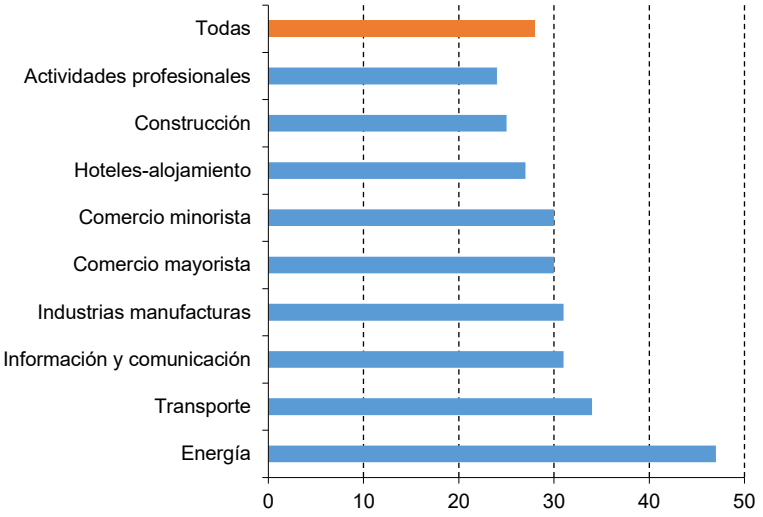
Las aplicaciones del IoT que ya se observan son de mucho interés[15]: desde la medicina a distancia a las cadenas de logística, desde las ciudades "inteligentes"- con sensores para una mejor gestión de la electricidad, energía o el tráfico-, hasta la gestión domótica de aparatos y gestiones de tareas en la casa o en la industria. Según datos de la OCDE, el 28% de las empresas en muchos de sus países miembros ya utilizan alguna forma de IoT en sus procesos productivos, con mayor intensidad en industrias como energía —con el auge del *smart metering*— y transporte y logística (véase el gráfico 1).

---

[14] La latencia o *roundtrip- time* es el tiempo que transcurre entre la emisión de un mensaje por el emisor, su llegada al receptor y el retorno al emisor. Una red 4G/LTE puede tener 50- 60 ms de latencia. Las redes 5G pueden reducir ésta a la mitad o mucho menos, dependiendo de la banda utilizada.

[15] Véase DotEcon (2018) y Axon Partners (2022).

Gráfico 1
Proporción de empresas en países de la OCDE que utilizan IoT según industria, 2021
(En porcentajes)



Fuente: Elaboración del autor sobre la base de OCDE, https://stats.oecd.org/Index.aspx?DataSetCode=ICT_BUS.

1. El IoT y la industria del transporte internacional y logística

Se ha visto en la sección anterior cómo una condición necesaria para el desarrollo del comercio y la aplicación de tecnologías 4.0 es la existencia de redes de telecomunicaciones con amplio despliegue y cobertura. Si bien las redes son esenciales, es su integración con la información generada por ellas y la aplicación de inteligencia sobre estos datos lo que aporta nuevos modelos de negocio, nuevos servicios y en suma, valor agregado. La integración de distintas tecnologías 4.0, como la IA, el Internet de las Cosas, la Realidad Aumentada, servicios en la nube o el big data, proporciona mejoras en la eficiencia de, por ejemplo el transporte y la logística, el comercio electrónico y en su cadena total de valor *(supply chain management).*

La industria del transporte y la logística suele incurrir en pérdidas, retrasos o costos derivadas de la falta de visibilidad del cargamento en algunas partes de la cadena. Con tecnologías como la geolocalización continua por GPS, IA, IoT o Realidad Aumentada (RA) se puede monitorizar el transporte a nivel muy preciso- incluso a nivel de container-, geolocalizado y con información suficiente para una mejor gestión del mismo. En una cadena logística cualquier reducción de pérdidas y mejora en la duración de un transporte o fase del proceso trae reducciones de costos directas. Y con redes 5G, a diferencia de las redes móviles o celulares hoy existentes, se consigue que un gran número de objectos y altos volúmenes de datos puedan ser gestionados y en tiempo real.

Las redes 5G en el transporte inteligente se utilizan para una variedad enorme de tareas y procesos: la monitorización y gestión óptima de rutas y carreteras a seguir por los camiones, la localización de módulos de transporte en cualquier rincón, la gestión integrada de diferentes transportes modales, el monitoreo de la temperatura y grado de humedad de los cargamentos de alimentos o productos perecederos, la optimización de rutas y carreteras alternativas a elegir, o en el uso de múltiples modos de transporte, incluso el uso de la realidad aumentada con el fin de visionar las cargas/descargas en puerto o centro logístico de mercancía o la supervisión por video de procesos donde la seguridad del tráfico sea relevante.

Para la gestión de flotas de transporte el *predictive maintenance* (mantenimiento predictivo) puede ser de utilidad en cuanto que gracias a la red, a los sensores insertados en los vehículos y a la IA

se puede identificar e incluso predecir en ciertos casos, averías o mal funcionamiento de partes de un vehículo y programar mejor su reparación.

Un aspecto importante y que se sitúa en otra capa de valor diferente de la conectividad es la posibilidad de gestionar varias redes simultáneamente y, en especial, de poder compartir datos generados por objetos en una red con otra subsiguiente y poder explotar esos datos con un fin preciso. Esto es probablemente lo que mejor define el IoT: la explotación de datos procedentes de muchas fuentes y elementos conectados que pueden interactuar entre sí y ser explotados en tiempo real.

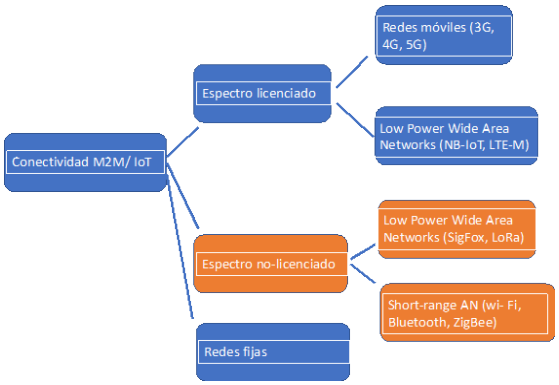
Es esta integración de redes, datos y capacidad de análisis lo que conduce también a cambios profundos en la integración de servicios hasta ahora segmentados, provistos por agentes diferentes e incluso en los modelos de negocio. Un ejemplo claro es el comercio electrónico. Es normal observar como una plataforma de comercio electrónico se centra en la presentación, búsqueda, comparación de productos y en facilitar su contratación a través de una plataforma. Con el desarrollo de redes de telecomunicaciones es posible la integración de estas fases en la contratación con otras, como la fase del pago del producto o con la fase de distribución del producto al usuario final —allí donde el reparto deba ser presencial o físico—. Actualmente grandes plataformas de comercio electrónico, como Alibaba o Amazon, suelen integrar en sus servicios todos estos servicios y usan redes de telecomunicaciones y tecnologías 4.0 avanzadas en su gestión.

La demanda que estas mejoras implican sobre la conectividad universal de las redes por desplegar es enorme. No solo hay demanda de nuevas redes licenciadas de espectro de frecuencias, como las redes 5G/IMT-2020[16], también se está haciendo uso de redes inalámbricas —no licenciadas, ya sean estas de amplio alcance (LoRa, SigFox) o de muy corto alcance— útiles por ejemplo para el *smart metering* —como en consumos de agua o de electricidad en los hogares— como las redes wi- Fi, Zigbee o Bluetooth (véase el diagrama 3). Estas últimas son muy relevantes ya que la parte de la red más "capilar" o de "última milla" es la más costosa de desplegar. Los usos industriales de las redes 5G se espera sean de tal importancia que ha conducido a que algunos gobiernos hayan asignado partes del espectro de frecuencias destinado a 5G para agentes y usos industriales exclusivamente.

Para las empresas que realizan operaciones de comercio internacional, la "logística conectada" es un gran avance. Monitorear y medir todos los elementos de una cadena de suministro implica grandes inversiones, pero también grandes mejoras en eficiencia y ahorro de costos. Esto es posible si se dan tres condiciones: i) conectividad global o universal hasta niveles muy locales (de "punto de entrega o distribución final"), ii) potencia de cálculo y sensores integrados entre sí, y iii) uso de IA (inteligencia artificial) para conocer lo que ocurre entre los sensores, terminales y máquinas conectados entre sí.

[16] Para una visión del espectro puesto a disposición de operadores móviles en Latino América, véase GSMA (2020).

Diagrama 3
Redes utilizadas para Internet de las Cosas



Fuente: BEREC (2019), "Internet of Things Indicators", BoR (19) 25, BEREC, marzo.

2. La logística conectada

Para conseguir una "logística conectada" es necesario garantizar conectividad en el vehículo o medio de transporte, así como garantizar también esta conectividad dentro del vehículo —para gestionar internamente mercancías, plazos y almacenamientos, calidad del producto—, así como garantizarla en las vías por donde circulen los vehículos (carreteras, vías de tren, espacio aéreo o marítimo). Con 5G es posible conectar todas estas fases de la cadena logística y elementos con el fin de mejorar las rutas, los procesos, las entregas, los imprevistos, desde que el bien es producido o almacenado, hasta la entrega final al consumidor. Dado que el transporte internacional frecuentemente es multimodal, las posibilidades de optimizar las rutas, medios y tránsitos mediante una red que conecte todos los puntos de una cadena se multiplican.

El desarrollo de los vehículos autónomos o conectados es un ejemplo del avance tecnológico que impacta directamente en las cadenas de suministro. Ya sea un automóvil, un camión, un dron o una embarcación, con la implementación de sensores, antenas, una red o varias redes de soporte e inteligencia adicional es posible el transporte de mercancías de un modo autónomo, esto es, sin necesidad de intervención humana. Esto impacta en los costos de gestión, en la demanda de trabajo—que pasa del menos al más cualificado—, a la optimización de rutas, puertos, puntos de carga/descarga, coordinación multi-modal, redes y en definitiva, a un mejor servicio al cliente final[17].

D. El comercio electrónico transfronterizo

El comercio electrónico transfronterizo es una fuente de evidencia clara para conocer mejor los flujos digitales de comercio internacional, aunque tiene limitaciones claras en cuanto a su comparabilidad con, y encaje en, las estadísticas oficiales. Es importante delimitar en primer lugar lo que se entiende por comercio electrónico para conocer con exactitud lo que se mide y ver así las ventajas e inconvenientes de esta aproximación al comercio internacional.

El comercio electrónico se define como "*la compra o venta de un bien o servicio, realizado a través de redes informáticas mediante métodos diseñados específicamente para el propósito de recibir o realizar pedidos. Una transacción de comercio electrónico puede hacerse entre empresas (Business-to-Business, B2B), entre empresas y consumidores (Business-to-Consumers, B2C) entre empresas y gobierno o instituciones públicas (Business-to- Government, B2G)*" (OMC-OCDE-FMI-ONU, 2023).

[17] Para una panorámica de posibles casos de uso e impactos de 5G y de IoT en la gestión de ciudades inteligentes y transporte urbano, en el sector de logística y en el coche autónomo/conectado, ver Banco Mundial (2020).

Eurostat, en el ámbito de la UE, tiene una definición prácticamente idéntica del comercio electrónico, por el cual entiende "*la venta o compra de bienes o servicios, ya sea entre empresas, hogares, individuos o entidades privadas, a través de transacciones electrónicas que se conducen o realizan por internet o por otra red de computadores*" (Eurostat, 2022).

Normalmente se entiende por comercio electrónico toda orden de compra o de venta que se haya realizado por internet. La definición más comúnmente utilizada abarca también contrataciones u órdenes realizadas por el intercambio electrónico de datos (EDI, por sus siglas en inglés), software utilizado por las empresas con formatos estandarizados de mensajes entre redes de computadores. Esto es, tanto el comercio B2C, como el ocurrido entre empresas (B2B), con la administración pública (B2G) o entre individuos (C2C), se engloban en lo que se mide por comercio electrónico.

Hay que aclarar la diferencia entre las ventas (valores brutos) transados en el comercio electrónico y el valor agregado. Normalmente en las cuentas nacionales, incluido la Balanza de Pagos, se busca calcular el valor agregado generado por las empresas residentes en un país. Por ejemplo, para el cálculo del Producto Interior Bruto (PIB), solo se toman en consideración las transacciones finales, a nivel minorista (retail). Todas las transacciones a nivel intermedio (wholesale), las que tienen lugar entre empresas por lo general, tienen su valor incorporado en el valor (de mercado) del producto final, cuando se vende al mercado final, y por tanto no se suman.

El comercio electrónico se calcula para las transacciones entre particulares (C2C), entre empresas y particulares (B2C), entre empresas (B2B), y entre empresas y gobiernos (B2G). Nótese que tanto el B2B como, en ocasiones al menos, el B2G, no forman parte del PIB. Esto, ya que se trata de valor agregado generado a nivel intermedio, y con el fin de evitar la doble contabilidad, se ignoran en el cómputo final del PIB. Luego, la suma de todo el comercio electrónico medido potencialmente (B2C, C2C, B2B, B2G) sobreestima el valor del producto interior (valor agregado) generado por las empresas en un país, y se trata por tanto, de una cifra no comparable directamente con el PIB.

Otra razón por la cual comercio electrónico y el PIB no son comparables es el mercado de producto usados o de segunda mano. Si un individuo vende una bicicleta usada a otro a través de una plataforma de venta por internet, esto es una transacción medible potencialmente de comercio electrónico —es comercio C2C— pero no forma parte del PIB por no tratarse de una nueva producción generada. Y este tipo de intercambios cada vez se dan con mayor frecuencia y volumen.

Hay otra consideración a hacer que conduce también a resaltar la diferencia entre comercio electrónico y valor agregado: la doble contabilidad de una misma transacción. Aunque depende del método de medición utilizado, en principio comercio electrónico es todo lo que se haya contratado/ordenado en línea, tanto ventas como compras. Así, si una ONE recoge en una muestra amplia de empresas su volumen de comercio electrónico, puede pedir, por un lado, el valor de compras en línea efectuado, y, por otro lado, el valor de ventas efectuado. En otra muestra a los consumidores, se puede pedir también estas dos piezas de información: compras y ventas en línea. Sin embargo, no se puede agregar, por ejemplo, las ventas domésticas de las empresas y sumar las compras efectuadas por los individuos, ya que se estaría contabilizando una misma transacción dos veces. Si bien esta doble contabilización puede ser tratada y reducida, dependiendo del tipo de fuente de información disponible y detalle, lo normal es que ocurran dobles conteos al obtener la cifra agregada al país.

1. Comercio electrónico y comercio internacional

El comercio electrónico puede darse a nivel doméstico o bien entre países, también llamado comercio electrónico transfronterizo. Se trata del conjunto de órdenes de compra o de venta que den lugar a una importación o a una exportación de bienes o servicios[18]. Su medición tiene retos importantes[19]. En primer lugar, el alcance de lo que se entiende por comercio electrónico puede cambiar según la organización

[18] Definición en línea con los principios establecidos por las Naciones Unidas y otros (2012).
[19] Véase Capítulo 13 en Comisión Económica para Europa, Eurostat y OCDE (2011).

que se encargue de su medición. La red de soporte de las órdenes de compra puede ser solo internet (red abierta) o puede incluir también redes privadas, utilizadas por grandes empresas. Con base en la definición de la OCDE, tiende a adoptarse un concepto amplio de redes "de computadores".

La cifra de comercio electrónico recogida, por ejemplo, por plataformas de comercio en línea tenderá a sobre ponderar el volumen real de comercio internacional efectuado entre países tal y como se viene recogiendo en las estadísticas oficiales.

