



# Colonialismo de Datos



**CLAUDIO GUTIÉRREZ**

Ph.D. Computer Science por la Wesleyan University. Magíster en Lógica Matemática por la Pontificia Universidad Católica de Chile. Licenciado en Matemáticas por la Universidad de Chile. Profesor Titular del Departamento de Ciencias de la Computación de la Universidad de Chile. Investigador Senior del Instituto Milenio Fundamentos de los Datos. Líneas de investigación: fundamentos de los datos, ciencia de datos, lógica aplicada a la computación, semántica de la Web.

✉ [cgutierr@dcc.uchile.cl](mailto:cgutierr@dcc.uchile.cl)



**RESUMEN.** El siguiente artículo analiza cómo las grandes empresas de datos, como Google y Meta, actúan de manera similar a los estados coloniales al capturar y explotar datos personales. Este “colonialismo de datos” se caracteriza por la centralización y el uso de datos humanos, a menudo sin el conocimiento de quienes les proveen los datos, reflejando los métodos de explotación del modelo colonial. También describe la infraestructura que estas empresas han desarrollado para capturar, almacenar y analizar datos, y la falta de regulación y la asimetría de poder entre estas compañías y de quienes les proveen los datos. Se sugiere la necesidad de modelos alternativos y regulaciones que devuelvan el control de los datos a las personas, con iniciativas como el proyecto SOLID de Tim Berners-Lee, que busca mejorar la privacidad y la propiedad de los datos personales.

“Colonialismo” parece ser un término sacado de una discusión política sesentera. La RAE lo define como “régimen político y económico en el que un Estado controla y explota un territorio ajeno al suyo”. La noción es fácil de ejemplificar con casos icónicos: los tres siglos coloniales de España en Latinoamérica; Inglaterra en la India; Francia, Alemania, Bélgica y varios más en África; etc.

Este modelo de apropiación, uso y centralización de recursos ajenos ayuda

mucho a entender un fenómeno muy actual: la lógica de funcionamiento de grandes compañías de datos, organizaciones muy diferentes a los Estados, aunque más poderosos que muchos de ellos.<sup>1</sup> Es la metáfora que aprovechan Couldry y Mejías en su tesis de “colonialismo de datos” [1] y que presentamos en este artículo.

## Datos y compañías de datos

Una compañía o empresa de datos (ED) es aquella donde los datos están al centro de sus operaciones, cuya fuente de negocios es la captura, almacenamiento, análisis y generación de valor a partir de datos de las personas y sus actividades. Buenos ejemplos de grandes EDs son Google, Amazon, Meta, Oracle, Microsoft, Apple, Alibaba, Tencent, Baidu, etc.<sup>2</sup>

La empresa tradicional es una organización que combina un conjunto de factores para producir un bien o servicio para la venta en el mercado. Esos factores tradicionalmente son la tierra, el trabajo y el capital; muy pronto se le agregó la “capacidad empresarial” (tecnología, conocimientos, organización). Las empresas de datos como las mencionadas, aparentemente destacan por esto último (de hecho, ellas se autopromocionan así). Sin embargo, su característica principal es su “materia prima”: los datos, que conforman la esencia de su flujo productivo y de servicios. Y técnicamente los datos los generan los usuarios. Entonces, ¿a qué factor productivo corresponden los datos? Son

“tierra” o “trabajo” o “capital”? Mucho del debate sobre cómo considerarlos tiene que ver con esta pregunta.

Tradicionalmente los datos eran “naturaleza” (“tierra” en lenguaje de los factores de producción). Obtener conocimiento por medio de datos (llamadas “observaciones”) está en el corazón de los inicios del método científico y de la tecnología en la producción industrial. Francis Bacon, uno de sus impulsores, sostenía por ello que “el conocimiento (obtenido de los datos) es poder”. El método se desarrolló durante varios siglos siempre considerando los datos como un asunto “natural”, una propiedad de la naturaleza que los científicos lograban “arrancarle”.

Bastante más tarde, durante el siglo XX, se comenzó a “recoger” (y procesar) sistemáticamente datos sociales y de humanos, esto es, estadísticas de Estado y población.<sup>3</sup> Se desarrollan métodos estadísticos, así como preocupaciones técnicas y metodológicas para tratarlos. Y entrado el siglo XX comienzan tímidamente las preocupaciones éticas (con el mundo de las encuestas, los “trabajos en terreno” de los antropólogos, etc.) [4].

