



Gobernar el nuevo mundo de los datos





CLAUDIO GUTIÉRREZ

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

cguetierr@dcc.uchile.cl



RESUMEN: Gobierno, gobernanza, datos. Tres nociones tan lejanas a nuestra experiencia diaria que intuyo que el grueso de los lectores dirá: “No me interesan esas divagaciones inútiles” y llegará hasta aquí. Antes, permítanme listar las conclusiones, cuyo argumento para cada uno de sus puntos desarrollaré en este artículo.

Primero, los datos han pasado a constituir elementos fundamentales del mundo en que vivimos (esto significa que no podemos soslayar su presencia). Segundo, éste es un fenómeno nuevo, en el sentido que los marcos organizativos, sociales y políticos con que abordamos el mundo material se hacen insuficientes para abordarlo. Tercero, lo anterior repercute en muchas facetas de nuestra vida: relaciones sociales, educación, industria, comercio, salud, política, etc. Por todo lo anterior, y éste es el objetivo del artículo, se hace necesario determinar cómo decidiremos como sociedad la manera cómo abordaremos este nuevo mundo de los datos, esto es, su gobierno y gobernanza.

Introducción: el abordaje de los datos hoy

Estamos viviendo una época donde existe una capacidad tecnológica tan grande para capturar datos, almacenar datos, procesar y transformar datos, comunicar datos, analizar datos, que los datos se han transformado en una mina de oro para los “emprendedores”. Hoy es moda hablar de *big data*; florecen los emprendimientos de datos; estamos llenos de diplomados y magister en “*data science*”; quien quiera tener un trabajo decente, en casi en cualquier área debe incluir en su CV una línea que indique que tiene experiencia en el mundo de los datos.

La aproximación a los datos ha sido hasta hoy muy similar a la conquista de las riquezas de América, el *rush* del oro en California, o el petróleo en el medio oriente. Una carrera por un bien “natural” que está ahí y hay que “explotar”. Quien lo haga primero, de la manera más eficiente, “triumfa” (como las hoy empresas más rentables Google, Amazon, Facebook, Microsoft, etc.). El resto funcionamos como materia prima.

En este cuadro uno puede hacerse unas preguntas aparentemente ingenuas, motivadas por otras explotaciones intensivas de materias primas. ¿Es infinito ese recurso? ¿A quién pertenece? ¿Quién lo regula? ¿Es sustentable? ¿Puede provo-

car daños? ¿Cómo nos afecta su desarrollo? etc. Todas preguntas que tienen que ver con lo que se conoce como “gobierno” o “gobernanza”.¹ Estas cuestiones no han importado hasta ahora mucho a esos conquistadores de la riqueza de datos.

Pero son preguntas muy relevantes a la hora de abordar una actividad que parece tomar autonomía como los productos del hechicero de la leyenda. Pero, aclaro inmediatamente, el problema con los datos es mucho más fundamental que los grandes temas de explotaciones del petróleo, de los bosques, el mar o la energía nuclear, que giran en torno a la responsabilidad social. El gran tema de fondo con los datos no trata sólo, ni fundamentalmente, de que la falta de escrúpulos con un recurso lo podría acabar (como el petróleo, los bosques o los peces) o podría producir daños a la humanidad (como la energía nuclear o los plásticos). Esos son problemas importantes. Pero tenemos aquí otro problema, de órdenes de magnitud más complejo, que es el que quiero presentar en este artículo.

Resumo la tesis en un párrafo: los datos han pasado a ser parte sustancial del ecosistema que nos conforma como humanos (a diferencia del petróleo, de las mineras, de los bosques, las energías). Su poder incide sobre la cultura; sobre las instituciones; en la organización económica, social y política; en la educación y

en los hábitos; y finalmente, algo no menor, en los instintos y la biología de los humanos. Esto pues, los datos son lejos algo que va más allá de una simple materia prima a explotar, o de marcas y signos de fenómenos a manipular. Aclarar que hay datos y datos. Hasta ayer, los datos que lográbamos recolectar eran datos “inofensivos”: pocos, irrelevantes, estáticos y monolíticos (cartas, llamadas telefónicas, fotografías, periódicos, libros, estadísticas varias, etc.). Hoy todo ello y miles de otras facetas nuestras y de la naturaleza, están en modo virtual. Y cada día con más alcance. Tenemos prácticamente una copia virtual (digital) de nuestro viejo y querido mundo material. Y además, las herramientas para jugar con él y transformarlo y experimentar a nuestro gusto. Por eso necesitamos que, cada una y cada uno de nosotros, entre todos, como sujetos, podamos decidir en cada momento qué hacer con ellos.