2. Distinción entre bienes y servicios en el comercio internacional y el comercio electrónico

En la contabilidad del intercambio de productos entre países ocurre una distinción clara entre bienes, por un lado, y servicios, por otro. El criterio básico de distinción entre ambos se basa en el modo de reparto o distribución al usuario final: si éste ocurre digitalmente o no (físicamente). Si ocurre en físico, la tradicional distinción entre bienes y servicios es de aplicación: si el producto conlleva un cambio de propiedad en favor del consumidor de tipo permanente, se considera un bien; si la propiedad se transfiere de un modo temporal, i.e., por licencia temporal o de uso, se trata en cambio de un servicio. Si el reparto o distribución final del producto ocurre por un medio digital, como internet, entonces el intercambio cae en categoría de servicios, lo mismo que los cargos, si los hay, del pago o de la intermediación financiera que la transacción puede acarrear (véase el cuadro 3)[20].

Cuadro 3
Tratamiento del comercio electrónico transfronterizo en bienes y en servicios según tipo de transacción

Tipo de transacción	Clasificación
1. Venta o compra de bienes o servicios ordenados electrónicamente y entregados físicamente	Bienes
2. Venta o compra de bienes ordenados y entregados electrónicamente	Servicios
3. Venta o compra de servicios ordenados y entregados electrónicamente	Servicios
4. Venta o compra de servicios relacionados con informática realizada a través de redes de computadores, entregada físicamente y con derecho de uso permanente	Bienes
5. Venta o compra de servicios relacionados con informática realizada a través de redes de computadores, entregada físicamente y con derecho de uso temporal basado en pago de licencia	Servicios
6. Venta o compra de servicios relacionados con informática realizada a través de redes de computadores, entregada digitalmente o descargado por usuario final	Servicios

Fuente: Comisión Económica para Europa, Eurostat y OCDE (2011).
Nota: Las transacciones 1-3 no implican el uso de propiedad intelectual, mientras que las transacciones 4-6 implican el uso de derechos de propiedad intelectual, ya sea de software, servicios electrónicos, audiovisual o de contenidos.

3. Clasificaciones del comercio electrónico y las estadísticas oficiales de comercio

Un reto que ha recibido cierta atención reciente es el tratamiento que recibe el comercio electrónico en los sistemas de clasificación de actividades, productos o comercio internacional.

Normalmente la colecta de datos de comercio electrónico la realiza un intermediario, bien sea financiero (banco o *merchant* de crédito/débito) o una plataforma de comercio electrónico (como Alibaba, Amazon o e-Bay). En su colecta es normal que utilicen clasificaciones o denominaciones de los bienes y servicios transaccionados distinta de las clasificaciones oficiales al uso. Es necesario entonces, una vez se accede al listado de transacciones o pagos hechos a través de alguna plataforma, la reclasificación de los bienes y servicios en base a una clasificación oficial.

[20] Estos criterios de delimitación son acordes también con FMI (2009).

Una salida puede ser utilizar las grandes clasificaciones existentes de ramas de actividad, ya sea la CIIU Rev. 4 (Naciones Unidas, 2011) o por ejemplo la NACE (Eurotat, 2008) en el espacio de la UE, para esta reclasificación, las cuales ofrecen distintos niveles de agregación (2, 3 o 4 dígitos). Un reto con el uso de estas clasificaciones tiene que ver con la categoría 479x de NACE- categoría de 3 o 4 dígitos a la que muchas empresas asignan no ya su actividad principal, sino también sus ventas. Esta categoría 479x engloba la actividad de "ventas al por menor no físicas (no realizadas en tiendas)"; por ejemplo 4791 recoge "ventas al por menor realizadas vía e- mail o vía internet".

La asignación de empresas en base a su actividad principal a un código CIIU/NACE de cuatro dígitos es muy útil como primera aproximación a la actividad general a la que la empresa dedica su producción; sin embargo, tiene sus limitaciones para conocer con más precisión los bienes que han sido ordenados electrónicamente y su volumen. La categoría 479x no da detalle alguno del tipo de bien o servicio transaccionado y además puede ser que una empresa con distribución básica en línea utilice también otros canales (como el físico) para la distribución de sus productos simultáneamente (empresas multicanal), situación cada vez más frecuente en el sector de venta minorista.

Otra alternativa es utilizar directamente la clasificación estándar de productos para comercio internacional llamado el Sistema Armonizado (SA), que en su última versión 2022 presenta más de 200.000 productos[21]. Esta clasificación es compatible a distintos niveles de agregación con la CIIU y NACE. También existe la Clasificación Central de Productos (CPC) o la Nomenclatura Combinada (NC) de la Unión Europea[22], entre otros sistemas de clasificación. Adicionalmente, las Naciones Unidas y otros (2012) disponen de una clasificación específica para el comercio internacional de servicios.

4. Comercio electrónico, sus fuentes y retos de medición

Hoy las ONE no disponen de un registro de transacciones de comercio electrónico realizadas en su país. Los datos pertenecen a empresas en general privadas, ya sea del sector financiero, plataformas de comercio electrónico, cadenas grandes de retailers o incluso tiendas de menor tamaño que trabajan en multicanal (offline y en línea) y que disponen de su propia plataforma. Por ende, si una agencia pública quiere conocer con algún detalle las características del comercio electrónico que tiene lugar en el país, solo puede hacerlo de dos modos: i) conducir sus propias encuestas a hogares/individuos o empresas —lo cual tiene limitaciones—, o bien ii) llegar a un acuerdo con las empresas que dispongan de los microdatos para su explotación posterior.

Sin perjuicio de las limitaciones ya señaladas, el comercio electrónico supone una parte muy importante del comercio total (en muchos países por encima del 20% de la recaudación total minorista) y creciendo, no solo en volumen, sino en ámbitos donde antes no se contrataba en línea y ahora se hace de modo masivo (véase el gráfico 2).

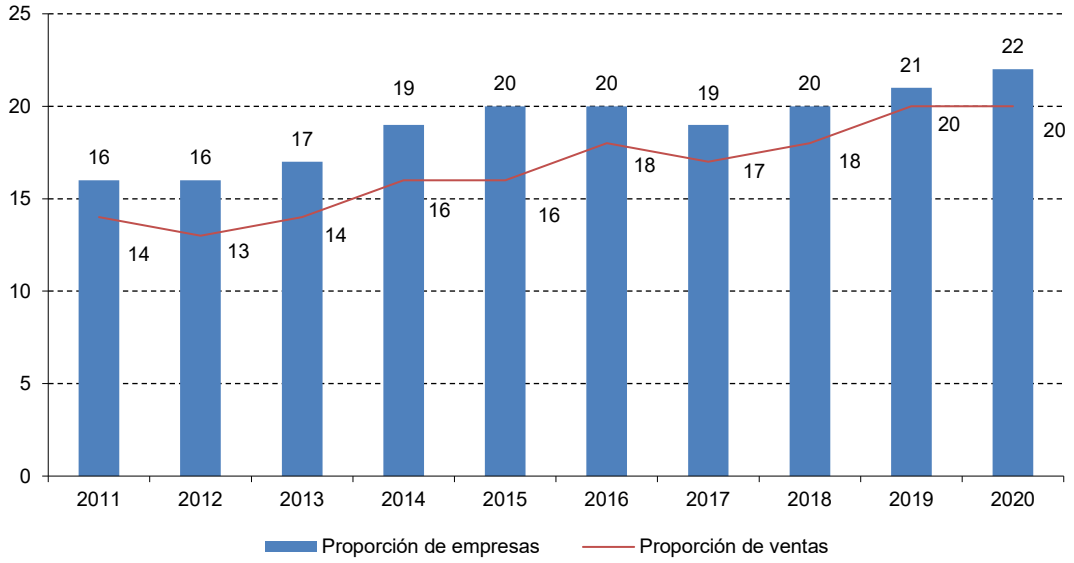
En el comercio de cualquier tipo hay tres fases diferentes: i) la orden de compra/venta, ii) el pago, y iii) el reparto o distribución final. El comercio electrónico, como se definió con anterioridad, se centra exclusivamente en la fase i) de contratación en línea. Los agentes implicados en cada parte de esta cadena de valor suelen ser diferentes, aunque en ocasiones tengan integradas verticalmente varias actividades en una sola organización. Cada agente en esta cadena de valor tiene acceso a datos diferentes, que pueden resultar más o menos útiles para la medición del comercio electrónico. Un agente de reparto del bien final, por ejemplo una empresa postal o courier, no tiene por qué conocer el valor o tipo de bien que reparte, ni el precio al cual el consumidor lo adquirió. Su fuente de datos permitiría mediciones solo aproximadas del volumen agregado de comercio. Un gestor de tarjetas de crédito, por el contrario, dispone de mayor contenido informativo en cuanto al comercio electrónico. Puede saber si el consumidor estaba presente con su tarjeta bancaria o no en el momento del pago; sabe el importe de la transacción, momento, origen y destino, y alguna denominación del bien o servicio pagado.

---

[21] Para más información, véase [en línea] https://www.wcoomd.org/en/topics/nomenclature/instrument-and-tools/hs-nomenclature-2022-edition/hs-nomenclature-2022-edition.aspx.

[22] Para más detalles, véase [en línea] https://ec.europa.eu/eurostat/web/international-trade-in-goods/methodology/classifications.

Gráfico 2
Proporción de empresas que utilizan comercio electrónico y proporción de ventas en línea sobre ventas totales en la Unión Europea (27 miembros), 2011-2020
(En porcentajes)



Fuente: Eurostat, [en línea] https://ec.europa.eu/eurostat/web/digital-economy-and-society/data/database.

Para el estadio del pago, los agentes más relevantes son los bancos, instituciones financieras que gestionan las tarjetas de crédito/débito o gestionan sistemas de pago por medio del smartphone. Una ONE que quiera acceder a los datos de comercio electrónico puede llegar a un acuerdo de colaboración con el intermediario financiero. Se detallan más adelante dos experiencias de colaboración que existen en España y Argentina.

E. Uso de datos de plataformas

Las plataformas intermediadoras que facilitan el maridaje o *matching* entre los usuarios demandantes de un servicio y los ofertantes de los mismos manejan datos tanto del lado de la demanda como de la oferta, haciendo uso de algoritmos y de IA con datos de alta frecuencia. Existen diferentes plataformas con modelos de negocio bien diferenciados que deben ser tenidos en consideración a la hora de acometer mediciones con los datos que manejan. Hay plataformas que facilitan la compra/venta de un bien (plataformas de intermediación), mientras que otras facilitan tan solo la comunicación entre individuos (redes sociales). OMC y otros (2023) definen una plataforma de intermediación como "Interfaces en línea que facilitan, pagando una tarifa, el acceso directo e interacción entre múltiples compradores y múltiples vendedores, sin que la plataforma tome propiedad económica de bienes o la prestación de los servicios que se venden (intermedio)". A su vez, se definen servicios de intermediación digital como "servicios de intermediación en línea que facilitan transacciones entre múltiples compradores y múltiples vendedores a cambio de una tarifa, sin la unidad de intermediación asume la propiedad económica de los bienes o la prestación de los servicios que se venden (intermedio)".

Existen plataformas de intermediación para muchos servicios y bienes, y con modelos de negocio diferentes. Hay plataformas que se especializan en la compra/venta de bienes de segunda mano. Las hay que posibilitan la contratación de servicios de alojamiento o de empleo, financieros, de auditoría, consultoría o de software, entre otros. Muchas de ellas ofrecen su servicio de intermediación sin tener

en cuenta fronteras o regímenes arancelarios o fiscales diferentes. Estas plataformas recogen información de ambos lados del mercado. Los datos suelen ser de alta frecuencia y pueden incluir identificadores locales o de grupos sociales o de edad de interés. Parece natural contar con los datos obtenidos y gestionados por estas plataformas para conocer mejor ciertas actividades económicas que de otro modo son medidas de un modo parcial.

Una experiencia de interés es la protagonizada por Eurostat/Comisión Europea y las principales plataformas de servicios de alojamientos en la UE (Airbnb, Booking, Expedia y TripAdvisor). Con el fin de aproximar mejor los flujos de turistas en la UE, Eurostat ha establecido colaboraciones diversas en los últimos años con estas plataformas. Para hacerse una idea del volumen de actividad a analizar, solo en 2019 en la UE se registraron más de 512 millones de personas/noche en este tipo de establecimientos. El desafío que existe con los métodos tradicionales —fundamentalmente basados en encuestas a alojamientos turísticos de frecuencia anual— es que no recogen bien los alojamientos en establecimientos pequeños o muy pequeños (de menos de 10 camas), como tampoco los alojamientos de muy corta duración ni el detalle geográfico de la actividad. Tampoco es posible controlar por otras dimensiones socioeconómicas relevantes, ya sea del usuario (demanda) o del establecimiento (oferta).

Eurostat comenzó la colaboración con Airbnb en 2015 pero ella no dio grandes frutos al requerir Eurostat demasiados datos y desagregaciones que la plataforma no estaba dispuesta a dar. En 2017 Eurostat decidió crear un grupo multidisciplinar de estadísticos, académicos y expertos legales y ampliar la discusión a todas las plataformas de contratación relevantes en la UE. En 2020 Eurostat y las cuatro plataformas llegaron a un acuerdo de compartición de datos por el cual las plataformas proveen a Eurostat de datos agregados a nivel de población, país y región con frecuencia trimestral de los alojamientos reservados para estancias breves[23].

El acuerdo otorga la tarea de recogida de los metadatos, validación y tratamiento a Eurostat, quien a su vez, una vez tratados los datos los traslada a las ONE de cada país miembro[24]. Es importante señalar que solo existe una "puerta de entrada" de los datos, que es Eurostat, *vis a vis* las plataformas. Por lo tanto, los criterios de recogida y tratamiento inicial de los datos fuente son comunes.

Los datos que se recogen trimestralmente son: número de estancias, número de noches por estancia reservada y número de noches por persona. Los datos se proporcionan con desagregación por ciudad (unas 250 ciudades importantes) y el establecimiento se identifica como grande (más de 10 camas) o pequeño. No se recogen datos sobre el precio pagado, ni el monto total pagado por la estancia.

Los datos procedentes de las plataformas tienen sus limitaciones y características propias, en especial si se quieren comparar con, o complementar, las estadísticas tradicionales de turismo en cada país. Por ejemplo, no se proporcionan datos de la capacidad de cada establecimiento —solo se dan las noches reservadas por individuos concretos—, lo cual unido al hecho de que un mismo establecimiento puede anunciarse en varias plataformas a la vez puede conducir a la doble contabilidad de dicho establecimiento. Además, no se puede agregar las reservas (noche) como aproximación a la capacidad total de este tipo de estancias, dato importante para el análisis con respecto a establecimientos tradicionales como los hoteles. Un avance en desarrollo acometido por las ONE en Finlandia y España hace una recogida por internet (webscraping) de todos los establecimientos, para después hacer corresponder las reservas —provistas por las plataformas directamente— por día, local y ciudad.

---

[23] Para los detalles metodológicos, acceso a datos y análisis hechos, ver [en línea] https://ec.europa.eu/eurostat/web/experimental-statistics/collaborative-economy-platforms.

[24] Como ejemplo de análisis y puesta a disposición de datos a nivel de los países miembros, véase el ejemplo del INE en España [en línea] https://www.ine.es/experimental/ocupacion/experimental_ocupacion.htm?L=0.

Otra fuente de doble contabilidad surge por el hecho de que un establecimiento puede anunciarse por una plataforma a la vez que participa en la encuesta tradicional de turismo que suele llevar a cabo la ONE en el país concreto. Así, la misma reserva puede ser contabilizada dos veces, y los métodos existentes no permiten una limpieza precisa.

