

**UN ECLAC en colaboración con el Departamento de Estadística del FMI y la
UNCTAD**

Noviembre 11, 2022

**Seminario virtual sobre “Medición de la economía y comercio digital
en Latinoamérica y el Caribe”**

**“El uso del Big Data para medir el comercio de
servicios”**

Iñigo Herguera
Consultor para CEPAL

Indice

1. Nuevos métodos de recogida y de explotación de datos: técnicas y algunas aplicaciones de interés

1.1. Datos de Posicionamiento Móvil (MPD)

1.2. AIS- Sistema de Identificación Automática de barcos

1.3. 5G y el Internet- de las- Cosas (IoT)

1.4. Medios de pago y medición del comercio electrónico

2. Problemas en la recogida, tratamiento y acuerdos de colaboración público- privados

3. Pequeño cuestionario a Oficinas Nacionales de Estadística y otros agentes públicos

La recolección de datos a través de herramientas digitales en grandes volúmenes, Big Data,

Ventajas:

- recogidas y procesamiento más rápido y menos costoso
- repetir “muestras” es menos costoso, con mayor frecuencia
- permite recogida de micro- datos, con muchos mayores desgloses/ desagregaciones
- muchas veces no son “muestras”, son registros (aunque puede haber problemas)
- permiten aplicación de Inteligencia Artificial, Big Data... nuevos modos de manejar y explotar los datos (para predicción, simulación,)

Y también problemas:

- nuevos métodos para el tratamiento y explotación de los datos
- proteger la anonimidad de los ciudadanos / usuarios
- acceso a los datos
- comparabilidad don datos de fuentes oficiales

Nuevos métodos de recogida y de tratamiento/ explotación

- Datos de posicionamiento móvil (MPD)
- las redes 5G y el Internet de las Cosas,
- los Sistemas de Identificación Automática (AIS)
- comercio electrónico como indicador de los flujos de comercio

“datos” necesitan de una red de telecomunicaciones

Aplicaciones/ experiencias de comercio internacional en servicios

- Turismo
- Logística, gestión de puertos, cadenas de transporte
- Comercio electrónico

1. Nuevos métodos de recogida y de explotación de datos: técnicas y algunas aplicaciones de interés

1.1. Datos de Posicionamiento Móvil (MPD)

1.2. AIS- Sistema de Identificación Automática de barcos

1.3. 5G y el Internet- de las- Cosas (IoT)

1.4. Medios de pago y medición del comercio electrónico

1.1. Datos de Posicionamiento Móvil (MPD)

- Son 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.

Fase 1: la extracción de los micro-datos que suelen ser confidenciales, privados y/o personales (Call Detailed Records, CDR)

Fase 2: los datos en esta fase ya están agregados o procesados de tal modo que ya no muestran datos privados (datos disociados)- ya no es posible identificar individuos concretos a partir de los campos ofrecidos

Fase 3: es la fase del macro-dato. Son datos agregados como indicadores que ya no tienen carácter ni privado ni confidencial

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

1. 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.

2. 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 (localización por satélite), datos de itinerancia internacional en redes móviles, o datos recogidos a través de una red de área local (bluetooth, wi- Fi)

3. 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- e

Tabla 7: ventajas e inconvenientes del uso de MPD para la elaboración de estadísticas de turismo. *Fuente:* Eurostat 2014.

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” transfronteriza que identifique redes de turismo a través de varias fronteras, si existe coordinación en la colecta de datos entre países	

Casos de interés:

MPD para flujos turismo dentro/ fuera	http://ec.europa.eu/eurostat/documents/747990/6225717/MP-Consolidated-report.pdf	Estonia	2014
Eventos turísticos y desportivos 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	https://www.researchgate.net/publication/280599074_Tourism_analytics_with_massive_user-generated_content_A_case_study_of_Barcelona	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://onlinelibrary.wiley.com/doi/abs/10.1002/jtr.2360	Africa	2020
Using tourist footprints: comparison of different big data sources	https://arxiv.org/abs/1705.07951	España	2018

1.2. AIS- Sistema de Identificación Automática de barcos

- AIS provee datos, ya sea vía frecuencia de radio (VHF) o via 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, nombre) y su movimiento.