Pero el gran salto comienza con la “automatización” del tratamiento de datos, debido a avances tecnológicos como computadores, sensores, redes, digitalización de imágenes, audio y video, etc. Sin embargo, el real “boom” de los datos tiene que ver con la creación de la Web a comienzos de 1990, un espacio que permitió y facilitó que diferentes personas y organizaciones pudieran comunicarse por medio de ella en relaciones de amistad, comerciales, de

- 1 No abordamos aquí las facetas de soberanía (a nivel de naciones) que involucra este fenómeno, tanto de los datos como de las grandes empresas de datos. Ver por ejemplo [2,3].
- 2 Cuando hablamos de grandes, hablamos de su infraestructura de hardware, software y de su personal. Por ejemplo, Meta tiene 80.000 empleados full time. Amazon emplea sólo en Estados Unidos alrededor de un millón de personas.
- 3 Desde siempre los Estados recogieron datos de su población para impuestos, fines militares, etc., pero no eran objetos que recibieran atención sistemática per se.

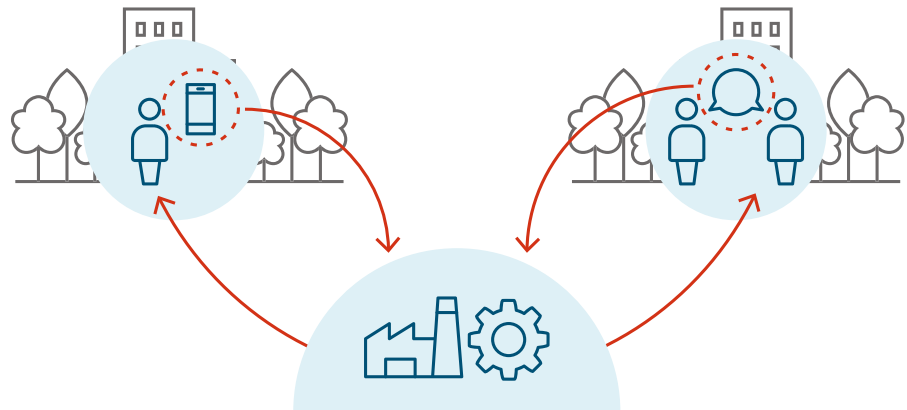


## Entonces los datos más requeridos por las grandes empresas de datos comenzaron a ser los de humanos.

trabajo, etc. En esa infraestructura universal circulaba la “sangre” de esas comunicaciones: los datos. El paso siguiente consistió en capturarlos y almacenarlos y así nacieron las grandes empresas de datos (ver Figura 1).

La novedad es que el grueso de la actividad de estas empresas de datos se ha concentrado en cierto tipo de datos. Para efectos pedagógicos (la realidad es más sutil) podemos organizarlos en dos dimensiones: datos de la naturaleza versus datos (de los) humanos por un lado<sup>4</sup>, y datos explícitos versus datos tácitos por el otro.<sup>5</sup>

Los datos más requeridos por las grandes ED comenzaron a ser los de humanos y tácitos. Primero, no es gran sorpresa que todos los datos interesantes giren alrededor de lo humano, directa o indirectamente. Un dato geográfico —piénsese en las grandes exploraciones— es “valioso” en tanto sirva para mapear el territorio; un dato meteorológico para la actividad diaria o indirectamente para la agricultura, el transporte, etc., esto es, usos o impacto finales sobre los humanos. Sin embargo, son los datos humanos, aquellos que permiten “conocer” e incidir directamente sobre los humanos, los que han ido tomando



**Figura 1.** Las grandes empresas de datos y sus “materias primas”.

el centro de la atención. Y mejor si son tácitos, esto es, capturados sin que los sujetos lo adviertan, pues de esa forma capturan el “funcionamiento en modo automático” de personas y grupos sociales, los hábitos, las costumbres, su interacción con aparatos (el Internet de las cosas), etc. sin tener el sesgo de la respuesta “pensada”, “consciente”. Esos datos son realmente el “nuevo petróleo” de las grandes compañías y son complementados con los de transacciones comerciales, financieras, administrativas y legales.<sup>6</sup> Y los siguen aquellos que directamente inciden en la sociedad y la economía humanas (datos agrícolas, meteorológicos, geográficos, de recursos naturales, etc.). A todo lo anterior, podríamos agregarle una tercera dimensión: datos estructurales o estáticos, versus datos funcionales o dinámicos. Los primeros son relevantes, pero absolutamente insufi-

cientes pues se necesita su dinámica, su “fisiología”.