En síntesis, las plataformas que facilitan el comercio electrónico son una fuente importante de evidencia sobre el comercio internacional en servicios o en bienes, con registros precisos no solo de la transacción sino también de los individuos o agentes que participan en ella. La colaboración de estas plataformas ya sea con Bancos Centrales o con las ONE de cada país redunda en una mejor aproximación al comercio internacional y en las mediciones de los componentes de la Balanza de Pagos.

II. Mejores prácticas en el uso de nuevas tecnologías en la recopilación de datos relacionados con el comercio internacional: experiencias internacionales

A. Experiencias de Big Data de ONE

A continuación, se describen algunos ejemplos de ONE de varios países con respecto al uso e integración de Big Data en su producción estadística periódica. Algunas ONE han creado unidades específicas especializadas en Big Data. Por ejemplo, la Oficina Central de Estadísticas de los Países Bajos (CBS, por sus siglas en inglés) creó el Centro para Estadísticas de Big Data (CBDS, por sus siglas en inglés) en 2016. Este centro cuenta con acuerdos de colaboración con las ONE de otros países. Colabora estrechamente con Eurostat en varios proyectos y tiene colaboraciones con distintos proveedores de datos (como Google) y otros agentes involucrados en la cadena de valor (Dell, IBM, KPN, Microsoft, SURFSara, TMO, entre otros). Asimismo, tiene también acuerdos de colaboración con universidades y centros de investigación.

El CBDS tiene tres objetivos: i) producir datos en tiempo real, ii) proveer un mayor nivel de desglose en las estadísticas existentes, como por ejemplo estadísticas a nivel regional o urbano, y iii) reducir la carga que recae sobre las empresas e individuos en la colecta de datos sobre sus hábitos, características y comportamientos. Para conseguir estos objetivos el CBDS ha llegado a acuerdos de colaboración con agentes privados o públicos para conocer mejor las técnicas y retos encontrados en el tratamiento de Big Data.

El CBS participa también en la Colaboración en Investigación y Metodología de la Estadística (CROS, por sus siglas en inglés). Es una colaboración mediante un portal entre investigadores, ONE y la Comisión Europea, donde se intercambia información y metodologías sobre los distintos aspectos del Big Data. Como fruto de esta colaboración, en 2016 se creó la red de ONE de la Unión Europea para el trabajo conjunto y compartición de experiencias en Big Data, ESSnet. Desde esta red se analizan desde nuevas propuestas para la arquitectura de producción de Big Data, extracción de datos de empresas a

partir de sus sitios web mediante web scraping y metodologías para la implementación de contadores inteligentes de energía (Smart Energy) hasta el seguimiento marítimo de buques de mercancías[25].

Una iniciativa de interés, apoyado por la CEPAL, en Brasil es una colaboración entre el Instituto Brasileño de Geografía y Estadística (IBGE) y un centro experto en economía digital y autoridad que concede los dominios de internet (Nic.br). Ambos buscan medir el tamaño y características intrínsecas de la economía digital en Brasil utilizando técnicas de Big Data y combinándolas con estadísticas oficiales[26].

Una dificultad clara en este proyecto es medir, con fuentes distintas, de lo que conforma la "economía de internet" (el dominio). Es un concepto más amplio que el comercio por internet o la actividad económica basada en actividades de información y comunicaciones, tal y como viene recogida en la contabilidad nacional. Se utilizan dos grandes fuentes de datos para delimitar el dominio.

Por un lado, el IBGE dispone de un registro de empresas, donde cada empresa del país debe registrarse a efectos administrativos y fiscales. En este registro cada empresa se identifica, entre otros campos, con su actividad principal según la CIIU (Rev.4). Dicha clasificación no está adaptada a la economía digital, ya que recoge solo la actividad en cuanto a producción y venta de bienes y servicios principales de la empresa, pero no el modo en que esta actividad principal es conducida[27]. Esto es, una empresa que gestione contratos financieros de otorgamiento de créditos pertenece a la clase 6492 según la CIIU independiente si realiza toda su actividad por internet o no. Si un analista quiere saber cuán intensivo es el uso de la contratación en línea en Brasil, esta clasificación no le aporta información alguna.

Por otro lado, y de un modo innovador, el proyecto utiliza otra gran fuente de datos: ii) los dominios de internet (en este caso com.br) o registro de sitios web de empresas. Con técnicas de web scraping, se puede conocer si el sitio web de la empresa ofrece créditos u otro servicio financiero en línea. Si ofrece servicios en línea de créditos a consumidores, entraría dentro del dominio de "economía digital". Dado que el número de sitios web habilitados es enorme, es necesario hacer uso de técnicas de Big Data para el manejo de este volumen de información.

Resultan claras las dificultades que plantea el matching de estas dos fuentes de datos: por un lado, la fuente oficial del Registro de Empresas con la clasificación tradicional de actividades económicas, y por otro lado, los dominios de internet que permiten la extracción de la actividad subyacente (no necesariamente la registrada). Ni las clasificaciones de actividades son iguales, ni los dominios de empresas que se obtienen de una y otra fuente son siquiera parecidos. Hay múltiples desafíos en el uso de los dominios de internet (duplicidades, varios sitios web pertenecientes al mismo grupo empresarial pero cada uno gestionado por un distribuidor diferente, información parcial, imposible de extraer, distinto nivel de "actividad" del sitio web, entre otros). Otro conjunto de retos muy importantes se refiere a la compatibilidad de ambas fuentes y llegar a un dominio de empresas de internet único y relevante con el cual aproximar el tamaño de la economía de internet.

A continuación, se revisan tres experiencias con acuerdos de colaboración de una ONE o agencia internacional con entidades financieras de medios de pago con el objetivo de conocer el volumen y bienes transaccionados por comercio electrónico.

---

[25] Una experiencia en la medición de contadores inteligentes se revisa más adelante.

[26] Big Data para la medición de la economía de internet, véase [en línea] https://www.cepal.org/sites/default/files/presentations/bigdata-medicion-economia-internet-ibge-brasil-2020.pdf.

[27] Ver por ejemplo ONU (2011).

1. Acuerdo de la autoridad reguladora y de competencia de España (CNMC) con entidades de medios de pago

A inicios de los años 2000, la autoridad reguladora de las telecomunicaciones de España (actualmente Comisión Nacional de los Mercados y la Competencia, CNMC) llegó a un acuerdo con las principales entidades de medios de pago-gestores de tarjetas de crédito/débito en España. El acuerdo implicaba que estas entidades entregarían datos anonimizados de todas las transacciones pagadas con tarjetas bancarias con periodicidad trimestral. La CNMC, por su parte, agregaría y explotaría estos datos con el fin de proporcionar informes de seguimiento del volumen de comercio electrónico en España (doméstico e internacional) y de los bienes o servicios más intercambiados. Por tanto, la CNMC recibe un censo de todas las transacciones electrónicas a través de terminales de punto de venta virtuales localizados en España. Nótese que no se recopilan pagos efectuados por transferencia bancaria ni en efectivo.

Los campos a compartir consisten en:

- Transacciones satisfechas con tarjetas de crédito o de débito, cuando la tarjeta (y el usuario) no están presentes físicamente. Es importante resaltar que esto limita el volumen de pagos hechos a B2C, esto es transacciones entre empresas e individuos. Se ignora por tanto el intercambio B2B electrónico, o transacciones entre empresas.
- Número de transacciones y montos totales.
- Por transacción: denominación global del bien o servicio intercambiado, importe total, origen de la operación (país) y destino de la operación de pago (país).
- Día/mes del año.

Los datos de las operaciones son sometidos a un proceso de anonimización por parte de las entidades de medios de pago, de tal modo que no resulta posible identificar transacción con el individuo (o cuenta de débito o crédito).

La CNMC, al recibir los datos de todas las transacciones, las reclasifica en base a la NACE Rev.4 hasta 4 dígitos de agregación. Este nivel de detalle resulta conveniente para la correcta clasificación de muchos bienes y servicios, pero no es suficiente para otros. Con el fin de dar el mayor detalle posible, la CNMC utiliza simultáneamente la Clasificación de Productos por Actividad (CPA), que es compatible con la NACE y ofrece la posibilidad de asignar adicionalmente hasta 5 o incluso 6 dígitos de actividad o industria a cada producto. La CPA se utiliza para un subconjunto reducido de productos dada la importancia relativa que éstos tienen sobre el total.

La explotación trimestral de los datos de comercio electrónico y el detalle metodológico de colecta, tratamiento y explotación de estos microdatos están disponibles al público, con posibilidad de descarga de datos y principales análisis descriptivos disponibles[28].

2. Acuerdo BBVA-OCDE

Otra experiencia de interés relativa al intercambio y explotación de datos con el fin de ofrecer un análisis no solo el volumen sino los determinantes socioeconómicos individuales de comercio electrónico es el acuerdo entre un banco privado, BBVA, y la OCDE. El acuerdo consistía en el intercambio, una vez anonimizados, de los datos sobre transacciones de pago con tarjeta efectuadas en línea por residentes en España (los denominados *payments with card-non present*, o pagos con tarjeta donde la tarjeta bancaria no está físicamente presente). Los campos a compartir consistían en: fecha y lugar, pago total efectuado,

[28] Para más detalles, véase http://data.cnmc.es/datagraph.

origen-destino del pago y bien o servicio intercambiado. Este acuerdo tenía un objetivo —la medición del volumen y características de los usuarios del comercio electrónico— y una vez cumplido, caducó.

Con esta colaboración fue posible conocer los volúmenes y los productos más frecuentemente intercambiados en línea. Además, con bases de datos alternativas —sobre condiciones socioeconómicas de individuos/hogares— y modelos logísticos multinomiales, fue posible descubrir los determinantes más relevantes en la adopción del comercio electrónico (OCDE, 2017a).

3. La medición del comercio digital en Argentina

Recientemente el Instituto Nacional de Estadísticas y Censos (INDEC) ha realizado una medición de un dominio de transacciones más amplio que el puro comercio electrónico. Se trata de medir el volumen de **importaciones de comercio digital**, esto es, cualquier transacción facilitada por medios electrónicos que permite bien sea la orden de compra (contratación) o el despacho del bien o servicio hasta el consumidor final. De un modo similar al comercio electrónico, se trata de un tipo de comercio cuya medición se basa en el tipo de transacción, esto es, en cómo se ha acometido la misma, no en el tipo o naturaleza del bien transaccionado (Ministerio de Hacienda de Argentina e INDEC, 2019).

Como fuente fundamental de información, el INDEC contó con un registro de pagos por tarjeta de crédito o débito hechos desde Argentina por la contratación de servicios digitales, para lo cual debieron definir en primer lugar los ofertantes (extranjeros) de estos. Se trata de servicios digitales demandados por hogares o individuos, y para definir el dominio inicial de estos servicios se partió de la Encuesta Nacional de Consumo de Bienes Culturales (ENCBC). El valor monetario de estas transacciones se estima a partir del Impuesto al Valor Añadido (IVA) que se aplica a dichos servicios. En la legislación argentina los servicios con valor igual o inferior a 10 dólares no pagan tributo alguno y se eliminaron del análisis. El resto de los productos culturales pagados por tarjeta de crédito pagan un IVA común de 21%.

Para cada transacción se recogían los campos siguientes: IVA pagado, país de residencia de la empresa ofertante del bien, fecha, nombre de la empresa ofertante. Una dificultad encontrada es que una misma entidad legal o empresa puede aparecer con varios nombres o denominaciones en distintas transacciones, por lo que el INDEC debió crear un listado de empresas ofertantes de bienes culturales fuera de Argentina.

Las actividades recogidas corresponden a: servicios financieros, de telecomunicaciones, informáticos, de oferta de contenidos, juegos en línea, servicios de información, de publicidad, estudios de mercado, de empleo, de educación, y otros servicios personales. Para aquellos servicios que son de pago recurrente (mensual), se utilizó el stock de suscriptores en un momento concreto y el pago medio por suscripción fue obtenido a través de una encuesta a los usuarios.

En cuanto a los resultados, fueron los servicios "audiovisuales y conexos" los más comúnmente contratados de modo digital, seguidos por los servicios de informática. El país de procedencia de la mayor parte de las importaciones de bienes culturales fue Estados Unidos, seguido de Alemania, España y Reino Unido.

Este es un claro ejemplo de colaboración entre una autoridad pública y entidades privadas que proveen de datos necesarios para la medición del comercio internacional realizado en línea. La evidencia obtenida sobre las categorías de servicios intercambiados no son un sustituto de las partidas recogidas por el Banco Central o la ONE para la contabilidad de la balanza de pagos. La propia definición de transacción recogida en esta experiencia abarca más que el dominio de transacciones imputadas en la balanza de pagos. Pero es un ejemplo útil de cómo ambas fuentes, la tradicional de balanza de pagos, y la nueva de transacciones electrónicas de servicios, se pueden complementar y muestra claramente la potencialidad del Big Data: mayor detalle por transacción, mayor frecuencia de recogida y reducción de costos de colecta y tratamiento.

B. Experiencias de Big Data en áreas concretas

En esta sección se describen experiencias concretas de explotación de Big Data en áreas diversas como la medición y gestión de flujos turísticos, la gestión de puertos, el tráfico marítimo de mercancías y la logística y el comercio electrónico.

1. Turismo

En un buen número de países donde la industria del turismo es relevante se vienen observando experiencias de diversos tipos utilizando MPD[29]. Al poder registrar el operador no solo la localización, sino los tráficos y detalle de los mismos de los usuarios, ya sean nacionales o foráneos, de una red concreta, es posible el análisis y posterior uso de dicha información para la mejora de las condiciones que se ofrecen a los turistas. A través del seguimiento de la localización (mediciones repetidas), por ejemplo, se identifican los lugares con mayor número de visitas, si puede haber cuellos de botella ya sea en el transporte público y o en redes de carreteras o en el stock de plazas turísticas en oferta, si hay correlaciones entre los lugares visitados y otras variables de entorno (clima, temperatura, atracciones turísticas, conciertos), o si hay fidelización por parte de turistas. Además, y gracias a los datos CDR de consumo móvil efectuado por usuarios, se sabe el consumo de itinerancia, las llamadas recibidas/enviadas, y con qué países se han comunicado los visitantes. También se puede analizar posibles efectos-arrastre, o el efecto de ciertas medidas de gestión (precios, descuentos, campañas de publicidad) sobre el éxito de un evento.

En flujos de turistas se utilizan tres tipos de Big Data principalmente, cada uno con sus fuentes diversas:

i) Datos generados por los usuarios (UGD). Son datos que generan y activan los propios usuarios finales al, por ejemplo, subir una fotografía en una red social, participar en un chat o utilizar la aplicación de mapas.

ii) Datos de terminales (smartphones). Pueden venir de distintas fuentes. Se trata de colecta de datos directamente del terminal a partir de algún uso o red que se utilice, como datos de GPS, datos de itinerancia internacional en redes móviles, o datos recogidos a través de una red de área local (bluetooth, wi-fi).

iii) Datos de transacciones. Son datos recogidos a partir de una actividad como puede ser el número de visitas a un sitio web, el número de transacciones de pago electrónico, o de contrataciones de comercio electrónico.

La colecta de Big Data de estas fuentes alternativas puede también utilizarse como complemento de las estadísticas oficiales que la ONE o el Banco Central publican periódicamente sobre el turismo y sectores relacionados. El Banco Central de Estonia, por ejemplo, utiliza desde 2012 Big Data para producir estadísticas de mayor frecuencia sobre la balanza de pagos, estimando casi en tiempo real los flujos de pagos que ocurren debido a la entrada de turistas.