Aplicaciones:

- aproximación al volumen de comercio de mercancías marítimo gracias a la explotación de los datos procedentes de AIS: como el volumen de comercio de petróleo (Adland, 2017) o de gas licuado (Shibashaki, 2020) o para el análisis de tráfico en tiempo real (Arslanalp, 2019), la estimación de carga en buques de gran tamaño (Jia, 2019) o la planificación de tráfico en puertos (Arifin, 2018)

Caso de interés: aproximación del flujo global de comercio marítimo de hierro, fertilizantes, grano y carbón en los distintos puertos de Japón para 2016, explotando distintas fuentes de datos (Kanamoto et. al. 2021)

Combinación de fuentes de información

AIS- datos localización, movimiento, calado



datos de geolocalización de los barcos se puede utilizar el **Geohash**: técnica de geo- codificación con cadenas alfanuméricas de distintas longitudes y proporciona distintos grados de precisión en la geolocalización

AXS: atraque, descarga de mercancías (tipo), fecha, puerto

Datos oficiales (1): Entrada/salida puerto

Datos oficiales (2): estadísticas flujos mercancías Instituto Estadística, UNComTrade

Problemas encontrados:

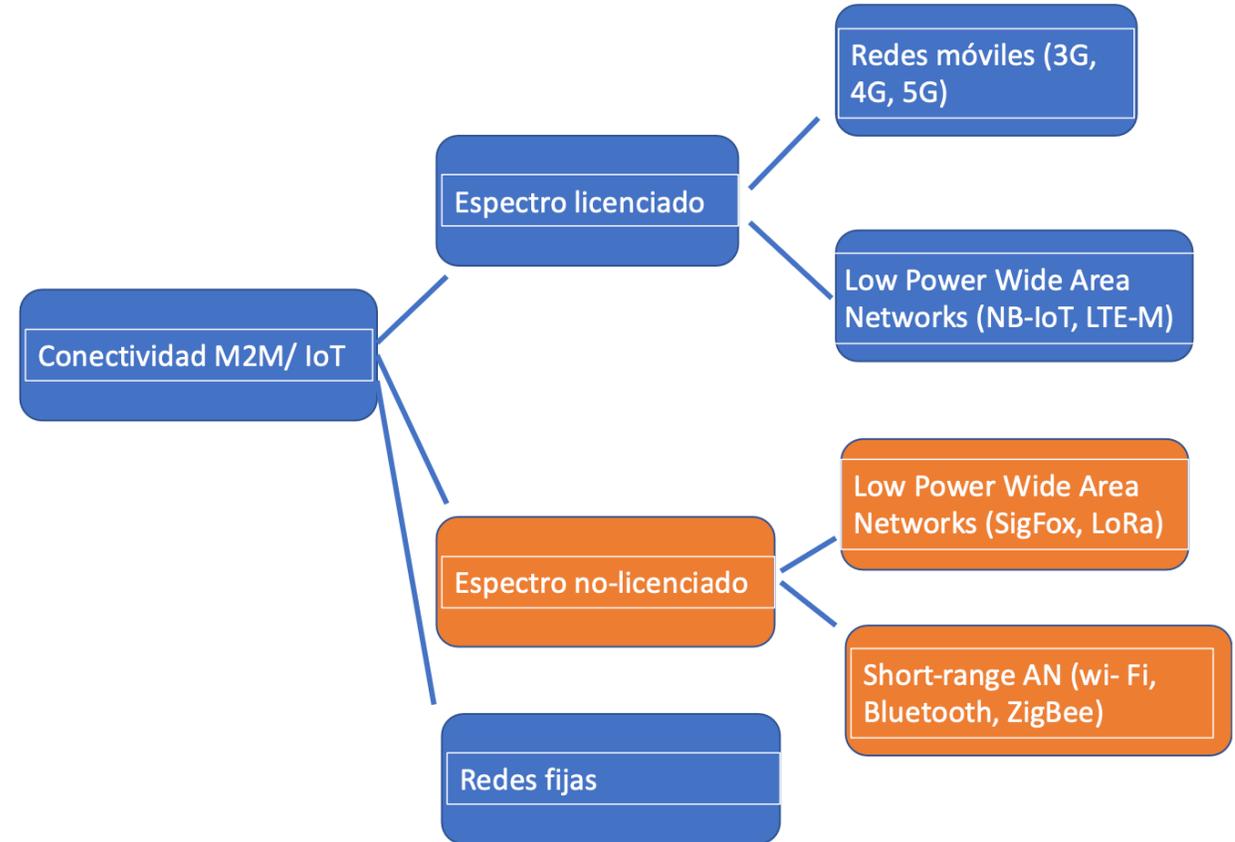
- cada fuente de datos utiliza una clasificación de mercancías/ bienes diferente:
 - (1) autoridades nacionales (Ministerio, Puertos),
 - (2) fuentes privadas (AXS, Clarkson) y
 - (3) fuentes públicas (UN ComTrade)
- necesidad de machear (hacer corresponder) distintas informaciones a través de fuentes diferentes:
 - * problemas de formatos diferentes,
 - * identificación de sujetos a medir con escalas o métricas distintas

1.3. 5G y el Internet- de las- Cosas (IoT)

5G (3GPP) o estándar IMT-2020 (UIT) puede proporcionar velocidades hasta 100 veces mayores que el estándar 4G/LTE, y con niveles de latencia mucho menores

El mayor ancho de banda posibilita grandes flujos de datos entre millones innumerables objetos-Internet de las Cosas (IoT)-conectados de continuo e interaccionando entre sí.