Para aprovechar este “nuevo petróleo” (los datos), una ED requiere una infraestructura que revisaremos brevemente. No abordamos aquí el relevante tema de las regulaciones, sobre todo de los datos de humanos, que son una materia que sólo recientemente se ha comenzado a incorporar.

## Cuatro facetas esenciales de una ED

Veamos cómo los datos ocurren en el flujo de producción de esas compañías (y en muchas otras de menor envergadura). Como decíamos, su modelo de negocio es la creciente captura, control

4 En lo que sigue hablaremos directamente de “datos naturales” y “datos humanos”. Esto puede confundir, pero queremos resaltar la contraposición natural–humano. Los datos de humanos realmente se refieren a datos que involucran o son generados —conscientemente o no— por los humanos (su cuerpo, su mente, sus actividades) directamente: datos biométricos, datos de actividad corporal, de gustos, de sentimientos, razonamiento, relaciones sociales, etc. Particularmente los datos del lenguaje son hoy cruciales para los modelos de inteligencia artificial generativos.

5 No confundir esta dimensión con el consentimiento. Dar consentimiento es permitir consciente y expresamente la entrega de datos. Ese consentimiento puede ser para captura de datos explícita (ejemplo: responder una encuesta) o tácita (ejemplo: permitir que lo filmen mientras desarrolla alguna actividad).

6 Sobre la deriva de cómo las grandes ED consideran estos datos, ver [5].



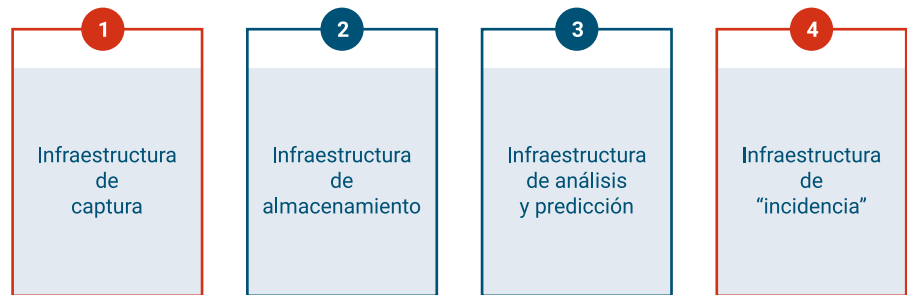
y explotación de los datos de las personas, grupos sociales y organizaciones. En esta lógica, una empresa de datos consta de cuatro unidades indispensables (ver Figura 2).

## Y poco a poco cambió su eje de actividades: de capturar datos para mejorar servicios a desarrollar servicios para capturar datos.

### 1. Infraestructura para capturar, conseguir, datos.

Dada la centralidad de los datos de humanos, desde el punto de vista técnico esto significa, por una parte, la posibilidad de llegar a cada una de las personas (su individualidad), y por otra, capturar sus interacciones. Veámos que la red universal (Internet y la Web) es clave pues allí circula una gran (cada día mayor) parte de las interacciones humanas y datos de individuos. Las grandes ED comenzaron sus negocios ofreciendo servicios para aplicaciones en la Web. De capturar en sus inicios datos de sus usuarios para mejorar sus procesos y servicios, muy pronto advirtieron que hacerse de esos datos que circulaban era un tremendo negocio. Y poco a poco cambió su eje de actividades: de capturar datos para mejorar servicios a desarrollar servicios para capturar datos.