Cuatro países de la Unión Europea (Alemania, Estonia, Finlandia y Francia) publicaron un estudio de viabilidad sobre el uso de indicadores de movilidad para la aproximación de los flujos turísticos y su posible uso por las estadísticas oficiales. Encontraron que el Big Data conlleva sesgos propios, diferentes de los encontrados en encuestas tradicionales y que la compatibilidad de ambas fuentes aún no era precisa. El análisis de los sesgos que fuentes concretas de Big Data pueden tener en la medición de turistas ha sido analizado extensivamente en los últimos años (Schmuecker y Reif, 2022; Paul i Agusti, 2020; Salas-Olmedo et. al., 2018 y Saluver et. al. 2020).

---

[29] Datos de posicionamiento móvil provistos por operadores de redes móviles al localizar el smartphone (Mobile Positioning Data).

Un área original de aplicación del Big Data a la gestión de sitios turísticos utiliza a los turistas como generadores de contenidos- al subir una fotografía o comentario a una red social, por ejemplo, con el fin de detectar impresiones o percepciones del viaje, del lugar visitado o del entorno (Paul i Agusti, 2020; Yang, et. al., 2022; Marine-Roig, 2019). Con técnicas de Big Data hay autores que explotan esta gran cantidad de información —especialmente ayudándose de *Deep Learning* para el reconocimiento de imágenes— e inferir así la percepción de los turistas en un evento o localidad precisa. Es lo que se denomina "content perception", la inferencia de las valoraciones de los individuos a través de un medio que se ha compartido en una red social (Reza Alaei, et. al, 2017).

Cuadro 4
Ejemplos de uso de Big data para el análisis del turismo

Tipo de experiencia	Fuente	País	Año
MPD para flujos turismo dentro/fuera de UE	http://ec.europa.eu/eurostat/documents/747990/6225717/MP-Consolidated-report.pdf	Estonia	2014
Eventos turísticos y deportivos y visitas repetidas	https://ojs.utlib.ee/index.php/TPEP/article/download/878/855	Estonia	2010
MPD pasivo y relación entre temperatura y flujo de turistas	https://www.semanticscholar.org/paper/WEATHER-DEPENDENCE-OF-TOURIST%27S-SPATIAL-BEHAVIOUR-Järv-Aasa/f2d5966c48a4a1db8f3a3f3828604737444371aa	Estonia	2016
Sesgos en fuentes de datos tradicionales y Big Data con GPS y MPD	https://www.sciencedirect.com/journal/annals-of-tourism-research-empirical-insights	Alemania	2022
Datos masivos generados por turistas y análisis de flujos		España	2015
Análisis de las percepciones de los turistas: análisis con deep learning de fotos compartidas	https://www.researchgate.net/publication/335421354_Tourism_Management_Discovering_the_tourists%27_behaviors_and_perceptions_in_a_tourism_destination_by_analyzing_photos%27_visual_content_with_a_computer_deep_learning_model_The_case_of_Beijing	China	2019
Contenido generado por turistas en redes sociales y análisis de "content sentiment"	https://journals.sagepub.com/doi/abs/10.1177/0047287517747753		2017
Mapping tourism hot spots using Instagram images in Africa	https://en línealibrary.wiley.com/doi/abs/10.1002/jtr.2360	África	2020
Using tourist footprints: comparison of different big data sources	https://arxiv.org/abs/1705.07951	España	2018

Fuente: Elaboración propia del autor.

2. Experiencias de explotación de datos AIS

a) Los "corredores 5G" transfronterizos en la UE

Una experiencia interesante es la de los corredores transfronterizos que se están implementado en algunas fronteras de países de la Unión Europea (UE) con el fin de facilitar, gestionar y optimizar los tráficos de personas y de mercancías por carretera y de promover la conducción autónoma.

Son muchas las rutas incluidas en 2018 y algunas más en el programa Horizon 2020 de la UE para el despliegue de redes 5G: *5G Carmen* —uniendo del tránsito por carretera entre Boloña (Italia) y Munich (Alemania)—, *5G-Mobix*: uniendo carreteras entre España y Portugal y otro entre Grecia y Turquía, además de crear zonas 5G en cinco grandes concentraciones urbanas en la UE (Versalles, Berlín, Stuttgart, Eindhoven y Espoo), entre otros muchos; *5G- Blueprint*- soluciones de red entre Bélgica y

Países Bajos para las vías de transporte por canales, ríos, autopistas y puertos, y 5G- Rail: diseñado una red de comunicaciones 5G en diversos puntos de líneas de ferrocarril en la UE[30].

Muchas de estas iniciativas se han desarrollado gracias a la colaboración público-privada para el 5G que promovió la Comisión Europea (5GPPP)[31], donde participan empresas, operadores y gobiernos diseñando estrategias y pruebas de experiencias 5G para la movilidad.

b) Pruebas de "pelotón de camiones" en Japón

El Ministerio de Comunicaciones de Japón ha promovido la realización de testeos del 5G, como por ejemplo la experiencia de *truck platooning*, por la cual un conjunto de camiones grandes son conducidos en fila simultáneamente por una autopista a través de una unidad central que dispone de sensores de control de todos los camiones, información que recoge en tiempo real y después esta unidad envía ordenes de conducción a toda la fila. Todas estas comunicaciones se realizan con redes 5G de alta fiabilidad (baja latencia). Gracias a la alta fiabilidad de la red 5G utilizada, se consigue que los camiones no se separen más que unos pocos metros entre ellos, reduciendo de este modo la resistencia al viento y por tanto el consumo de gasolina total. También se consigue reducir la congestión que generaría ese mismo número de camones conduciendo independientemente por la misma vía, aumentando la seguridad del transporte.

En la UE una experiencia similar está siendo llevada a cabo por un consorcio de fabricantes de grandes camiones (como Daimler Benz, Iveco, MAN, Scania y Volvo), junto con socios tecnológicos, de gestión de redes y de componentes[32]. Es una iniciativa financiada parcialmente por la UE en su programa Horizon 2020. Por ahora se han realizado solo pruebas de pelotones de camiones. Se trata de una experiencia abierta a todos los fabricantes de grandes camiones para el transporte en "flota" (o pelotón) simultáneo con el fin de mejorar la eficiencia en el tráfico, reducir el consumo de petróleo y aumentar la seguridad.

c) Gestión de los puertos con 5G

En Italia el puerto de Livorno, puerta marítima de entrada a la Toscana, en colaboración con un socio tecnológico (Ericsson), ha desplegado una red de 5G, que constituye la base de todo un sistema de transporte y logística integrado que abarca no solo las operaciones de carga/descarga en puerto, sino la gestión de mercancías hacia las principales ciudades y polos industriales de la Toscana. La idea se originó en encuentros entre el puerto, Ericsson y un consorcio de universidades italianas que eligieron como base de experimentación de las posibilidades del 5G a uno de los mayores puertos del Mediterráneo. Uno de los objetivos es lograr una mayor sostenibilidad ambiental, en línea con los Objetivos de Desarrollo Sostenible de las Naciones Unidas.

El despliegue del sistema logístico inteligente supone la instalación de cámaras, sensores y objetos, todos conectados a una red 5G de alta fiabilidad que conforma el sistema integrado de comunicaciones. A este sistema se le añaden tecnologías como IA y Realidad Aumentada. De este modo, se posibilita realizar cargas/descargas de mercancías con robots y emitir al sistema la situación de stock en cada barco en tiempo real con el fin de coordinar la salida de la carga por camión o tren. La información generada por todas conexiones es utilizada con IoT y con IA con el fin de asegurar el uso rápido de las facilidades del puerto y sus conexiones con el exterior.

---

[30] Forman parte de la iniciativa Connecting Europe Facility 2022-27 [en línea] https://digital-strategy.ec.europa.eu/en/policies/cross-border-corridors.

[31] Para más detalles, véase [en línea] https://5g-ppp.eu/trials-and-pilots-for-connected-and-automated-mobility/.

[32] Para más información, véase ENSEMBLE, ENabling SafE Multi-Brand pLatooning for Europe [en línea] https://cordis.europa.eu/project/id/769115.

Aunque el proyecto lleva una duración escasa, los resultados ya indican una reducción significativa de los costos operativos del puerto, menor uso de combustibles fósiles y energía y menores tiempos de almacenaje. Ericsson estima que se ha conseguido una reducción de 2,5 millones de Euros solo en el tiempo que se ha conseguido ahorrar de atraque de los barcos y un 25% más de productividad en el uso de las grúas en muelles gracias al control y gestión remotos de las mismas. Otro aspecto interesante de esta experiencia es la generación de información que el sistema inteligente ha supuesto para los socios y para el ámbito académico, con la posibilidad de explotar esta información y analizar mejoras en la eficiencia de la gestión.

Otro resultado de interés es la creación de un "gemelo digital" (*digital twin*), esto es, una réplica del puerto de Livorno en 3D que simula todas las actividades sometidas a control y gestión de interés. Esta experimentación se ha realizado también en otros puertos, como los de Singapur y Shanghai.

A modo de ejemplo de cómo el "gemelo digital" —o la recreación por Realidad Aumentada— puede ayudar en el proceso de carga/descarga en puerto, se describe la secuencia de acciones hombre-máquina— sistema que tienen lugar:

- Se recoge una mercancía en un contenedor, y se la identifica y registra por el sistema de control en una base de datos.
- Una grúa o carretilla transporta la carga a la zona de almacenamiento. En cuanto una cámara detecta el cargamento llegando al almacén comienza el seguimiento del cargamento.
- Utilizando Realidad Aumentada, el conductor o vehículo es guiado al lugar donde debe descargar el cargamento.
- Una vez que el cargamento ha sido almacenado, las cámaras controlan que haya sido debidamente instalado. La posición final es identificada y cargada en la base de datos.
- Cualquier reposicionamiento del cargamento en almacén es a su vez grabado por la cámara y registrado en la base de datos.
- En la fase de carga del barco, los vehículos son monitoreados por cámaras. Al entrar en almacén, el conductor de la grúa recibe por Realidad Aumentada en 3D la posición del cargamento a llevar. Las cámaras de control pueden controlar zonas más claras u oscuras, con lo que la operación se puede realizar en cualquier escenario (día/noche).
- Al disponer de un "gemelo digital" del cargamento, con sus características propias de dimensiones, contenido, necesidades de temperatura, entre otras, y del espacio en terminal de almacén y carga, es posible replicar cualquier proceso con Realidad Aumentada. Por ejemplo, el conductor puede, antes de recoger el cargamento y gracias a la Realidad Aumentada, identificar el cargamento y conocer sus características: peso, dimensiones, necesidades específicas, tiempo en almacén, etc.

Los operadores de telecomunicaciones también han activado acuerdos con instalaciones de logística. Vodafone, por ejemplo, está inmerso en proyectos de redes 5G en los puertos de Huelva, Barcelona y Algeciras, en España, y en Plymouth en el Reino Unido, entre otros.

Las posibilidades que ofrecen las redes 5G son claras, pero hay experiencias interesantes explotando el estándar anterior (4G/LTE). Por ejemplo, en el puerto de Oakland (California) se ha conseguido unir puntos de embarque y transporte en el puerto con red privada 4G, de costo menor en cuanto al despliegue. Esta red ha sustituido a la que existía compuesta por diferentes puntos de conexión wifi, que utiliza también espectro no licenciado (de uso libre).

Cuadro 5
Ejemplos de uso de 5G en puertos, rutas transfronterizas y tráfico

Tema	Fuente	País	Año	Socios
Comunicación directa entre vehículos y entre vehículos e infraestructura vial	https://www.t-press.cz/en/press-releases/press-news-archive/t-mobile-is-testing-cv2x-data-technology-for-vehicles-and-infrastructure.html	República Checa	2019	T-Mobile, Skoda
Vehículos autónomos utilizando redes 5G	https://technode.com/2019/07/11/bmw-china-unicom-5g/	China	2019	China Unicom, BMW
Redes privadas de 5G para la inspección de componentes de aviones	https://www.globenewswire.com/news-release/2020/02/27/1991601/0/en/Nokia-deploys-5G-private-wireless-network-for-Lufthansa-Technik-virtual-inspection-trial.html	Finlandia	2020	Lufthansa Nokia
Redes 5G privadas para gestión del puerto de Zeebruegge	https://www.nokia.com/about-us/news/releases/2020/03/25/nokia-completes-phase-one-of-belgian-port-of-zeebrugge-digitalization-with-5g-ready-private-wireless-network/	Bélgica	2020	Nokia
Corredores transfronterizos en la UE	https://digital-strategy.ec.europa.eu/en/policies/cross-border-corridors	Unión Europea	2020	
Geolocalización de barcos puerto de Barcelona y Huelva	https://smartports.tv/vessel-geolocation-system	España	2020	Vodafone, IBM, Huawei

Fuente: Elaboración propia del autor.

III. Los desafíos en la explotación de Big Data de movilidad

Obtener información a partir de las trazas que dejan los usuarios de redes móviles tiene ventajas claras con respecto a las fuentes más tradicionales de colecta, como la precisión de la información que se consigue y la frecuencia en la recogida de los datos, que permite un análisis dinámico de un comportamiento social. Pero también existen dificultades a superar (UN Global Working Group on Big Data for Official Statistics, 2019):

- Retos administrativos en acceder a los datos de movilidad.
- Dificultades técnicas del procesamiento de datos y retos metodológicos.
- Datos de referencia, evaluación y estimación: nuevos métodos para el tratamiento de Big Data.

En los siguientes párrafos se presentan brevemente estos conjuntos de desafíos.

A. Retos administrativos en acceder a los datos de movilidad (MPD)

Los propietarios o los gestores y responsables de los datos de movilidad son los operadores móviles con red propia. Son los responsables de su colecta, almacenamiento, tratamiento y procesamiento. Estos operadores están sometidos a la legislación de datos personales, ya sea la legislación general sobre la privacidad o la legislación específica aplicada al sector de las telecomunicaciones. Por regla general la legislación establece que un operador móvil puede hacer uso de sus datos para su propia gestión de la red, facturación, calidad del servicio u otro servicio asociado.

Se considera que los datos son privados, y por tanto confidenciales, cuando son personales, esto es, cuando directa o indirectamente pueden identificar a una persona usuaria del servicio. Una vez que los datos han sido agregados o anonimizados de algún modo, ya no son considerados como privados, aunque aún pueden ser considerados como "confidenciales" dependiendo de la legislación específica del país.

Cualquier empresa que obtenga datos privados, como por ejemplo un operador de telecomunicaciones, está sometida a la legislación nacional de privacidad y datos personales. Por lo general, tiene prohibido tanto explotar los datos privados con fines que no sean su propia gestión del negocio principal como proveer esos datos privados a terceros. Lo que sí es posible, comunmente, es trasladar o vender los datos que tenga siempre que haya anonimizado con algún procedimiento los datos fuente. Procedimientos de anonimización de datos hay muchos, y lo que en ocasiones está difuso es el nivel de anonimización que se exige. Adicionalmente a la legislación de datos personales, los operadores de telecomunicaciones están sometidos a la legislación específica sobre telecomunicaciones, que a menudo establece condiciones estrictas para el uso de datos de los clientes. Con el fin de dotar de mayor seguridad a los agentes que manejan datos personales y los procesos a seguir y grado de anonimización a obtener, las Agencias de Protección de Datos Personales en varios países publican guías que intentan orientar dichos procesos[33].