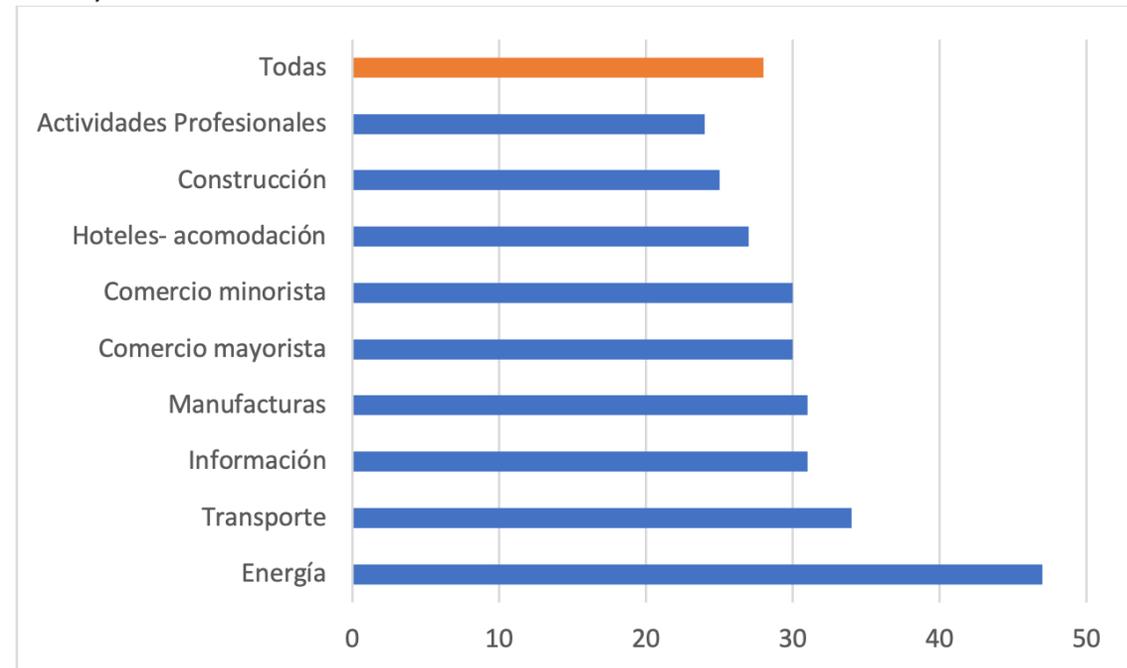
Figura 2: redes utilizadas para Internet de las Cosas. Fuente: BEREC, 2019



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

- (1) Banda ancha móvil mejorada (eMBB),
- (2) Conexiones de máquina masivas (mMTC)
- (3) Comunicaciones ultra-fiables y de baja latencia(ULLC)

Figura 1: Proporción de empresas en países seleccionados de OECD que utilizan IoT, 2021. *Fuente:* OECD, 2022



Casos de interés:

1. La logística conectada

- Monitorizar y medir todos los elementos de una cadena de suministro implica inversiones, por un lado, y también grandes mejoras en eficiencia y ahorro de costes. Esto va a ser posible si se dan tres condiciones:
 - (1) **conectividad** global o **universal** hasta niveles muy locales- de “punto de entrega o distribución final”,
 - (2) potencia de cálculo y **sensores integrados** entre sí , y
 - (3) **uso de IA** (inteligencia artificial) para conocer lo que ocurre entre los sensores, terminales y máquinas conectados entre sí.
- es esta integración de redes, datos y capacidad de análisis lo que conduce a cambios profundos en la integración de servicios hasta ahora segmentados, provistos por agentes diferentes

Casos de interés:

2. Gestión de los puertos con 5G

- En Italia el puerto de Livorno en colaboración con Ericsson ha desplegado una red de 5G, 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 más allá, hacia las principales ciudades y polos industriales de la Toscana
- El despliegue del sistema logístico inteligente supone la instalación de cámaras, sensores y objetos todos conectados a una red 5G de alta fiabilidad
- A este sistema se le añaden tecnologías como IA, Realidad Aumentada y de este modo se posibilita realizar cargas/ descargas de mercancías con robots, gestionar el transporte multi- modal, controlar calidad o estado de mercancías....