Así, de esos orígenes “modestos” se escaló a la idea de capturar digitalmente toda posible faceta de la realidad y de la interacción social, sea dato simple, texto, imagen, sonido, video. Las apps y los aparatos digitales, que resuelven problemas de la gente u ofrecen nuevas actividades, en ese proceso capturan sus interacciones con ellos, esto es, sus maneras de comportarse ante diferentes situaciones, como vitrinear un producto, el proceso de compra, de elección, los horarios de actividad, los



**Figura 2.** De estas cuatro facetas, la (1) y la (4) son las claves, las que hacen la novedad y son las que “rozan” con la legislación; (2) y (3) son indispensables, pero son soporte para las otras dos.

gustos, las relaciones sociales de amistad, de trabajo, de amor, las expresiones artísticas, científicas o políticas, los juegos y distracciones, etc.<sup>7</sup>

Las presiones por parte de estas ED sobre la neutralidad de la red (de Internet) y de los protocolos y estándares de la Web han sido brutales.<sup>8</sup> Aquí se confabulan empresas de datos con proveedores de servicios de internet (ISP) y compañías telefónicas [6].

### 2. Infraestructura para integrar, almacenar, organizar y disponibilizar datos.

Estas grandes EDs no serían posibles sin una integración de los datos capturados. Esta centralización descomunal

se da tanto a nivel privado (por ejemplo, Estados Unidos) como a nivel estatal (por ejemplo, China). Se requiere una capacidad casi inimaginable de hardware, software y sistemas para coordinar las miles de millones de fuentes, de integrarlas coherentemente, de organizarlas para sus diferentes usos.<sup>9</sup>

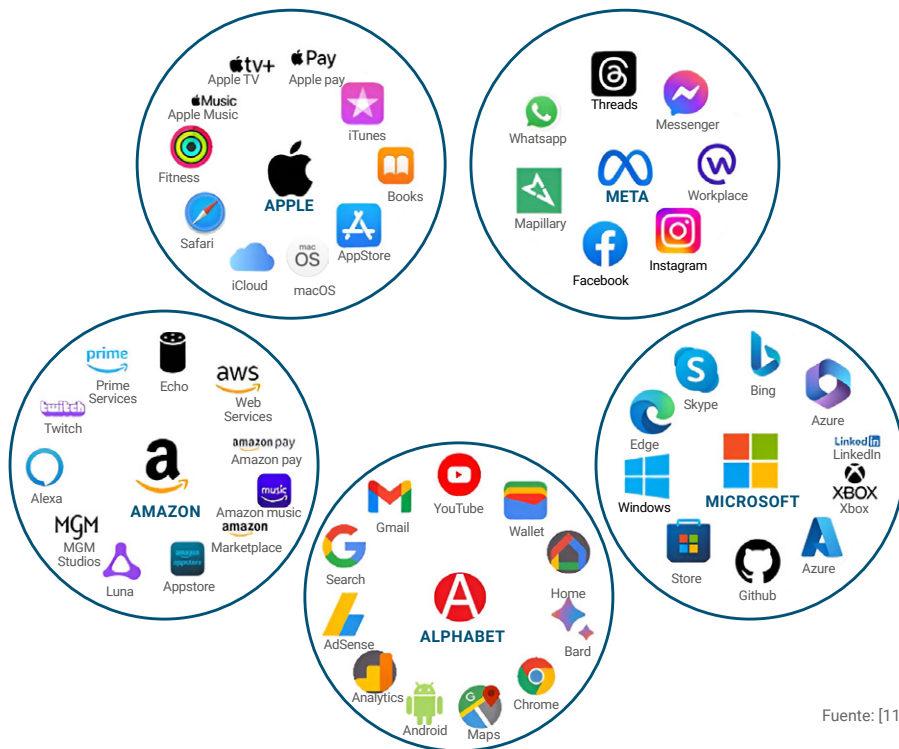
Esas colosales infraestructuras muestran el abismo que se ha ido creando entre las grandes compañías y el resto de las organizaciones como universidades, agencias de gobierno, empresas medianas y pequeñas, y por supuesto, los usuarios finales. Sus centros de datos son un buen ejemplo: son una red global que involucra millones de servidores distribuidos en todo el mundo (en decenas de países) para asegurar baja latencia y alta disponibilidad. Esta

7 Shoshana Zuboff le ha llamado a este proceso el paso al “capitalismo de vigilancia” [7, 8]. Destaca 5 maneras de extraer datos hoy: (1) de transacciones económicas mediadas por computadores; (2) sensores incrustados en objetos, cuerpos y lugares; (3) bases de datos de corporaciones y agencias de gobierno; (4) vigilancia pública y privada a través de cámaras, celulares, satélites; y (5) actividades de usuarios de juegos, búsquedas, música, redes sociales, etc.