Una ONE de cualquier país está, a su vez, sometida a la legislación propia de Estadísticas Oficiales. Esta le otorga la potestad en la colecta de datos de cualquier individuo en el país y establece la obligación de colaboración por parte de individuos y empresas, pero a su vez impone condiciones en la explotación y uso de la información recogida. En especial, queda prohibida la publicación de datos que pudieran identificar a algún individuo o empresa encuestado o participante en alguna colecta. Si bien todo individuo o empresa en un país tiene la obligación de colaborar con la ONE en sus tareas, cuando se trata de compartir Big Data las ONE de cualquier país han seguido un camino de acuerdos de colaboración con los agentes que disponen de esos datos. Hay ONE que han solicitado mayores poderes en la legislación nacional con el fin de poder obtener Big Data de los operadores privados cuando así lo necesiten en sus planes estadísticos.

Una de las dificultades que enfrentan las ONE es precisamente el heterogéneo y en ocasiones difuso marco normativo que existe en su relación con otros agentes privados cuando solicita el uso de Big Data de éstos. Si se quiere promover el uso de Big Data por parte de las ONE, o de instituciones públicas o privadas en general, sería útil un marco normativo lo más homogéneo posible que garantice la confidencialidad de los datos personales a la vez que facilite los acuerdos entre agentes para la explotación del Big Data.

El nivel de agregación de los datos aportados por el operador o empresa que dispone de Big Data es relevante. Por un lado, a mayor nivel de agregación mayor facilidad en la transmisión y uso de Big Data. Por otro lado, altos niveles de agregación precisamente eliminan dimensiones críticas en la explotación del Big Data. Un operador de red móvil puede, por ejemplo, proveer a una ONE datos agregados de visitantes extranjeros en una ciudad concreta, incluso su evolución diaria por un periodo de 15 días. Pero si estos datos no son longitudinales, esto es, si no permiten seguir individuo a individuo a lo largo del periodo relevante, no será posible hacer estadística alguna de número de visitantes —para lo cual saber la llegada y salida en la ciudad de cada usuario de terminal móvil es crítico—, ni tampoco identificar si se trata de viajes de turismo o por otros motivos.

Un informe contratado por Eurostat donde se analiza la explotación de Big Data en el turismo en los países miembros de la Unión Europea (Ahas et. al., 2014) concluía que, a pesar de las muchas iniciativas lanzadas por las ONE con agentes proveedores de Big Data, el marco normativo aún no era lo suficientemente homogéneo en toda la UE. Las metodologías de colecta, explotación y resultados vistos no permiten aún considerar a las fuentes de Big Data como un sustituto, sino más bien como un complemento interesante, de las estadísticas oficiales de la ONE. Los principales ventajas e inconvenientes de MPD para la elaboración de estadísticas de turismo se indican en el cuadro 6.

---

[33] Como ejemplo de una guía, ver AEPD (2016).

Cuadro 6
Ventajas e inconvenientes del uso de MPD para la elaboración de estadísticas de turismo

Ventajas	Inconvenientes
Consistencia del número de viajes y noches de estancia comparado con la evidencia derivada de estadísticas "tradicionales"	Complejos procesos para el acceso a los datos de los operadores de redes móviles e incertidumbre en colaboración a futuro
Mejor cobertura de estancias en sitios de no- pago gracias al MPD	Falta de información adicional al viaje y lugar que es de interés (gastos efectuados, precios, medios de transporte utilizados…)
La posibilidad de hacer desgloses mucho más detallados, ya sea en el tiempo (día/hora), por ciudad, área, comarca	Sesgos en algunas clasificaciones como resultado del uso de definiciones distintas (visitas de muy corta estancia)
Posibilidad con Big Data de identificar eventos turísticos de interés, visitas repetidas, frecuencia en las visitas	Sobre ponderación de visitas que no son estrictamente de turismo
Proporciona estadísticas en tiempo real (casi instantáneamente)	Sobre o infra representación de flujos reales de turistas (los que no utilizan terminal móvil nunca; los que hacen mucho uso de itinerancia internacional de varios operadores a la vez o varios terminales)
Producción de estadísticas de un modo más automatizado	
Posibilidad de utilizar "estadísticas espejo" transfronterizas que identifiquen redes de turismo a través de varias fronteras, si existe coordinación en la colecta de datos entre países	

Fuente: Eurostat (2014).

Otro desafío encontrado con el uso de MPD para medir los flujos de turismo tiene que ver con la clasificación utilizada en las estadísticas públicas, por una parte, y la información que de MPD se puede extraer, por otra (Eurostat, 2013). Tradicionalmente una ONE mide en sus estadísticas oficiales "estancias" o "pernoctaciones" —medidas en días— por residentes no domésticos en sitios de pago (hoteles, pensiones, campings). Pero el MPD aportado por operadores de telecomunicaciones provee una información distinta: la presencia de un no residente en un lugar, sin poder conocer si se hospedó en un hotel (u otro sitio de pago) o no. Ambas magnitudes se correlacionan muy estrechamente, pero no miden lo mismo.

Como se ha señalado antes, el uso de MPD junto con otras fuentes de Big Data —como la reportada por usuarios de tarjetas de pago—, puede solucionar este reto con métodos estadísticos/econométricos específicos. En proyectos de Big Data el operador es quien provee los datos base, de movilidad en red, por ejemplo, pero suele ser el caso que otros agentes proveen otras capas de datos complementarias que les dan utilidad. La coordinación entre los distintos agentes que participen en un proyecto amplio de este tipo es crucial, tanto en la definición exacta de los indicadores a colectar, como en la transmisión anonimizada pero que permita explotar el corte trasversal de los datos, y en los formatos a utilizar.

B. Desafíos técnicos del procesamiento de datos y retos metodológicos

Gestionar grandes volúmenes de datos puede requerir conocimientos, servidores y una infraestructura específica para hacerlo. Dependiendo de los tipos de datos a explotar, la metodología para su tratamiento y procesamiento puede ser diferente. Algunos aspectos metodológicos que deben ser tenidos en cuenta son los siguientes:

i) **Anonimización:** los datos de localización o de uso de los individuos, ya sea de consumos móviles o de aplicaciones, caen dentro de la protección de la privacidad en cualquier jurisdicción. Es necesario el consentimiento del usuario final, o bien disponer la capacidad legal por parte del operador para recolectar y tratar esos datos, o transformarlos en otros

que garanticen la anonimidad de tal modo que los registros individuales no puedan hacerse corresponder con individuos identificables.

Por ejemplo, en la UE hay al menos tres piezas legislativas que se deben respetar en relación con los datos de movilidad: i) el Reglamento General de Protección de Datos (2016), ii) la legislación de las comunicaciones electrónicas (Directiva 2022/58/EC y Código de Comunicaciones Electrónicas (2018) —que aplica solo a los operadores—), y en su caso, iii) la Ley de Estadísticas Oficiales —que afecta solo a los organismos públicos responsables de la producción estadística (las ONE).

La Agencia Española de Protección de Datos (AEPD), al igual que agencias similares en otros países como en Colombia (Dirección de Regulación, Planeación, Estandarización y Normalización, 2014), ha publicado una guía orientativa del procedimiento a seguir con el objetivo de transformar datos privados en datos anonimizados. Conviene resaltar algunas fases y tareas de dicho proceso que se detallan en esta guía (AEPD, 2016):

- El objetivo de esta fase es "eliminar o reducir al mínimo" la posibilidad de re-identificar a los individuos recogidos en una base de datos, a la vez que posibilitar que cualquier tratamiento de los datos posterior "no implique una distorsión de los datos reales" (microdatos). Se debe tener cuidado no solo de la re-identificación directa —con base a los datos de la base de datos específica— sino también a la posibilidad de re-identificación indirecta de los individuos, esto es, utilizando otras fuentes de datos distintas e independientes.
- Principios a seguir en el proceso:
  - De proactividad: el objetivo es proteger la privacidad de datos personales, y para ello los procesos deben integrar esta garantía desde el inicio del proceso de colecta y tratamiento de los datos. Puede ser útil clasificar los datos de entrada en base a los niveles: i) datos privados, ii) microdatos[34], iii) datos de identificación indirecta[35], y iv) datos sensibles.

    En la evaluación del impacto en la protección de datos personales se debe evaluar el riesgo de re-identificación, que se debe poner en conocimiento del agente que reciba los datos.
  - De funcionalidad: los datos anonimizados deben ser útiles para los fines a los que se vayan a utilizar y deben, por tanto, no distorsionar la realidad. Puede ser necesario una reclasificación de ciudades, localidades, nombres, pero se debe acometer de tal modo que para el fin último del proyecto se trate de una re-clasificación adecuada a la realidad.
  - De privacidad a lo largo del ciclo de vida de la información: si los datos van a ser utilizados a lo largo del tiempo, o se van a integrar o explotar junto con otros conjuntos de datos, se debe garantizar la anonimización en todo el proceso y tiempo.
- Separación de tareas e independencia de equipos humanos que acometen cada parte.
- Publicación de los pasos a dar, tareas conseguidas y evaluación de riesgos.

ii) **Procesamiento de datos**: la etapa de procesamiento de datos, consistente en la limpieza de las observaciones. Pueden aparecer mediciones solo parcialmente recogidas, incompletas, erróneas; pueden aparecer datos atípicos, duplicados, o datos faltantes.

---

[34] Los microdatos permiten la identificación directa de las personas al medir características que permiten identificar al individuo.

[35] Identificadores indirectos son características medidas que por sí solas no permiten identificar al individuo pero que en conjunto con otros indicadores posibilitan la identificación, como lugar, estado civil, y edad.

iii) Necesidad de **aplicar nuevos softwares** que gestionen el gran volumen de datos. Usualmente el volumen de datos que se puede recoger a través de canales digitales es enorme, y las herramientas tradicionales para el uso y manejo de datos no están adaptadas a dicho volumen. El almacenamiento de datos requiere de servidores con garantías de seguridad y el trabajo posterior exige frecuentemente computación en la nube.

iv) Necesidad de investigar y contar con **nuevas técnicas y métodos estadísticos** para el tratamiento de Big Data.

C. Datos de referencia, evaluación y estimación: nuevos métodos para el tratamiento de Big Data

El método tradicional de recogida de datos por una ONE se basa en el diseño muestral, esto es, se genera un universo de población y se elige una muestra en base a un diseño muestral probabilístico. Se recogen las observaciones y se realiza la inferencia a partir del diseño muestral hecho con anterioridad, esto es, con media y varianza controladas ex ante gracias al diseño. No es necesario hacer hipótesis sobre la distribución de la variable de interés. Por el contrario, los datos procedentes de Big Data generalmente no vienen recolectados en un entorno con diseño muestral previo, lo cual puede conducir a sesgos muy importantes en la recogida de las observaciones. Este aspecto debe ser tratado de algún modo. Se exponen a continuación dos experiencias ilustrativas.

Los datos de movilidad, al proceder de un operador de red móvil, recogen la información de los suscriptores de ese operador, los usuarios extranjeros que durante itinerancia utilicen la red del operador, y posiblemente los clientes/suscriptores de un operador móvil virtual (OMV) que, al no tener red propia, se hospeda y descansa en los servicios del operador de red anfitrión. Es una muestra, por tanto, de usuarios lo que dispone el operador, y hay que delimitar si ella es representativa o no del total de usuarios o de ciudadanos sobre los cuales se quiere hacer una inferencia.

Un primer ejercicio consiste en contrastar ciertas dimensiones poblaciones de la muestra del operador a utilizar, con respecto a la población en su conjunto —dato que procede generalmente de estadísticas oficiales publicadas por la ONE del país—. Si se detectaran sesgos en los datos recogidos y se logran identificar, existen métodos estadísticos para corregirlos, parcialmente.

De un modo más general, la recogida de datos de Big Data plantea retos nuevos en cuanto a la representatividad y utilidad de los datos obtenidos. Por ejemplo, en una experiencia conducida por la ONE de los Países Bajos, cuyo objetivo era medir los flujos de tráfico en carreteras principales que conectaban localidades diferentes. Para la medición del flujo, se centraron en camiones de transporte de mercancías, el proyecto desplegó un alto número de sensores de tráfico en puntos concretos, bien de cruce de carreteras o bien en tramos largos. Pero ni todas las rutas punto a punto tenían instalado un sensor inteligente, ni las rutas que sí lo tenían habían sido seleccionadas de un modo aleatorio. Esta ausencia del diseño muestral, crucial en cualquier trabajo de campo de la estadística tradicional, plantea dificultades en cuanto a la inferencia posterior que se vaya a hacer, ya que la "muestra" no ha sido diseñada de modo aleatorio y podría haber sesgos en la recogida de información de los sensores. Dado que no es conocida la probabilidad ex ante de que un tipo de camión concreto sea observado en un tramo concreto en el proceso de recogida de observaciones, esta probabilidad debe ser aproximada. Esto exigió calcular con modelos logísticos la probabilidad de que un camión fuera observado como una función concreta de las características del tramo de carretera o vía donde se observó la ocurrencia del evento y, de modo complementario, la probabilidad de observar al camión en una carretera donde no se habían instalado sensores. Los modelos aproximaron con alta precisión los flujos de tráfico, pero los autores avanzaron que el uso de más conjuntos de datos- como por ejemplo las características propias de los camiones observados- podía dar lugar a mejoras en la estimación.

1. Sesgos en estudios de población a partir de Big Data

Muchas colectas de datos se basan en los usuarios de redes sociales. Estas son muy extendidas hoy entre la sociedad aunque, y esto resulta importante, son más utilizadas por ciertos segmentos de la población que por otros. Cuando la recogida de datos se centra en los usuarios de una red social concreta es fundamental conocer el segmento de población que más frecuentemente utiliza esa red social. Se sabe que el estatus socioeconómico, edad, género o las habilidades con internet afectan claramente a la elección de si utilizar o no una determinada red social. Esto introduce un sesgo en cuanto a que los individuos observados en la colecta no son individuos que aleatoriamente suscribieron la red social de partida; se trata, por el contrario, de individuos que se preseleccionaron en base a atributos sociales varios.

Para identificar si los datos de partida muestran sesgos es fundamental el chequeo cruzado de ellos con fuentes, bien estratificadas y generales a toda la población, como son los estudios y diseños muestrales efectuados por una ONE. Una vez identificados los posibles sesgos, hay salidas metodológicas para tratar los atributos que inducen el sesgo y tratarlos convenientemente en la colecta y en la inferencia posterior.

En un estudio reciente donde el objetivo es la fiabilidad de las aproximaciones con Big Data al número de visitantes en varias localidades en el norte de Alemania, Schmuecker y Reif (2022) comparan tres fuentes distintas de Big Data: i) los datos procedentes de mediciones pasivas de localización obtenidas a través de los smartphones de visitantes, ii) mediciones a través de sistema GPS asignado a los visitantes, y iii) el registro de pernoctaciones en hoteles y centros de alojamiento (registro oficial). Los autores encontraron que cada método tenía sus sesgos propios.

En cuanto al número de visitas con pernoctación, los autores encontraron que los datos procedentes de redes móviles sobreestiman su volumen si se comparaban con la evidencia obtenida de las visitas registradas en hoteles y otros alojamientos (fuentes oficiales). Esta evidencia no es sorprendente dado que se sabe que los registros en hoteles y otros centros de pago subestiman el volumen real de visitantes. Esta es una ganancia clara en los métodos de Big Data comparados con las estadísticas tradicionales. Por otro lado, los autores analizaron la evolución en el tiempo de las distintas mediciones, dado que en el periodo considerado ocurrieron días festivos, fines de semana y vacaciones específicas. La evolución en el tiempo de las pernoctaciones medidas por redes celulares se correlacionó muy de cerca con los datos de registros oficiales. Encontraron una correlación muy alta entre las mediciones por red móvil y el registro tradicional de pernoctaciones, pero una correlación mucho menor con las mediciones por satélite GPS.