- Ericsson estima que se ha conseguido una reducción de 2,5 millones de Euros solo en el tiempo de atraque de los barcos y un 25% más de productividad en el uso de las grúas en muelles gracias al control y gestión en remoto de las mismas
- Menor uso de combustibles fósiles, de energía y menores tiempos de almacenaje
- Se crea un “gemelo digital” (*digital twin*): una réplica del puerto de Livorno en 3D que simula todas las actividades sometidas a control y gestión de interés.
- Experiencia en también en otros puertos, como el de Singapore y Shanghai, Málaga, Barcelona.....

Casos de interés:

3. Pruebas de “pelotón de camiones” en Japón

- el Ministerio de Comunicaciones de Japón ha promovido la realización de tests del 5G, como por ejemplo la experiencia de *truck platooning*
- Similares pruebas en la UE

4. Corredores trans- fronterizos en la UE

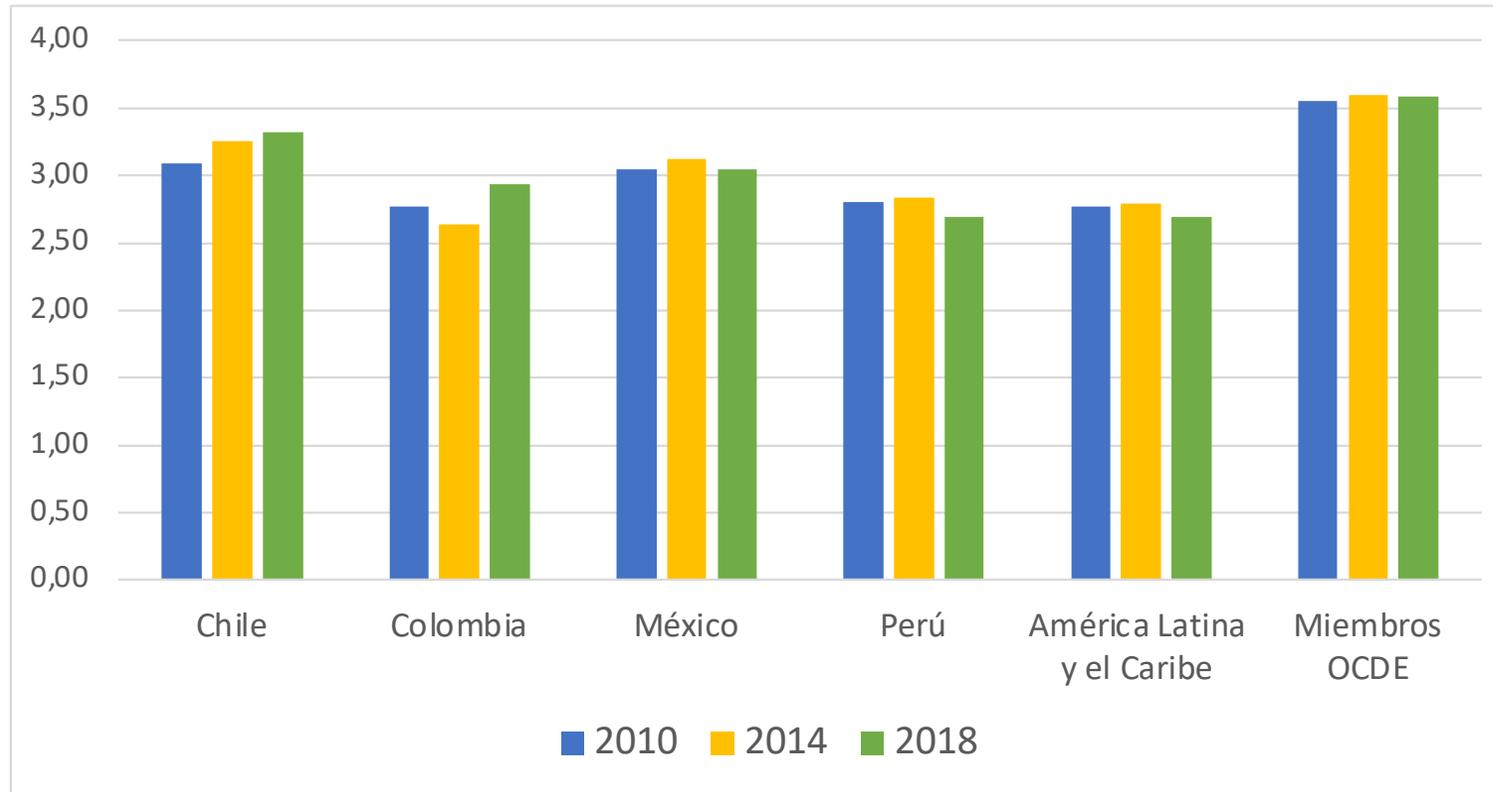
- para facilitar, gestionar y optimizar los tráficos de personas y de mercancías por carretera y de promover la conducción autónoma
- colaboración público- privada