8 El principio de neutralidad de la red establece que los ISPs (Proveedores de Servicios de Internet) traten todo el tráfico en línea igual y de manera abierta, sin discriminación, bloqueo, limitación o priorización.

9 <https://www.statista.com/topics/4213/google-apple-facebook-amazon-and-microsoft-gafam/>.





Fuente: [11]

**Figura 3.** Infraestructura digital de servicios provista por las grandes EDs.

“nube” de centros de datos también es usada para prestar servicios a empresas y gobiernos en todo el mundo. Almacenan y trabajan con datos del orden de los exabaytes.<sup>10</sup> A todo ello hay que sumarle la gestión y organización del almacenamiento, la redes de comunicaciones, la capacidad de disponibilizar estos datos.

Sobre esta faceta de infraestructura, existen también iniciativas públicas (algo más modestas) que provienen sobre todo del mundo de los sistemas de

información, en la forma de repositorios abiertos de organizaciones y de gobiernos.<sup>11</sup> La idea es disponibilizar datos con las correspondientes regulaciones para que cumplan estándares de calidad y queden al servicio de todos y no sólo de quienes tengan poderes particulares. Se han generado muchos estándares para ello como FAIR,<sup>12</sup> OASIS,<sup>13</sup> certificaciones como CoreTrustSeal,<sup>14</sup> y los principios TRUST (transparencia, responsabilidad, foco en usuarios, sostenibilidad, tecnología adecuada) [9].

### 3. Analítica para predecir (futuro) en diferentes facetas.

No basta con almacenar los datos. Las grandes EDs usan sofisticadas técnicas estadísticas, de optimización y computacionales para analizar esos datos en lo que se conoce como “analítica” de datos. Es una combinación de poderosa infraestructura computacional de hardware y software, con sofisticadas técnicas de optimización y herramientas estadísticas que han permitido “entrenar” gigantescas redes neuronales con complejas arquitecturas y fantásticas cantidades de datos de entrenamiento.

Un buen ejemplo son los datos necesarios para entrenar LLMs (*Large Language Models*). La iniciativa Common Crawl ha sido muy relevante en ello.<sup>15</sup>

El software base que usan es generalmente abierto (sus servidores corren Linux y Apache), sus centros de datos usan MySQL y otros lenguajes abiertos, el software de procesamiento de datos como Hadoop, Spark, etc. es conocido y abierto. Aunque todo esto está teóricamente al alcance de todos, la escala que se requiere está lejos de poder alcanzarse por cualquier organización o universidad hoy día. También aquí las grandes EDs lideran en sus servicios [10].

### 4. Infraestructura de servicios para “aplicar” esas predicciones en diferentes ámbitos.

De lo que trata el negocio es una vez analizados los datos, poder “volver” a

10 Un notebook tiene capacidad para almacenar varios terabytes; para almacenar un petabyte necesita 1.000 de esos notebooks; para almacenar un exabyte necesitaría un millón de notebooks adecuadamente integrados. Un data center está limitado por capacidades físicas de refrigeración y energía.

11 Un buen ejemplo es el European Open Science Cloud (EOSC): <https://eosc-portal.eu/>.

12 Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* 3, 160018 (2016).

13 Reference Model for an Open Archival Information System (OAIS). Recommended Practice CCSDS 650.0-M-2. Consultative Committee for Space Data Systems, (2012).

14 CoreTrustSeal, <https://www.coretrustseal.org/why-certification/certified-repositories/> (2020).

15 <https://commoncrawl.org/>.



## De lo que trata el negocio es una vez analizados los datos, poder “volver” a los usuarios con recomendaciones y sutiles maneras de conducirlos y orientarlos.

los usuarios ya sea con recomendaciones y sutiles maneras de conducirlos y orientarlos. Esta es la característica que materializa el “poder” de los datos. Estos “servicios” se enlazan con el sistema de captura de datos. En la Figura 3 se muestra la infraestructura digital de servicios provista por las grandes EDs.