Una posible fuente de sesgo en las mediciones por red móvil se debe a la alta proporción posible de visitantes extranjeros, que pueden no ser del todo fiables si hay una alta proporción de ellos que por alguna razón no utilicen las comunicaciones móviles en itinerancia internacional (debido, por ejemplo, al alto precio). En este caso, de todos modos, la presencia de visitantes extranjeros era muy reducida.

Los autores concluyen que las tres fuentes de datos utilizadas mostraron sesgos: la tradicional de registros, la de localización por tecnología móvil, y los datos de satélite (GPS). Proponen más análisis, con series temporales y de sección cruzada más largas, con mediciones repetidas del mismo lugar para poder acometer un análisis de las bondades y limitaciones de cada fuente de datos. Asimismo, proponen utilizar estadística Bayesiana, usando los registros tradicionales como estadístico a priori y evaluando las ganancias en las probabilidades aproximadas ex post añadiendo fuentes alternativas.

2. Ejemplo: obtención de indicadores a partir de Big Data y comparación con indicadores tradicionales

La Unión Internacional de Telecomunicaciones (2020, p. 59) construye un indicador para varios países desde hace años sobre la cobertura de redes 4G en la población, que se define como "población cubierta (en % sobre total) por alguna red 4G/LTE o de estándar superior". Este indicador suele venir recogido

con una frecuencia anual por la ANR de telecomunicaciones en cada país. Aunque los métodos de recogida pueden variar de país a país, muy frecuentemente la autoridad pide a cada operador que en base a una unidad geográfica del territorio pre-elegida (puede ser un nivel de unidad geográfica muy detallada, como la calle, el código censal o el código postal, o puede tratarse de una unidad más agregada, como la comarca) declare la cobertura que sobre la población tiene cada una de sus redes o estándares desplegados. Cada operador provee a la ANR de sus coberturas y la autoridad agrega las coberturas individuales declaradas, lo cual presenta un reto de agregación: el solapamiento de redes. Las redes de distintos operadores suelen cubrir, al menos parcialmente, a las mismas poblaciones. Utilizando la hipótesis de solapamiento máximo, la autoridad elige para cada unidad geográfica la cobertura del operador con mayor cobertura en esa área. Este es el modo "tradicional" de aproximar la cobertura de una red móvil para una tecnología o estándar concreto.

Con Big Data de movilidad se puede llegar a aproximar con mucha mayor precisión la verdadera cobertura, cualquiera sea la unidad geográfica elegida como unidad de referencia. Cada operador dispone de la localización aproximada de todos sus suscriptores (siempre que tengan la terminal encendida) gracias al registro de antena del cual depende el usuario en un momento. Este registro es una aproximación a un polígono de cobertura de la antena, que pueden ser varios kilómetros cuadrados, aunque hay posibilidad de utilizar el gradiente del usuario con respecto a la posición de la antena y calcular con mayor precisión aún la localización. Si el usuario no tuviera el terminal activo, el operador puede repetir el registro de localización vía la antena más cercana en otro momento del día, o de la semana. De este modo puede registrar la localización de todos sus usuarios, incluidos los que se encuentren en itinerancia en el extranjero. La posibilidad de repetir en momentos diferentes la medición de los usuarios dentro de un área concreta posibilita reducir los errores significativamente en la aproximación al área cubierta y sin importar de un modo tan crítico la extensión de la unidad geográfica de referencia.

El error de agregación que se menciona en el párrafo anterior —varios operadores cubriendo una misma área geográfica— se puede corregir en gran medida con localizaciones más precisas y solapando la huella de cobertura de un operador con otro.

De todos modos, es importante saber que pueden existir sesgos de partida en los datos utilizados si proceden de Big Data cuando no se trate de datos censales y la elección de unidades, de puntos de observación o las categorías básicas para la inferencia posterior no están diseñadas de modo aleatorio.

IV. Resultados del cuestionario sobre el uso de nuevas tecnologías digitales para la compilación y procesamiento de estadísticas de comercio de servicios en los países de la Alianza del Pacífico

El Cuestionario fue enviado a los Bancos Centrales y a las ONE de los países integrantes de la Alianza del Pacífico: Chile, Colombia, México y Perú. Se recibieron respuestas por parte de las siguientes instituciones:

- Banco Central de Chile
- Banco de la República de Colombia
- Instituto Nacional de Estadística y Geografía de México (INEGI)
- Banco de México (Banxico)
- Banco Central de Reserva del Perú

Se exponen de modo sintético las respuestas más relevantes agrupadas en bloques temáticos.

A. Uso de Big Data

Mientras una institución declaró no hacer uso de Big Data (el Banco de República de Colombia), los bancos centrales de México y Perú declararon hacer uso de un tipo concreto de fuente de Big Data: web scraping y colecta de datos de transacciones financieras. El Banco Central de Chile es la institución que más intensamente utiliza Big Data. Declaró tener cierta experiencia en la colecta ya sea de datos de imágenes/satélites, webscraping, sobre transferencias electrónicas o incluso extracción de datos de documentos (con reconocimiento de imágenes o textos). Ninguna institución declaró utilizar microdatos de fuentes abiertas procedentes de terceros.

Aun cuando varias instituciones, especialmente bancos centrales, disponen de experiencia con Big Data, declaran de todos modos no producir datos abiertos con procedencia de Big Data. Esto es, si bien acometen ciertas experiencias, no lo hacen de un modo regular.

En cuanto a la pregunta sobre si la organización dispone de un equipo de personas especializado en técnicas de Big Data o hay planes de formación, casi todas las instituciones declararon que tenían planes específicos de formación para sus recursos humanos al respecto. Resalta la atención que demandan las "técnicas de recopilación y tratamientos de datos de alta frecuencia". Dos bancos centrales declararon disponer ya de técnicos especializados en técnicas de Data Science, y uno de ellos disponía de un equipo específico dedicado a Big Data, específicamente a la tarea de "técnicas de recopilación y tratamientos de datos de alta frecuencia".

Todas las instituciones consultadas reconocen que el Big Data puede ser un excelente complemento de las estadísticas tradicionales. Sin embargo, solo una de ellas[36] había alcanzado un acuerdo de colaboración con empresas para acceder a este tipo de fuentes (Banco Central de Chile). Las razones aportadas se centran en el hecho de que con Big data se puede conseguir un complemento a las tareas tradicionales y una de ellas añade el atractivo "Porque podemos acceder a más muestra que con métodos tradicionales" y también porque "permite capturar información en tiempo real y fortalecer la medición de las transacciones internacionales de comercio de servicios que hoy no son captadas por los instrumentos tradicionales". Otra de las ventajas declaradas del Big Data es: "El uso de Big Data puede proveer de mayor granularidad en las estadísticas de servicios, al servir como información para validar los datos provenientes de fuentes tradicionales." Banxico y el Banco Central de Chile declararon que el Big data puede proporcionar datos a los que de otro modo no es posible acceder. El Banco Central de Reserva del Perú afirmó que el Big Data puede ser útil si proporciona microdatos, esto es, información acerca de las transacciones individuales realizadas.

Un aspecto interesante se refiere a las razones por las cuales no se ha llegado a algún acuerdo con empresas propietarias o gestoras de datos masivos. Un banco central informó: "Los agentes propietarios o gestores de los Big Data no muestran interés en colaborar".

En cuanto a la pregunta sobre si existe legislación de protección del consumidor/usuarios y temas relacionados suficiente en el país para cometer trabajos con Big Data, las instituciones consultadas de Chile, Colombia y México declararon que esta existe (tanto una Ley de protección de Datos reciente y una Ley del Sistema Nacional de Información Estadística y Geográfica). Por su parte, el Banco Central de Reserva del Perú declaró que consideraba insuficiente la legislación existente.

B. Uso de IoT o IA

En cuanto al uso de técnicas 4.0, como la IA o el IoT, para la colecta de datos y/o elaboración de encuestas, las instituciones en general declararon no haberlo utilizado aún, ni para reconocimiento de textos, imágenes, ni para generar datos en un sentido amplio. La excepción es el Banco Central de Chile, que utiliza "tecnologías que permiten el análisis de lenguaje escrito (minería de textos)". Esta institución respondió que desarrolló con recursos propios el software necesario explotando algoritmos programados en código abierto.

[36] El Banco Central de Chile ha llegado a acuerdos con empresas, bien para colecta de datos administrativos, o bien disponen de "contratos con empresas del sector privado que permiten hacer uso de información masiva disponible en páginas web".

Las razones dadas por las otras instituciones para no utilizar aún tecnologías 4.0 son que: "La organización no dispone la organización de conocimiento especializado suficiente sobre el uso de la IA", o bien que hoy por hoy "la IA no es útil para la construcción de estadísticas de comercio exterior de servicios".

C. Síntesis de resultados del cuestionario

El Big Data es utilizado aún de un modo moderado por parte de las ONE o Bancos Centrales en los países de la Alianza del Pacífico. Básicamente aquellos que lo hacen es de un modo exploratorio o experimental.

Excepto el Banco Central de Chile, ninguna otra institución ha llegado a acuerdos de colaboración con empresas privadas que colectan y gestionan estos datos. Se menciona el escaso interés por parte de estas empresas en la colaboración.

Todas las instituciones que respondieron al cuestionario declaran la utilidad clara de Big Data como complemento de estadísticas tradicionales, como fuente de contraste de datos regularmente obtenidos y como modo de obtener evidencia de alta frecuencia y de micro- unidades o agentes económicos que de otro modo no se pueden obtener.

Las barreras al uso de Big Data se centran en: i) escaso interés de agentes privados en esta colaboración, ii) falta de formación en Ciencia de los Datos (Data Science), aunque varias instituciones ya disponen de unidades o expertos en esta área, y iii) un marco legislativo que en algunos casos es insuficiente para avanzar con garantías en el uso de Big Data En este sentido, la heterogeneidad de legislaciones existente entre los países de la AP puede limitar un mayor uso de técnicas de Big Data para la compilación y análisis de estadísticas de comercio internacional de servicios.

V. Conclusiones y recomendaciones

Las nuevas tecnologías digitales para la recolección y procesamiento de datos han abierto un mundo de fuentes alternativas de datos, posibilidades de medir fenómenos sociales y económicos que antes eran muy difíciles y costoso de aproximar. El resultado de estos métodos nuevos de colecta son datos no estructurados en grandes volúmenes o Big Data. Estos tienen propiedades muy diferentes de las que se espera en datos recolectados con métodos tradicionales, métodos que se centraban en garantizar la representatividad y la comparabilidad.

Estas nuevas tecnologías de colecta —como Datos Posicionamiento Móvil, Sistema de Identificación Automática, scanned data, web scraping, 5G, IoT— tienen nuevos agentes protagonistas: los que acceden a través de una red o un software a estos datos. Los datos referidos a individuos u objetos que colectan estos agentes suponen una fuente de ventaja fundamental para su negocio y no es fácil que los quieran compartir. Además, los datos referidos a individuos están protegidos por una legislación de protección de datos estricta.

Los bancos centrales y las ONE han visto la utilidad de Big Data, ya sea para conocer fenómenos antes no medibles o para complementar las estadísticas tradicionales, ya que con Big data los gastos de colecta, almacenamiento y tratamiento de los datos se reducen y el detalle, precisión, frecuencia de estos datos hacen que sean un complemento ideal para su tarea.

Los retos para los bancos centrales y las ONE son grandes: desde la formalización de un acuerdo con un agente privado para el uso de estos datos, el gobierno de los mismos, la anonimización, tratamiento y análisis y la comparabilidad con fuentes tradicionales, son problemas aún no resueltos de un modo general.

Estas ventajas están siendo ya exploradas y utilizadas por las ONE de varios países en América Latina y otras partes del mundo. Algunas ONE incluso han creado unidades o institutos de explotación de Big Data, ya sea en colaboración con instituciones privadas o mediante acuerdos de colaboración con otras entidades generadoras de Big Data específicos en proyectos concretos. Seguir los pasos que han ido dando en esta colaboración público-privada, conocer los problemas en el tipo de dato especifico que gestionan, su tratamiento y análisis es la mejor forma de avanzar en la explotación de Big Data.

Ya sea en la colecta o en el análisis posterior, es frecuente combinar fuentes de datos —tradicionales y de recogida con herramientas digitales— en un mismo análisis. Los datos de posicionamiento móvil, los procedentes de sistemas automáticos de identificación o los que se obtienen de transacciones de comercio electrónico pueden traer sesgos importantes, ya que miden solo partes, no el universo, de los individuos o actividad relevantes. Por esta razón, las estadísticas tradicionales que compilan las ONE son importantes para contrastar la representatividad y utilidad de los datos procedentes de nuevas fuentes. Ambos mundos —el tradicional y el nuevo— se complementan.

También las técnicas de análisis cambian. Del diseño de una muestra representativa, con estratos bien definidos ex ante y una secuencia de muestreo bien definida, se ha pasado a la colecta masiva de datos a través de dispositivos digitales a los que se añade en ocasiones Inteligencia Artificial, Aprendizaje de Máquina, Procesamiento de Lenguaje Natural y otras técnicas ya sea en el estadio de la colecta o en el análisis posterior. Se trata nuevos métodos de análisis, muy distintos a los tradicionalmente utilizados por la estadística tradicional.

El mundo de aplicaciones que las nuevas fuentes y métodos de datos abren es enorme. Desde la medición de cargas/descargas de mercancías en puertos, hasta la medición del flujo de turistas o de ocupación de ciertos espacios públicos en tiempo real; desde la medición los flujos de comercio internacional contratados o pagados con medios electrónicos hasta el enrutamiento del comercio por mercancías multi- modal entre países.

De una encuesta breve enviada a Bancos Centrales y ONEs de los países de la AP es claro el interés por estas fuentes alternativas de datos. El Big Data es aún poco utilizado por estas instituciones y su uso, hoy por hoy, es principalmente exploratorio o experimental. Excepto un Banco Central, ninguna otra institución había llegado a acuerdos de colaboración con empresas privadas que colectan y gestionan este tipo de datos.

Las barreras al uso de Big Data que declaran estas instituciones se centran en: i) el escaso interés de los agentes privados en llegar a acuerdos de colaboración, ii) la falta de formación en ciencia de los datos, aunque varias instituciones de la AP ya disponen de unidades o expertos en esta área, y iii) en algunos casos se considera que el marco legislativo actual no es suficiente para avanzar con garantías en el uso de Big Data.

Los desafíos para las instituciones públicas en el uso de Big Data y nuevas tecnologías de colecta son grandes, tanto en la consecución y gobierno de los acuerdos con agentes privados, como en el tratamiento y explotación de los datos y en las técnicas al uso para su explotación. Para superar las dificultades que entraña el uso de nuevas fuentes, la colaboración público- privada y entre instituciones públicas es fundamental y en este aspecto se centran las recomendaciones.

Sobre la base de lo anterior, surgen las siguientes recomendaciones:

- Creación de una unidad especial internacional con carácter jurídico y presupuesto propio —como por ejemplo, un instituto o agencia— por los cuatro países de la AP centrada exclusivamente en la exploración y explotación de Big Data y de fuentes alternativas de datos, con poderes para suscribir acuerdos de colaboración con instituciones privadas o públicas. Los miembros de esta agencia pueden proceder de los bancos centrales, las ONE o ministerios, así como de Universidades, centros de investigación y otras instituciones o fundaciones con interés y conocimiento en las fuentes alternativas y explotación de datos. Se trata de crear una unidad de especialistas en Ciencia de Datos, donde cabe la investigación aplicada y la experimentación con fuentes y métodos alternativos. Un ejemplo de interés puede ser el Centro para Estadísticas de Big Data (CBDS) de los Países Bajos o bien la red colaborativa entre ONEs dentro de la UE(27), ESSnet, dedicada a compartir experiencias de Big Data.