Tabla 6: experiencias al uso de 5G en puertos, rutas trans-fronterizas y tráfico

Tema	Link	País	Año	Partners
Comunicación directa entre vehículos y entre vehículos e infraestructura vial	https://www.t-mobile.cz/en/press-releases/press-news-archive/t-mobile-is-testing-cv2x-data-technology-for-vehicles-and-infrastructure.html	Rep 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 Zeebrugge	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 trans-fronterizos en la UE	https://digital-strategy.ec.europa.eu/en/policies/cross-border-corridors	Unión Europea	2020	

Interés para países integrantes en la Alianza del Pacífico

Indicadores de Desarrollo logístico, Banco Mundial



Grados:

1 – bajo

5- alto rendimiento

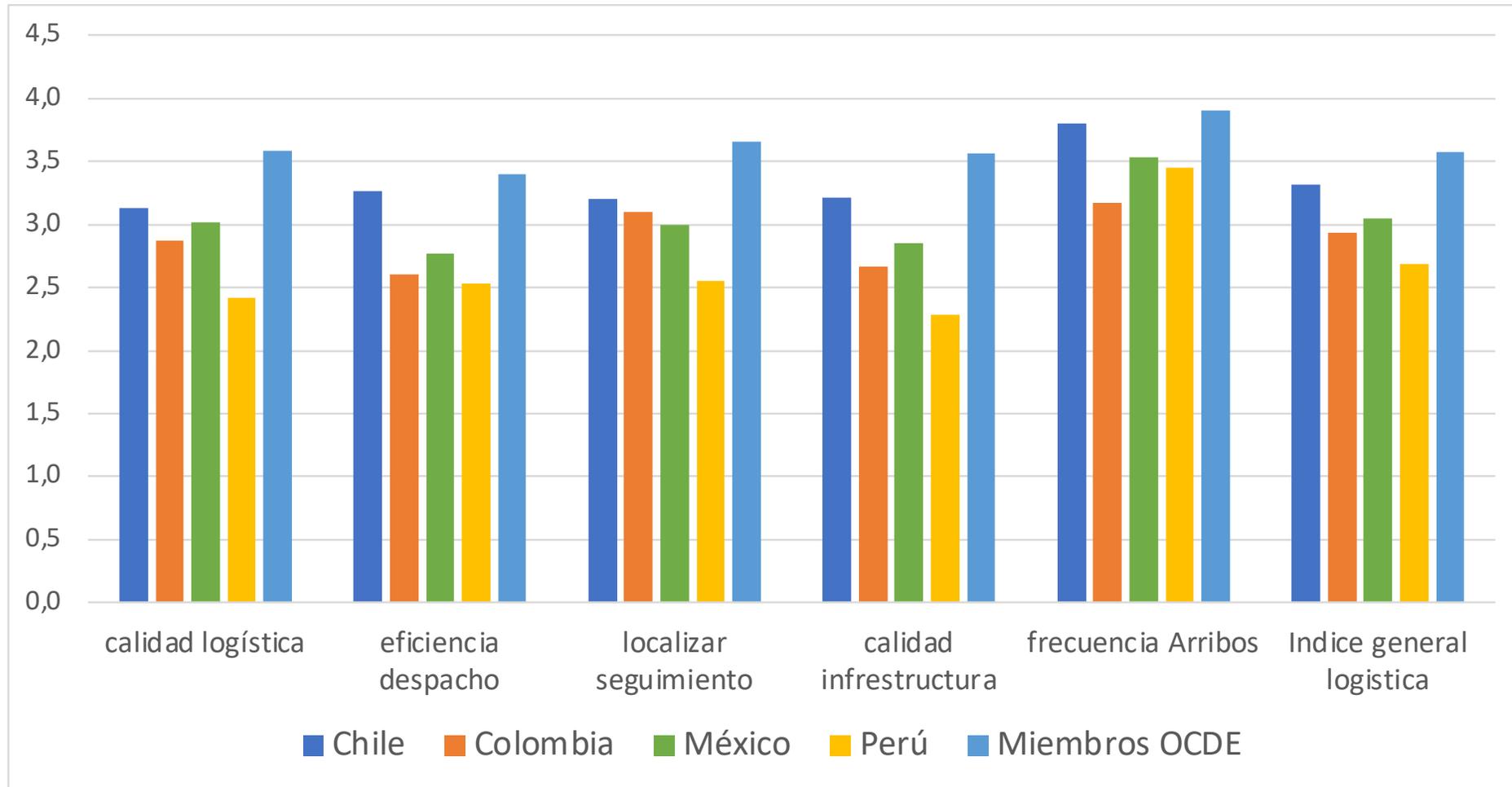
Universo:

> 5000 evaluaciones

> 1000 servicios de carga

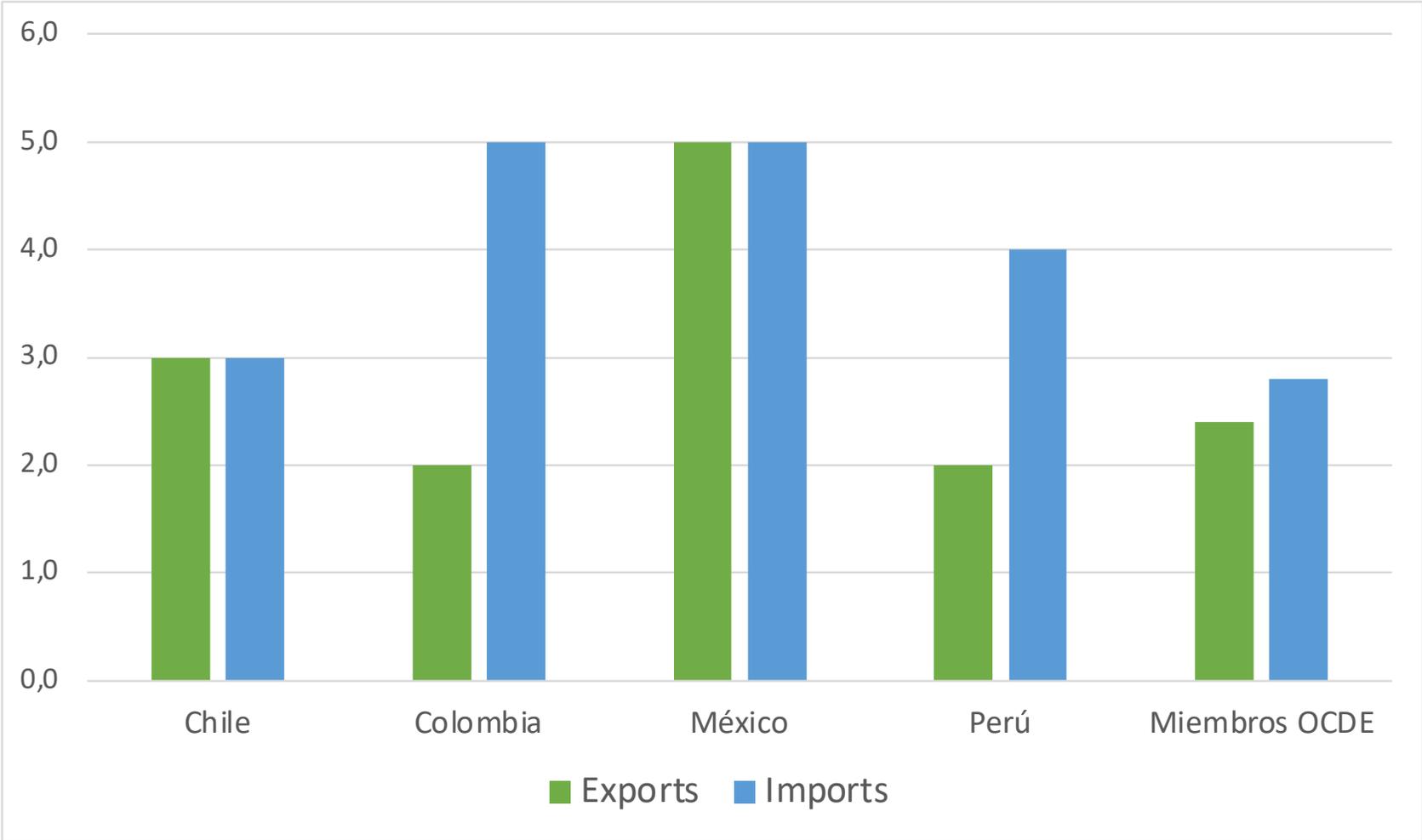
Índice de Desempeño Logístico refleja las percepciones de la logística de un país en función de la (1) eficiencia del proceso de despacho de aduanas, (2) la calidad de la infraestructura relacionada con el comercio y el transporte, (3) la facilidad de organizar envíos a precios competitivos, (4) la calidad de los servicios de logística, (5) la capacidad de rastrear envíos y (6) la frecuencia con la cual los envíos llegan al destinatario dentro del tiempo programado