Todo lo anterior comenzó de una manera “ingenua”: en Google y Amazon usando los clicks de los usuarios para determinar de mejor manera preferencias. En Facebook y Twitter usando las relaciones y comunidades que se iban formando entre los usuarios para mejorar el servicio y atraer más usuarios. En los diversos juegos, aprovechando los datos que dejaban los jugadores para diseñar nuevas funcionalidades. Todo muy “ingenuo” y “sano”... Pero en paralelo fue creciendo la industria del aprovechamiento de ese conocimiento del comportamiento de los usuarios para otros fines: desde la creación de nuevos productos, la ubicación y formas adecuada de la propaganda; hasta los usos políticos como en el escándalo de Cambridge Analítica.<sup>16</sup>

En una primera época, todo este proceso se realizaba sin ninguna regulación pues los datos eran asunto nuevo. Hoy se dan fuertes disputas con las empresas de la competencia, con las regulaciones de Estados (como las de la EU),<sup>17</sup> y con las organizaciones de usuarios que están haciéndose conscientes del tema. Suficiente para los datos. Vamos al colonialismo (de datos).

---

## Colonialismo de datos

---

La expresión “colonialismo de datos” fue popularizada por los investigadores Couldry y Mejías [1, 12]. La resumen así: “Se está construyendo un nuevo orden a través de la extracción continua de datos de nuestras vidas sociales”. Y entonces sostienen que este proceso, “esta cuantificación de lo social representa un nuevo movimiento colonial. Aunque los modos, intensidades, escalas y contextos de desposesión han cambiado, el impulso subyacente del colonialismo de datos actual sigue siendo el mismo: adquirir “territorio” y recursos de los que el capital pueda extraer valor económico” [1].

Lo crucial del colonialismo es la diferencia entre el colonizador y el colonizado, entre quien ejerce el poder y se apodera de los recursos, y quienes tienen pocos recursos y poder y permanecen desposeídos. En el ámbito digital, es una dimensión mucho más profunda que la “brecha digital”, que se definía como una “desigualdad en el acceso” entre grupos sociales, pero donde aparentemente quienes tenían más acceso no “explotaban”, no vivían de, quienes tenían menos acceso. Hay brecha entre quienes tienen acceso a los grandes conciertos en una ciudad capital, y los habitantes de un pueblo donde no llega nadie. Pero es difícil argumentar que aquellos explotan a éstos, que desarrollan su actividad “gracias” a la desposesión de los otros.

En el terreno de lo colonial, el “éxito” del colonizador no puede realizarse sin los recursos del colonizado y el uso de la coacción, sea física o económica. Es una relación directa, inmediata. El éxito de una compañía azucarera belga estaba en directa relación con la miseria, la desposesión y la sujeción de los esclavos en las colonias caribeñas. El éxito de las compañías inglesas del estaño estaba directamente ligado a la explotación de los mineros bolivianos.

Lo novedoso de las compañías coloniales de datos es que, aunque viven de los datos de sus “usuarios”, se enmarcan en un tipo de “contrato” donde pareciera haber un acuerdo entre quien entrega sus datos y quien se los apropia. A cambio del uso de una app, del correo, de una red social, etc. el “usuario” acuerda entregar sus datos a la compañía. El “éxito” de la compañía no está (al menos inmediatamente) relacionada con la desposesión del usuario. Algunas veces al contrario: el usuario parece salir beneficiado con este “contrato” (pongo las comillas, pues es un contrato muy asimétrico: no hay negociación, sino reglas que deben ser aceptadas si uno quiere entrar en ese sistema). Esto es lo complejo del nuevo modelo.

El gran “truco” de magia<sup>18</sup> de esas compañías consiste en aprovechar la esencia de lo que es un dato: una marca, algo inocuo, muchas veces inservible, que no afecta (directamente) la materialidad del objeto del que se obtiene, pero que en contextos adecuados, obtiene todo su valor. Así la empresa aparece recogiendo “basura” (marcas “inútiles”), que una vez reunidas con otras, con contextos adecuados, y procesados con técnicas estadísticas y analíticas sofisticadas, sus resultados pueden ser aprovechables para incidir o para afectar (negativa

---

16 <https://bipartisanpolicy.org/blog/cambridge-analytica-controversy/>.

17 La más conocida de las regulaciones directamente aplicables a los datos es la GDPR europea, que es de 2018. [https://commission.europa.eu/law/law-topic/data-protection/data-protection-eu\\_es](https://commission.europa.eu/law/law-topic/data-protection/data-protection-eu_es).