- Homogeneizar las legislaciones diversas de protección y privacidad datos a través de los países de la AP.
- Establecer un sistema de intercambio de experiencias en fuentes alternativas de datos entre las distintas ONEs de la AP, con reuniones periódicas, coordinando esfuerzos en capacitación en Ciencia de Datos y en el desarrollo de acuerdos con otras instituciones privadas. Dado el conocimiento y las experiencias que algunos bancos centrales y ONE de la AP disponen, compartirlo sería beneficios para todos, ahorraría un proceso de aprendizaje que es largo y costoso y ayudaría a superar las dificultades que se presentan.
- A nivel de cada ONE de cada Estado Miembro fijarse algunos objetivos, aunque sea limitados, de exploración de fuentes alternativas que se crea puedan ser un complemento para las estadísticas oficiales. Como ejemplos de posible interés: recogida de precios al detalle (minorista), estadísticas alternativas de turismo o de comercio internacional, la colecta de pagos realizados de comercio electrónico —como ha acometido la ONE de Argentina—, o la medición del alcance de la economía digital en el PIB —como han acometido tanto la ONE de Brasil junto con el centro público de investigación en TIC, Nic.br como la ONE de México.
- Acompañar cada experiencia en Big Data y nuevas técnicas de recolecta digitales con formación específica en Ciencia de Datos y con recursos humanos y de software suficientes.

Bibliografía

Adland, R., H. Jia, y S.P. Strandenes (2017), "Are AIS-based trade volume estimates reliable? The case of crude oil exports", *Maritime Policy & Management*, vol. 44, N° 5.

AEPD (Agencia Española de Protección de Datos) (2016), "Orientaciones y garantías en los procedimientos de anonimización de datos personales", Madrid.

Ahas, R., J. Armoogun, S. Esko y otros (2014), "Feasibility study on the use of mobile positioning data for tourism statistics" [en línea] https://ec.europa.eu/eurostat/documents/747990/6225717/MP-Consolidated-report.pdf.

Arifin, M. D., K. Hamada, N. Hirata, K. Ihara, y Y. Koide (2018), "Development of ship allocation models using marine logistics data and its application to bulk carrier demand forecasting and basic planning support", *Journal of the Japan Society of Naval Architects and Ocean Engineers*, Vol. 27.

Arslanalp, S., M. Marini, y P. Tumbarello, (2019), "Big data on vessel traffic: Nowcasting trade flows in real time", *IMF Working Paper* [en línea] https://www.imf.org/en/Publications/WP/Issues/2019/12/13/Big-Data-on-Vessel-Traffic-Nowcasting-Trade-Flows-in-Real-Time-48837.

Arvis, J. y otros (2010), "Connecting to Compete 2010 : Trade Logistics in the Global Economy--The Logistics Performance Index and Its Indicators", Washington, DC, Banco Mundial.

AGCOM (Autorità per le Garanzie nelle Comunicazioni) (2021), "Esiti delle attività del Gruppo di Lavoro per l'analisi delle tecnologie di comunicazione dei dati nei sistemi di smart metering", Roma.

Axon Partners (2022), "Assessment of the cost of providing mobile telecom services in the EU/EEA countries. A study prepared for the European Commission DG Communications networks, content and technology".

Batty, E. (2018), "Data analytics enables advances AIS applications", en C. Doulkeridis y otros (Eds.), *Mobility analytics for spatio- temporal and social data*, Springer Verlag.

BEREC (Body of European Regulators for Electronic Communications) (2019), "Internet of Things Indicators", BoR (19) 25, BEREC, marzo.

Clarkson Research (2019), Dry Bulk Trade Outlook.

Comisión Económica para Europa, Eurostat y OCDE (2011), *The impact of globalization on national accounts*, Ginebra.

Comisión Europea, "Connecting Europe Facility 2022-27", [en línea] https://digital-strategy.ec.europa.eu/en/policies/cross-border-corridors.

DIRPEN (Dirección de Regulación, Planeación, Estandarización y Normalización) (2014), "Lineamientos para la anonimización de microdatos", Bogotá.

DotEcon Ltd y Axon Partners Group (2018), "Study on Implications of 5G Deployment on Future Business Models", Informe para BEREC (Body of European Regulators for Electronic Communications), No BEREC/ 2017/02/NP3, marzo.

ENSEMBLE, ENabling SafE Multi-Brand pLatooning for Europe [en línea] https://cordis.europa.eu/ project/id/769115.

Ericsson (2021), "Ericsson Mobility Report", Noviembre [en línea] https://www.ericsson.com/es/press-releases/2021/11/ericsson-mobility-report-mobile-data-traffic-increased-almost-300-fold-over-10-years.

Eurostat (2021), "Community Survey on ICT usage and E-commerce in Enterprises", Luxemburgo.

_____ (2013), "Methodological Manual for Tourism Statistics", Luxemburgo.

_____(2011), "Código de buenas prácticas de las estadísticas europeas para los servicios estadísticos nacionales y comunitarios", Comité del Sistema Estadístico Europeo, Luxemburgo.

_____(2008), "Statistical Classification of Economic Activities in the European Community", https://ec.europa.eu/eurostat/documents/3859598/5902521/KS-RA-07-015-EN.PDF.

Fondo Monetario Internacional (2009), "Balance of Payments and International Investment Position Manual", Washington D.C.

GSMA (Asociación Mundial de Operadores Móviles) (2020), "5G y el rango 3,3- 3,8 GHz en América Latina", noviembre.

_____(2017), "An Introduction to Network Slicing" [en línea] https://www.gsma.com/futurenetworks/wp-content/uploads/2017/11/GSMA-An-Introduction-to-Network-Slicing.pdf.

Hargittai, E. (2015), "Is Bigger Always Better? Potential Biases of Big Data Derived from Social Network Sites", *The Annals of the American Academy of Political and Social Science*, vol. 659, Toward Computational Social Science: Big Data in Digital Environments.

Herguera, I., T. Pérez, T. Garín, R. López, A. Valarezo. (2020), "E-commerce by individuals in Spain using panel data 2008- 2016", *Telecommunications Policy*, vol. 44.

Hernández, R. y otros (editores) (2014), *Latin America´s emergence in global services: a new driver of structural change in the region?,* libros CEPAL, Santiago.

IDATE (2021), "Digiworld Yearbook 2021" [en línea] https://en.idate.org/digiworld-yearbook-2021/.

Jia, H., V. Prakash, y T. Smith (2019), "Estimating vessel payloads in bulk shipping using AIS data", *International Journal of Shipping and Transport Logistics*, vol. 11, N° 1.

Kanamoto, K.y otros (2021), "Can maritime big data be applied to shipping industry analysis? Focusing on commodities and vessels sizes of dry bulk carriers", *Maritime Economics & Logistics*, vol. 23.

López González, J. y M. Jouanjean (2017), "Digital trade: Developing a framework for analysis", *OCDE Trade Policy Papers*, N°. 205, Paris.

Marine- Roig, E. y S. Anton Clave (2015), "Tourism analytics with massive user-generated content: A case study of Barcelona", *Journal of Destination Marketing and Management*, vol. 4, N°.3.

Ministerio de Hacienda de Argentina e INDEC (2019), "Estimación de la importación de servicios digitales en la Balanza de Pagos de Argentina", *Documento de Trabajo*, N°. 27, Buenos Aires.

Montserrat, J. y otros (2020), "Envisioning 5G enabled transport", Washington, D.C.; Banco Mundial.

Naciones Unidas (2019), "Handbook on the Use of Mobile Phone Data for Official Statistics", Global Working Group on Big Data for Official Statistics.

_____(2016), "Recommendation for Access to Data from Private Organizations for Official Statistics. The way forward"; UN International Conference on Big Data for Official Statistics, septiembre.

_____(2013), "Global Pulse, Mobile Phone Network Data for Development", octubre.

_____(2011), "International Standard Industrial Classification of all Economic Activities", Departamento de Asuntos Económicos y Sociales, Nueva York.

_____(2010), "International Merchandise Trade Statistics: Concepts and Definitions 2010 (IMTS 2010)", Departamento de Asuntos Económicos y Sociales [en línea] https://www.un.org/en/development/ desa/publications/international-merchandise-trade-statistics-concepts-and-definitions-2010-imts-2010.html.

Naciones Unidas Global Pulse (2013), Mobile Phone Network Data for Development.

Naciones Unidas y otros (2012), "Manual de Estadísticas del Comercio Internacional de Servicios, 2010" [en línea] https://unstats.un.org/unsd/tradeserv/tfsits/msits2010/docs/MSITS%202010%20M86%20(Sp)%20web.pdf.

OCDE (Organización para la Cooperación y Desarrollo Económico) (2022), "Measuring the Internet of Things", Working Party on the Measurement and Analysis of the Digital Economy, Directorate General for Science, Technology and Innovation, DSTI/CDEP/MADE(2021)9/REV, Paris.

_____(2019), "Unpacking comercio electrónico: Business models, trends and policies", Paris, en línea: https://www.oecd.org/digital/ieconomy/unpacking-ecommerce.pdf.

_____(2017a), "An overview of e-commerce in Spain using BBVA Big Data", DSTI/CDEP/MADE(2017)2.

_____(2017b), "Measuring digital trade: towards a conceptual framework", Paris [en línea] https://unctad.org/system/files/non-official-document/dtl_eWeek2017c04-OCDE_en.pdf.

OMC (Organización Mundial del Comercio) (2019), "Joint Statement on Electronic Commerce", Ginebra.

OMC, OCDE, FMI y Naciones Unidas (2023), *Handbook on Measuring Digital Trade* [en línea] https://www.wto.org/english/res_e/booksp_e/digital_trade_2023_e.pdf.

Paul i Agusti, D. (2020), "Mapping tourist hot spots in African cities based on Instagram images", *International Journal of Tourism Research*, vol. 22, N°. 5.

Reza Alaei, A., S. Becken, y B. Stantic, (2017), "Sentiment analysis in tourism: capitalizing on Big Data", *Journal of Travel Research*, vol. 58, N°. 2.

Rodrigues de Lima, R. y otros (2022), "An Empirical Comparison of Portuguese and Multilingual BERT Models for Auto-Classification of NCM Codes in International Trade", *Big Data and Cognitive Computing*, 6, 8 https://doi.org/10.3390/bdcc6010008.

Salas-Olmedo, M., y otros (2018), "Tourist digital footprint in cities. Comparing Big Data sources", *Tourism Management*, vol. 66.

Saluver, E., y otros (2020), "Methodological framework for producing national tourism statistics from mobile positioning data", *Annals of Tourism Research*, vol. 81.

Schmuecker, D. y J.Reif (2022), "Measuring tourism with big data? Empirical insights from comparing passive GPS data with passive mobile data", *Annals of Tourism Research and Empirical Insights*, No. 3.

Shibasaki, R.; K. Kanamoto, T.Suzuki (2020), "Estimating global pattern of LNG supply chain: a port-based approach by vessel movement database", *Maritime Policy and Management*, vol. 47.

Tu, E., G. Zhang, L. y otros (2017), "Exploiting AIS data for intelligent maritime navigation: A comprehensive survey from data to methodology", *IEEE Transactions on Intelligent Transportation Systems*, vol. 99.

Unión Europea, H2020 EfficienSea2 project, [en línea] https://cordis.europa.eu/project/id/636329/es.

_____(2018), "Código de Comunicaciones Electrónicas", Directiva UE/2018/1972, 11 de diciembre.

_____(2016), Reglamento UE/2016/679 del Parlamento Europeo y del Consejo relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos.

_____(2002), Directiva (2002/58) relativa al tratamiento de datos personales y la protección de la intimidad en el sector de comunicaciones electrónicas.

UIT (Unión Internacional de Telecomunicaciones) Recomendación ITU-R M.1371-5, de 2014, [en línea] http://www.itu.int/rec/R-REC-M.1371-5-201402-I.

_____(2020), "Handbook for the collection of administrative data on telecommunications/ICT", Ginebra [en línea] https://www.itu.int/en/ITU-D/Statistics/Pages/publications/handbook.aspx.

_____(2017), "Big Data for measuring the information society. Methodology", Ginebra.

_____(2016), Recomendación ITU-R M.2092-0, Ginebra.

UN Global Working Group on Big Data for Official Statistics (2019), presentación [en línea] https://slideplayer.com/slide/16209110/.

UNCTAD (2016), "In Search of Cross-Border E-Commerce Trade Data", *UNCTAD Technical Notes on ICT for Development*, N° 6. Ginebra [en línea] https://unctad.org/system/files/official-document/tn_unctad_ict4d06_en.pdf.

Yang, Y., Fan, Y., Jiang, L., Liu, X. (2022), "Search query and tourism forecasting during the pandemic. When and where can digital footprints be helpful as predictors?", *Annals of Tourism Research*, vol. 93.

Anexos

Anexo 1

A. CEPAL para los países de la Alianza del Pacífico

Contexto: Desde 2020, la Comisión Económica para América Latina y el Caribe (CEPAL) está apoyando a la Alianza del Pacífico (AP) en la mejora de las estadísticas de comercio internacional de servicios (ECIS) en los cuatro países miembros. En 2022, este apoyo se concentra en dos áreas: la compilación de las ECIS sobre servicios profesionales y el potencial de las nuevas tecnologías en la compilación de las ECIS. Para cada uno de los dos temas, la CEPAL preparará un informe con un diagnóstico sobre la situación actual en los países de la AP, una revisión de buenas prácticas dentro y fuera de América Latina y recomendaciones para mejoras futuras.

A quién va dirigido: Técnicos de instituciones a cargo de la colecta o procesamiento de datos para las ECIS, principalmente los bancos centrales, las oficinas nacionales de estadísticas, ministerios u otras instituciones en el estado y/o el sector privado.

Objetivo: El presente cuestionario busca recabar información sobre el uso de las nuevas tecnologías (como Big Data, Geographic Information System -GIS-, Inteligencia Artificial, Web Scraping y otras técnicas digitales) para la recolección y procesamiento de datos para compilar las ECIS en los países miembros de la AP.

B. Contestar esta encuesta no debería llevar más de 10 minutos

1. Sobre la organización

a) Tipo de organización que representa:

i) Oficina Estadística Nacional

ii) Banco Central

iii) Ministerio con producción estadística

iv) Otra institución:

b) ¿Pone su organización a disposición del público conjuntos de datos? (por ejemplo, pone a disposición del público para descarga datos elaborados, como descriptivos ya agregados, o pone a disposición del público micro -datos ya anonimizados, tal y como se recogen en encuestas)

Sí/No:

i) La organización ofrece Cuadros ya elaborados con los principales indicadores estadísticos, con posibilidad de ver, graficar y descargar en formatos abiertos

ii) Ofrece datos en formatos abiertos con posibilidad de realizar desgloses más allá de los que se ofrecen en las Cuadros y gráficos disponibles

iii) Una vez anonimizados, ofrece la posibilidad directa o bajo petición de descarga de datos con todo el desglose posible tal y como se colectaron

2. Uso de Big Data

Big Data: conjuntos de datos voluminosos, complejos, estructurados o no, procedentes de fuentes tradicionales. Son datos generados por tecnologías modernas, como direcciones IP, web logs, datos de uso del espectro de radiofrecuencias, búsquedas en internet, procedentes de sensores, dispositivos GPS y otros.

a) ¿Tiene su organización experiencia con Big Data ¿(elegir varias si relevante):

Sí/No:

i) datos de imágenes de satélite o geodata (GIS)

ii) datos procedentes de red u operador móvil

iii) datos de usuarios (procedentes de redes sociales)

iv) datos de crowdsourcing (obtenidos de aplicaciones móviles o de OpenStreetMap)

v) datos procedentes de webscraping

vi) datos procedentes de scanner data/de sensores

vii) datos procedentes del sector privado (datos financieros, por ejemplo)

viii) Datos/información extraída de documentos

ix) Datos procedentes de elementos o de terminales asociados a cosas con conexión SIM/eSIM geolocalizados

x) No hemos utilizado aún Big Data

b) ¿Utiliza su organización alguna fuente de datos abierta obtenida de terceros como complemento para las estadísticas oficiales que publica en alguno de los ámbitos siguientes? (posible respuesta múltiple)

Sí/No:

i) A nivel geográfico, como por ejemplo Meta

ii) A nivel de localización, como por ejemplo OpenCellID

iii) Velocidad de conexiones de banda ancha, como ookla

iv) A nivel de población, como WorldPop o Meta

v) Sobre dimensiones socioeconómicas, como por ejemplo Meta

vi) Para la medición de cambios de los precios de productos

vii) Sobre volumen de comercio electrónico, como por ejemplo de intermediarios financieros, gestores de tarjetas de crédito/débito

viii) Sobre publicidad

ix) Sobre transacciones financieras o de pago

x) Sobre aspectos meteorológicos o espaciales

a) ¿Genera su organización Big Data, esto es, dispone su organización de otras fuentes digitales de datos —como puede ser datos de posicionamiento móvil, sistemas de identificación automática, sensores) tales que las pone a disposición del público al menos parcialmente?