Indice de desempeño logístico y sus componentes, datos para 2018. Fuente: Banco Mundial



<https://datos.bancomundial.org/indicador/LP.LPI.CUST.XQ?view=chart>

Plazos de entrega para las **importaciones y las exportaciones** de los países de la Alianza del Pacífico (días mdianos). *Fuente:* Banco Mundial



1.4. Medios de pago y medición del comercio electrónico

Fases en la compra:

- (1) la orden de compra/ venta,
- (2) el pago,
- (3) el reparto o distribución final

“toda contratación hecha online....”

- cada vez importa más (retail y wholesale)
- diferente contabilidad con respecto de “valor añadido” (PIB)
- B2C, B2B, B2G
- distintas fuentes (pagos electrónicos sin presencia de tarjeta física, EDI, plataformas de comercio- e....)
- dificultades para medir el comercio- e trans-fronterizo (mero tránsito, clasificaciones de “servicios” heterogéneas.....)

- Clase 479x de NACE (rev.4)/ ISIC, “ventas retail en comercios no -físicos...”

- Fuentes:

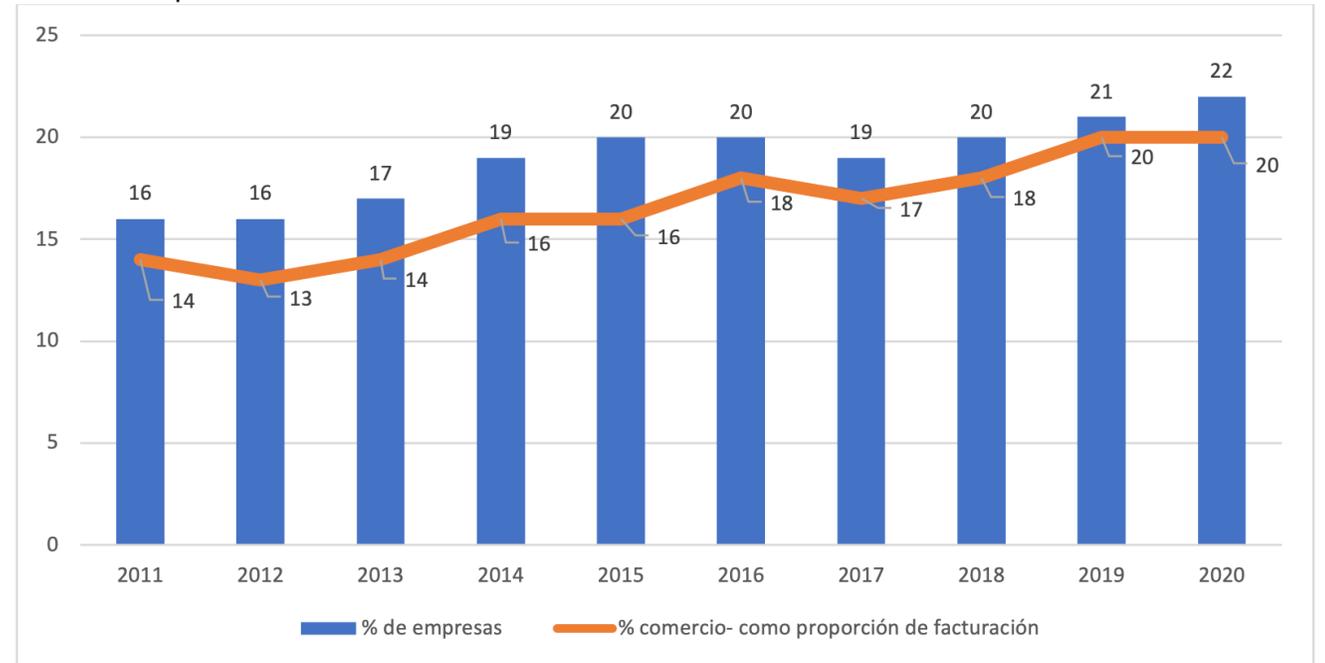
(1) encuestas (a Empresas, Hogares)

(2) registro de pagos online

(3) registros delivery (bienes)

(4) registros plataformas de comercio-e

Figura 3: Proporción de empresas que utilizan comercio-e y proporción de ventas online sobre ventas totales en la Unión Europea. Fuente: Eurostat