18 La esencia del arte del mago es distraer la atención con algo llamativo, para así poder hacer pasar desapercibido lo relevante que está haciendo y producir el efecto “mágico” cuyo proceso de generación nadie advertirá.



**Couldry y Mejías sostienen que el colonialismo "tradicional" posee cuatro características claves [...]. No es difícil advertir que es exactamente lo que ocurre hoy en el mundo digital de los datos sociales.**

o positivamente) al objeto del cual se obtuvieron. En tiempos del método científico "natural" nada de esto importaba. Sacar las muestras de un árbol para estudiar sus propiedades, analizarlas, y luego volver con algún producto o técnica para incidir sobre su crecimiento, producción, etc. no aparece como abusivo. Cuando en vez de árboles, tratamos con humanos, el asunto cambia.

Desde siempre está mal o prohibido por ley "afectar" a otras personas. ¿Cómo se logra esquivar esto en el mundo de los datos? Aprovechando (y desarrollando) el sentido común (de otra época) de que los datos no son relevantes, una "ideología" que permite que esta actividad se vea como una más, e inocua para las personas. Más aún, que esto se vea como "natural". Ello se aprovechaba usando la asimetría de información y poder. El jurista nazi Carl Schmitt justificaba así el colonialismo: "El descubrimiento europeo de un nuevo mundo en los siglos XV y XVI no se produjo, pues, por casualidad [...] fue un logro del recién despertado racionalismo occidental [...]. Los indios carecían del poder científico de la racionalidad cristiano-europea. La ventaja intelectual estaba totalmente del lado europeo, hasta el punto de que el Nuevo Mundo pudo ser simplemente 'tomado' " [12]. Nos interesa quedarnos con el fondo del argumento: en algunos momentos de la historia, ciertos grupos, organizaciones o países tienen alguna "ventaja" por sobre otros, lo que les per-



mite simplemente "tomar" los recursos de otros. Hoy, las grandes ED quienes tienen los recursos, usan el "sentido común" de la gente y no perturban "aparentemente" a nadie, y de esa forma pueden hacer sentir que sus actividades son naturales. Más aún, para muchos (como para Schmitt el colonialismo) sería un "premio" a sus "logros" tecnológicos.

Couldry y Mejías sostienen que el colonialismo "tradicional" posee cuatro características claves: (I) la apropiación de recursos; (II) la formación de nuevos órdenes sociales; (III) una concentración extrema de la riqueza; y (IV) la creación de ideologías para justificar las prácticas de apropiación. No es difícil advertir que es exactamente lo que ocurre hoy en el mundo digital de los datos sociales. Sostienen que lo que está en juego hoy es la apropiación de la vida social más que territorios y riquezas materiales como antes. Y concluyen: "Aunque los modos, intensidades, escalas y contextos de desposesión han cambiado, el impulso subyacente del colonialismo de datos actual sigue siendo el mismo: adquirir "territorio" y recursos de los que el capital pueda extraer valor económico" [1]

(traducción realizada gratuitamente con el traductor DeepL.com, previa entrega de mis datos...).

## A manera de conclusiones

Estamos viviendo tiempos interesantes (como dice el adagio chino). Entrando en un nuevo mundo lleno de posibilidades, pero amenazados también por prácticas conocidas que apropian esas posibilidades para dirigir las en función de intereses estrechos.

Hemos mostrado los problemas derivados del colonialismo de datos. ¿Cuáles son las alternativas no sólo para evitarlo sino para desarrollar modelos alternativos? Hay iniciativas muy interesantes al respecto. Una de ellas es el proyecto SOLID que lidera Tim Berners-Lee, el creador de la Web como espacio universal distribuido. SOLID es un proyecto que apunta a "cambiar radicalmente la manera en que las aplicaciones en la Web funcionan hoy, para que se tenga una real propiedad de los datos (personales) y



una mejora de la privacidad”.<sup>19</sup> Esencialmente es “un conjunto de convenciones y herramientas para construir aplicaciones descentralizadas basados en principios de datos enlazados”. Hay muchas iniciativas como éstas, pero son bastante desconocidas por el gran público.