Sí/No:

i) A nivel geográfico

ii) A nivel de localización

iii) A nivel de población

iv) Sobre dimensiones socioeconómicas

v) Para la medición de cambios de los precios de productos

vi) Sobre volumen de comercio electrónico

vii) Sobre transacciones financieras o de pago

viii) Sobre aspectos meteorológicos o espaciales

ix) Otro ámbito (especificar)

b) ¿Ha sustituido su organización parte de su producción regular de estadísticas por datos obtenidos mediante Big Data (por ejemplo, pasar de encuestas o recogidas censales a recogidas de datos provenientes de otras fuentes)?

Sí/No:

i) Por ejemplo, para la recogida de precios de productos y la medición del índice de precios periódico

ii) Por ejemplo, para los censos de población

iii) Por ejemplo, para los censos o registros de empresas

iv) Tenemos en marcha proyectos piloto

v) Otros

c) ¿Cuenta su organización con personal especializado en colecta, tratamiento y explotación de Big Data?

i) Tenemos un área especializada en ello

ii) Tenemos un número suficiente de expertos en ello

iii) No tenemos aún conocimiento ni expertos suficientes

d) ¿Tiene su organización planes de formación para sus funcionarios en técnicas de colecta, explotación y tratamiento de Big Data? (colecta en cuanto a nuevos métodos de recogida o fuentes de datos; explotación en cuanto a retos con el estadio de la publicación; tratamiento en cuanto a desafíos de sesgos, nuevos métodos a utilizar)

Sí/No:

e) ¿Encuentra fácil o costoso llegar a acuerdos de colaboración con agentes o empresas que tienen acceso a Big Data?

Sí/No:

Razones (posible respuesta múltiple):

i) Limitaciones derivadas de la legislación vigente sobre la producción estadística oficial impone muchas barreras y dificultades

ii) Los agentes propietarios o gestores de los Big Data no muestran interés en colaborar

iii) El costo monetario que supone monetario adquirir Big Data es muy alto

iv) El trabajo estadístico que supone compatibilizar los datos procedentes de fuentes de Big Data con los datos obtenidos mediante métodos de estadísticas oficiales es muy largo y costoso (por ejemplo, debido a los sesgos, diferencias en su calidad o formato)

v) La gobernanza, gestión y distribución de responsabilidades es costosa de diseñar y de implementar

f) ¿Cree que existe en su país legislación suficiente y detallada para la protección de los datos personales y la privacidad de los mismos?

Sí/No:

Si alguna sugerencia, por favor exponer sucintamente...

g) ¿Cree que el uso de Big Data puede ser un buen complemento o incluso sustituto de las fuentes de estadísticas tradicionales?

Sí/No:

Explicación sucinta:

h) ¿Considera su organización que el uso de Big Data puede ser útil para la mejor medición del comercio y de los flujos de pagos a nivel internacional?

Sí/No:

3. Sobre uso de Internet de las Cosas

a) ¿Utiliza su organización elementos o terminales interconectados (smart devices) conel fin de obtener mediciones específicas y recurrentes (tráfico, movimientos de personas, consumos de insumos —electricidad, agua—...) que sean monitorizados remotamente?

Sí /No:

Otros:

Exponer experiencia brevemente

4. Inteligencia Artificial

Inteligencia Artificial (AI) se entiende como el uso de sistemas que utilizan tecnologías del tipo: minería de textos o imágenes, reconocimiento de voz, imágenes por computadora, elaboración de lenguaje natural, aprendizaje de máquinas (machine learning), con el fin de recolectar datos de alguna actividad con el fin de predecir, recomendar o decidir con diversos niveles de autonomía la mejor acción para la consecución de un objetivo determinado (Eurostat, 2021).

a) ¿Utiliza su organización Inteligencia Artificial (AI) por ejemplo en sus actividades de recopilación o tratamiento de datos?

Sí/No:

b) ¿Usa su organización alguna de las siguientes tecnologías de IA? (posible respuesta múltiple)

i) Tecnologías que permiten el análisis de lenguaje escrito (minería de textos)

ii) Tecnologías que permiten trascribir lenguaje oral en medio que sea posible sea leído por una máquina o software?

iii) Tecnologías que generen lenguaje oral o escrito por si mismas?

iv) Tecnologías capaces de identificar personas u objetos basadas en la imagen (reconocimiento de imagen, procesamiento de imagen)

v) Aprendizaje de máquinas (machine learning) para el análisis de datos?

vi) Tecnologías que posibiliten el movimiento físico o desplazamiento de máquinas o elementos en base a decisiones autónomas basadas en la observación del entorno? (robots autónomos, vehículos autónomos, drones autónomos)

c) ¿Utiliza su organización software de IA para conducir cuestionarios/muestreos con personas telemáticamente/en línea?

Sí/No:

d) Si su organización utiliza software o sistemas de IA, ¿cómo los adquirió?

Se desarrolló dentro de la organización

Sistemas o software de código abierto que fueron modificados y adaptados para fines concretos

Adquisición de software o sistemas disponibles comercialmente

Se contrató a una empresa especialista externa para el desarrollo del software o sistema de IA utilizado

e) Si su organización NO utilizó sistemas o software de IA, ¿cuáles son las razones? (elegir todas las alternativas relevantes)

i) El costo es muy alto

ii) La organización no dispone la organización de conocimiento especializado suficiente sobre el uso de la IA

iii) Incompatibilidad con la filosofía y métodos de trabajo de la organización

iv) Incompatibilidad de sistemas instalados

v) Problemas con la legislación sobre violación de protección de datos y privacidad

vi) Incertidumbre en cuanto a las consecuencias legales de utilizar IA (responsabilidad civil, alcance de datos a recolectar, tratar, procesar...)

vii) Razones éticas

viii) La IA no es útil actualmente para la organización

Anexo 2
Glosario de términos

Algoritmos de cifrado: un algoritmo de cifrado permite realizar operaciones con datos cifrados de tal manera que el resultado de las operaciones es el mismo que si las operaciones se hubieran realizado con los datos sin cifrar (propiedad homomórfica). Este esquema de cifrado abre la posibilidad del tratamiento de datos personales anonimizados garantizando la privacidad del tratamiento y que los resultados de los tratamientos van a ser accesibles únicamente al poseedor de la clave de descifrado.

Algoritmo de hash: algoritmo matemático que permite generar la huella digital de un documento digital (archivo, imagen, cadena alfanumérica, etc.). Un algoritmo de hash es un mecanismo que, aplicado a un dato concreto, genera una clave única que puede utilizarse para representar un dato y posibilita que partiendo de un mismo dato podamos generar siempre la misma huella digital pero no permite el proceso inverso, esto es: partiendo de una determinada huella digital no se puede obtener el dato original.

Aprendizaje profundo (Deep learning): es un tipo de Machine Learning por el cual se entrena a una computadora para que realice tareas replicando el modo de hacer humano, como puede ser identificar imágenes, rostros o reconocer el habla. El Aprendizaje Profundo no parte de ejecutar ecuaciones predeterminadas con unos datos o input concreto, trata en cambio de dar solo unos parámetros básicos sobre los datos y deja que la computadora aprenda por cuenta propia a reconocer patrones o relaciones mediante el uso de muchas capas de procesamiento. Requiere de gran potencia de cálculo tanto por la gran cantidad de datos que se requiere para entrenar a una máquina, como por las distintas capas de algoritmos que se utilizan simultáneamente en el proceso de reconocimiento o tratamiento de la información.

Big Data: según IBM, "la analítica de big data es el uso de técnicas analíticas avanzadas contra conjuntos de datos muy grandes y diversos que incluyen datos estructurados, semiestructurados y no estructurados, de diferentes fuentes y en diferentes tamaños". Estas técnicas se utilizan para identificar correlaciones, patrones ocultos y relaciones entre los factores.

Blockchain: es un registro digital descentralizado con sello de tiempo de las transacciones en el que éstas se almacenan de forma permanente y casi inalterable utilizando diversas técnicas criptográficas. Se trata de una lista de registros en continuo crecimiento, llamados bloques, que se encadenan entre sí mediante herramientas criptográficas.

Capas de anonimización: es un proceso de anonimización secuencial. Una vez se han anonimizado todos los datos que puedan servir para re-identificar a las personas, el agente o la parte receptora de los datos puede, a su vez, acometer una segunda anonimización para evitar que pudiera producirse la re-identificación. El receptor de los datos asegura que sus procesos utilizan sus propios recursos de anonimización.

Computación cuántica: según el Instituto de Computación Cuántica de la Universidad de Waterloo, "la computación cuántica consiste esencialmente en aprovechar y explotar las sorprendentes leyes de la mecánica cuántica para procesar información. Un ordenador tradicional utiliza largas cadenas de "bits", que codifican un cero o un uno. Un ordenador cuántico, en cambio, utiliza bits cuánticos, o qubits", que son 0 y 1 al mismo tiempo. Los ordenadores cuánticos tienen el potencial de procesar exponencialmente más datos que los ordenadores clásicos. La computación cuántica puede considerarse tanto una tecnología como una infraestructura, porque requiere de hardware cuántico.

Distributed ledger technology (DLT): La tecnología de libro mayor distribuido (DLT) es un sistema digital de registro de transacciones de activos en el que las transacciones y sus detalles se registran en múltiples lugares al mismo tiempo. A diferencia de las bases de datos tradicionales, los libros de contabilidad distribuidos no tienen un almacén central de datos ni funciones de administración. Se generan bases de datos no centralizadas que pueden ser gestionadas por varios participantes.

Inteligencia Artificial, IA: se refiere a los sistemas que cambian el comportamiento sin ser programados explícitamente en base a los datos que se observan, se recogen y luego se analizan. La inteligencia artificial es un conjunto de técnicas que posibilitan generar algoritmos de modo dinámico que aprenden del contexto. Los algoritmos generan reglas de decisión que cambian. Es un término amplio que incluye el aprendizaje automático, el aprendizaje profundo, la visión por ordenador y el procesamiento del lenguaje natural.

Interfaz de programación de aplicaciones (Application programming interface, API): es un intermediario de software que permite que dos aplicaciones interactúen entre sí. Las API definen los tipos de llamadas o peticiones que pueden hacerse de una aplicación a otra, cómo hacerlas, los formatos de datos que deben utilizarse y las convenciones que deben seguirse. Son esenciales pues facilitan la comunicación entre aplicaciones, permitiendo utilizar los distintos servicios que diferentes aplicaciones posibilitan y explotar datos de distintas fuentes.

IoT: según la OCDE (2022) el IoT "…. Incluye todo tipo de terminal y objetos cuyo estado puede ser alterado via internet, con la participación o no de una persona". Se puede entender como una red de sensores y dispositivos inteligentes conectados a internet que pueden enviar y recibir datos y que suelen encontrarse en vehículos, edificios y artículos con electrónica incorporada. La IoT permite el seguimiento de los productos a lo largo de las cadenas de suministro y puede reducir el costo del comercio mundial al aumentar la eficiencia del envío y el transporte.

Procesamiento del lenguaje natural (NLP, por sus siglas en inglés): rama de la IA que se ocupa de la interacción entre los ordenadores y los seres humanos utilizando el lenguaje natural, esto es, el lenguaje humano evolucionado naturalmente, en contraposición al lenguaje informático binario. Es un área de la Inteligencia Artificial que trata de conseguir que un ordenador sea capaz de entender palabras escritas o enunciadas tal y como las entiende una persona y capaz de comunicarse por su cuenta. Gracias a NLP un ordenador puede traducir textos u oraciones de un lenguaje a otro o responder a mensajes emitidos por personas.

Redes neuronales artificiales (Artificial Neural Networks): Estructuras de redes virtuales que están diseñadas, en su topología y comportamiento, para asemejarse a las células neuronales y a las conexiones de éstas dentro del cerebro biológico. Por ejemplo, se utilizan redes neuronales artificiales con aprendizaje supervisado para el reconocimiento de imágenes.

Reducción de datos: según la AEPD, es el proceso por el cual se reduce el número de datos originales sin alterar los mismos, disminuyen el nivel de detalle de los datos originales evitando la presencia datos únicos o atípicos sin relevancia para el resultado final. Puede hacerse eliminando variables sensibles, reduciendo el número de registros, recodificando datos originales o agregando datos a nuevas categorías.

Scanner data: consiste en utilizar la información registrada por un comercio minorista en la caja registradora donde se utiliza el scanner para el registro de los productos comprados y hacer la facturación. Los campos recogidos son: número de unidades vendidas, el pago total por cada uno de los productos comercializados, el tipo de producto comprado, posibles descuentos y los impuestos relevantes a cada producto. Para la clasificación del producto cada tienda o empresa puede utilizar clasificaciones diferentes. El scanner data permite identificar productos con alto grado de detalle (variedades, tipos, tamaños), ofertas o campañas de descuentos específicos o empaquetamiento de bienes en una misma oferta comercial.

Web scraping: consiste en la obtención automatizada de información de algunas páginas web de empresas que prestan servicios a través de Internet. Se trata a través de software de extraer cierta información de los sitios web. Una aplicación frecuente es para recoger precios de productos en oferta de distintos sitios web y poder después compararlos.

En el presente estudio, se examina el uso de nuevas tecnologías digitales para la recolección de datos sobre para el comercio internacional de servicios. El uso de macrodatos (*big data*) y de nuevos métodos de recopilación —como las fuentes del Sistema de Posicionamiento Global (GPS), los datos de posicionamiento móvil, la extracción automática de contenido web (*web scraping*), el sistema de identificación automática o los datos escaneados— ha multiplicado el volumen de datos disponibles para medir la actividad económica y el comercio. Debido a este mayor volumen de datos, los bancos centrales y las oficinas nacionales de estadística enfrentan el requerimiento de introducir cambios importantes en su gestión. En concreto, se requiere lograr acuerdos de colaboración con los agentes privados que disponen de estos datos, así como contar con nuevos métodos de análisis. En este estudio, se revisan también los resultados de una encuesta aplicada a los bancos centrales y oficinas nacionales de estadística de los países miembros de la Alianza del Pacífico. Las respuestas recibidas indican que los macrodatos son aún poco utilizados por dichas instituciones, y que su uso es principalmente exploratorio o experimental.

CEPAL
Comisión Económica para América Latina y el Caribe (CEPAL)
Economic Commission for Latin America and the Caribbean (ECLAC)
www.cepal.org

LC.TS.2023/170