2. Problemas en la recogida, tratamiento y acuerdos de colaboración público- privados

4.1. Problemas administrativos en acceder a los datos de movilidad

4.2. Problemas técnicos del procesamiento de datos y problemas metodológicos

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

4.2. Problemas técnicos del procesamiento de datos y problemas metodológicos

Algunos aspectos metodológicos que deben ser tenidos en cuenta:

(1) **anonimización:** los datos de localización o de uso de los individuos caen dentro de la protección a la privacidad en cualquier jurisdicción. Es necesario consentimiento del usuario final, o bien es necesario disponer la capacidad legal por parte del operador para recolectar y tratar esos datos o transformar esos datos en otros tales que garanticen la anonimidad

(2) **procesamiento de datos:** limpieza de las observaciones. Pueden aparecer mediciones solo parcialmente recogidas, incompletas, erróneas; pueden aparecer datos atípicos, duplicados, o datos faltantes.

(3) necesidad de **aplicar nuevo software** que gestionen el gran volumen de datos. Usualmente el volumen de datos que se puede recoger es enorme y las herramientas tradicionales para el uso y manejo de datos no están adaptadas. El almacenamiento de datos requiere de servidores con garantías de seguridad y el trabajo posterior exige frecuentemente computación en la nube

y colaboración público- privada

- 1. Center for Big Data Statistics (CBDS)** en Holanda fue creado en 2016 y cuenta con acuerdos de colaboración con las ONE de otros países. Colabora estrechamente con Eurostat en proyectos concretos y tiene colaboraciones con proveedores de datos (como Google) y otros agentes involucrados en la cadena de valor (IBM, KPN, TMO, SURFSara, Microsoft, Dell, entre otros). CBS tiene también acuerdos concretos de colaboración con Universidades y centros de investigación.
- 2. Nic.br y el Instituto Brasileiro de Geografía y Estadística**, para medición de “economía de internet”

3. Pequeño cuestionario a Oficinas Nacionales de Estadística y otros agentes públicos

Objetivo: conocer el uso de Big Data, Geographic Information System (GIS), Inteligencia Artificial, Web Scraping y otras técnicas digitales para la recolección y procesamiento de datos.

¿Tiene su organización experiencia con Big Data?

- a) datos de imágenes de satélite o geodata (GIS)
- b) datos procedentes de red u operador móvil
- c) datos de usuarios (procedentes de redes sociales)
- d) datos de crowdsourcing (obtenidos de aplicaciones móviles o de OpenStreetMap)
- e) datos procedentes de webscraping
- f) datos procedentes de scanner data/ de sensores
- g) datos procedentes del sector privado (datos financieros, por ejemplo)
- h) Datos/información extraída de documentos
- i) Datos procedentes de elementos o de terminales asociados a cosas con conexión SIM/ eSIM geolocalizados
- j) No hemos utilizado aún Big Data

¿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?

- a) A nivel geográfico, como por ejemplo Meta
- b) A nivel de localización, como por ejemplo OpenCellID
- c) Velocidad de conexiones de banda ancha, como ookla
- d) A nivel de población, como WorldPop o Meta
- e) Sobre dimensiones socio-económicas, como por ejemplo Meta
- f) Para la medición de cambios de los precios de productos
- g) Sobre volumen de comercio electrónico, como por ejemplo de intermediarios financieros, gestores de tarjetas de crédito/ débito
- h) Sobre publicidad
- i) Sobre transacciones financieras o de pago
- j) Sobre aspectos meteorológicos o espaciales

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

Razones

- a) Limitaciones derivadas de la legislación vigente sobre la producción estadística oficial impone muchas barreras y dificultades
- b) Los agentes propietarios o gestores de los Big Data no muestran interés en colaborar
- c) El costo monetario que supone monetario adquirir Big Data es muy alto
- d) 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)
- e) La gobernanza, gestión y distribución de responsabilidades es costosa de diseñar y de implementar

¿Usa su organización alguna de las siguientes tecnologías de IA?

- a) Tecnologías que permiten el análisis de lenguaje escrito (minería de textos)
- b) Tecnologías que permiten transcribir lenguaje oral en medio que sea posible sea leído por una máquina o software?
- c) Tecnologías que generen lenguaje oral o escrito por si mismas?
- d) Tecnologías capaces de identificar personas u objetos basadas en la imagen (reconocimiento de imagen, procesamiento de imagen)
- e) Aprendizaje de máquinas (machine learning) para el análisis de datos?
- f) 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)

¿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)?

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

¿Considera que la legislación aplicable de protección de datos está bien diseñada para proteger a los individuos, por un lado, y para incentivar el uso de Big Data?

Muchas gracias!!!