Es importante recordar que el modelo colonial no sólo conllevaba la extracción de recursos, sino un marco de poder para impedir el desarrollo de otros modelos. Por ello es importante preguntarse: ¿cuáles son los desarrollos en que podemos ayudar? Probablemente lo central son aplicaciones que permitan que los usuarios (personas, organizaciones, comunidades) tengan control sobre sus propios “recursos”, esto es, sus datos. En salud, la ficha única electrónica;

en política y administración, el gobierno de datos y las infraestructuras distribuidas, en inteligencia artificial, los LLMs abiertos; en publicación de datos, los mecanismos de lo abierto, de lo común, de la privacidad diferencial, etc.

Para terminar, van aquí algunas sugerencias sobre maneras de enfrentar el colonialismo de datos<sup>20</sup>:

1. Identificar las maneras en las cuales los procesos de datificación están operando en su vida diaria.
2. Aborde el mundo de los datos como un bien común y eduque a sus comunidades en las posibilidades y las limitaciones y peligros del nuevo mundo de los datos.

3. Apoye las legislaciones y políticas que regulen los abusos de las plataformas, la concentración de mercados en el área, la profundización de las brechas digitales.
4. Promueva alternativas de gobernanza sobre los datos a nivel de organizaciones, comunidades y gobierno.
5. Desarrolle herramientas y modelos alternativos que aborden los problemas que generan para la población las arquitecturas centralizadas de las grandes compañías.
6. Preocúpese de los trabajadores que quedan en desventaja debido a los cambios que genera la economía de datos. ■

## REFERENCIAS

- [1] Couldry, Nick; Mejías, Ulises. The Costs of Connection How Data Is Colonizing Human Life and Appropriating It for Capitalism. En castellano: El costo de la conexión. Edic. Godot.
- [2] Ávila Pinto, Renata. (2018). Digital sovereignty or digital colonialism. *Sur International Journal on Human Rights*, 15(27), 15-28.
- [3] Gu, H. Data, Big Tech, and the New Concept of Sovereignty. *J OF CHIN POLIT SCI* (2023). <https://doi.org/10.1007/s11366-023-09855-1>.
- [4] Alain Desrosieres. *The Politics of Large Numbers. A History of Statistical Reasoning*. Harvard Univ. Press, 1998.
- [5] Birch, K., Cochrane, D., & Ward, C. (2021). Data as asset? The measurement, governance, and valuation of digital personal data by Big Tech. *Big Data & Society*, 8(1). <https://doi.org/10.1177/20539517211017308>.
- [6] La batalla por un Internet abierto. Políticas de la FCC y sus efectos. <https://fastercapital.com/es/contenido/La-batalla-por-un-internet-abierto-politicas-de-FCC-y-sus-efectos.html>.
- [7] Zuboff, Shoshana. Big Other: surveillance capitalism and the prospects of an information civilization. *Journal of Information Technology* (2015) 30, 75-89.
- [8] Zuboff, Shoshana. *La Era del Capitalismo de Vigilancia*. Paidós, 2020.
- [9] Lin, D., Crabtree, J., Dillo, I. *et al.* The TRUST Principles for digital repositories. *Sci Data* 7, 144 (2020).
- [10] Berisha, B., Mëziu, E. & Shabani, I. Big data analytics in Cloud computing: an overview. *J Cloud Comp* 11, 24 (2022). <https://doi.org/10.1186/s13677-022-00301-w>.
- [11] Shaleen Khanal, Hongzhou Zhang, Araz Taeihagh, Why and how is the power of Big Tech increasing in the policy process? The case of generative AI. *Policy and Society*, 2024; <https://doi.org/10.1093/polsoc/puae012>.
- [12] Couldry, N. & Mejías, U. A. (2019). Making data colonialism liveable: how might data's social order be regulated?. *Internet Policy Review*, 8(2). <https://doi.org/10.14763/2019.2.1411>.
- [13] Ricaurte Quijano, Paola. Resisting Data Colonialism: What Lies Ahead. En: *Tierra Común Network. Resisting Data Colonialism – A Practical Intervention*. Institute of Network Cultures, Amsterdam, 2023. <https://networkcultures.org/blog/publication/tod-50-resisting-data-colonialism-a-practical-intervention/>

19 Project SOLID <https://solid.mit.edu/>.

20 Tomamos algunos puntos del “decálogo” en [13], p. 97.