De la administración al gobierno de los datos

Hay varias disciplinas que históricamente se han preocupado de los datos. Ellas ejemplifican de muy buena manera los enfoques tradicionales hacia los datos. La *estadística* analiza datos. Usa datos como entrada y su tarea es estudiarlos,

¹ | Desde hace algunos años, las nociones “gobierno” y “gobernanza” tienden a oponerse: el término “gobierno” es reservado al viejo poder jerárquico, a la autoridad del Estado y a las concepciones centralistas, en tanto que, la “gobernanza” sugiere una nueva modalidad, horizontal, de la gestión del poder [1, 2].



analizarlos, etc. Sus productos estrellas son técnicas y métodos para recolectarlos de manera que tengan validez; para hacer inferencias a partir de ellos, etc. La disciplina de *bases de datos* históricamente se ha preocupado del almacenamiento, preservación, seguridad y disponibilización de los datos, y de la manera de consultarlos eficientemente. Sus productos estrellas son algoritmos y software para lograr eso. *Sistemas de información* (bibliotecología) se preocupa de estructurar los datos en su faceta semántica. Organizarlos y disponibilizarlos de acuerdo a las temáticas que involucran y los intereses de los usuarios. Finalmente, en *administración* (gestión) de datos, la idea es diseñar los procesos y flujos organizacionales para que la información circule de manera fluida en la organización o la empresa.

Todas ellas asumen los datos como datos, como recursos (naturales o producidos por los humanos), y trabajan a partir de su recolección. Como tradicionalmente los datos eran tratados como materias primas o como documentos de bóveda, había habido escasa o nula preocupación por el destino y el carácter de ellos. Los datos simplemente eran mudos y —literalmente— invisibles materiales de apoyo para el análisis y la toma de decisiones. Incluso muchas de las disciplinas de punta hoy, como *big data analytics*, *AI* y *machine learning* siguen esa lógica.

Cuando las personas, las redes sociales, el espacio público, las instituciones, las empresas, los gobiernos, comienzan a llenarse de datos (imágenes, sonido, videos, documentos, mensajes, señales de sensores, etc.), entonces comienza a aparecer el gran tema de cómo abordar nuestra relación con esas partículas que nos rodean e invaden. Y surgen las preguntas: ¿Quién es dueña o dueño de los datos y qué derechos le da esa propiedad? ¿Quién está autorizada o autorizado para recolectar esos datos y quién para usarlos? ¿Qué reglas existen para agregar a partir de, y usar, esos datos? ¿Cuáles son las reglas para los derechos sobre los productos de esos datos? etc.

Un dramático ejemplo de este hoyo negro es el que ocurrió con los datos de la pandemia. Ninguna de esas preguntas estaba clara. Es un caso de texto de ausencia completa de políticas de gobernabilidad de datos. Los ministros involucrados (Salud y Ciencia y Tecnología) entendían los datos del COVID como en el siglo XIX se entendía un reporte experimental, esto es, como un tesoro a resguardar del resto. No había ninguna claridad ni iniciativa sobre su destino, sobre quién y cómo accederlos, cómo enriquecerlos, cómo mejorar su recolección. Ningún esfuerzo de anonimización, y gran confusión sobre quiénes eran las y los titulares y qué derechos tenían sobre ellos. Suma y sigue. Ante este temor al vacío, se optó por mantenerlos encapsulados, embodegados, al margen de todo uso, salvo de quien en ese momento tenía el poder. (Y *bonus track* en el caso chileno: manipulación y engaño, aprovechando que nadie más tenía acceso a ellos).

Tener respuestas claras a esas preguntas y muchas otras significa tener políticas de gobierno (o gobernanza) de datos. Porque de aquí en adelante, los datos serán siempre parte de nuestro hacer. Nos interesa presentar aquí la dimensión pública del problema (el que más urge en mi opinión). Pero es importante aclarar: lo que desarrollaremos no importa sólo a aquel (hoy aún pequeño y extraño) grupo que se preocupa de los bienes públicos, de las políticas públicas. Estas preocupaciones que he indicado a nivel global, tienen carácter fractal: al interior de las organizaciones y empresas ocurre el mismo fenómeno, y a nivel de las personas también [3, 8].

Los diferentes niveles de gobierno de datos

Decíamos que, históricamente, las disciplinas que tenían que ver con datos estaban centralmente preocupadas de su gestión y administración. La diferencia crucial entre administración/gestión y gobierno, es la que hay entre medios y



Figura 1. Ciclo de los datos.

fines, entre el ¿cómo? y el ¿para qué? y ¿hacia dónde? Como tradicionalmente los datos fueron esencialmente medios para tomar decisiones, los fines se sustentaban. Ayer los datos sustentaban una investigación, un juicio, una afirmación; hoy los datos direccionan actividades humanas. Las compañías de redes sociales las usan para redireccionar a las personas y a la política; las compañías de juegos a la vida de los jóvenes; los gobiernos a los ciudadanos; las empresas a sus clientes; etc. Entonces se hacen relevantes las preguntas sobre qué hacer con los datos y quién lo decide, esto es, preguntas sobre su gobierno.

Antes de entrar en esas preguntas, recordemos esquemáticamente el ciclo de los datos, donde destacaremos cuatro hitos (ver Figura 1): generación o producción, almacenamiento, análisis y procesamiento, y utilización y consumo.

Estos hitos gruesos nos ayudarán a formular preguntas relevantes que surgen en el gobierno de datos y a reflexionar sobre quiénes debieran decidir sobre sus respuestas. Dejamos abiertos a la reflexión ambos temas, pues el objetivo de este artículo es simplemente despertar el interés de los y las lectoras por estos asuntos que hay que resolver en comunidad.



A. Datos personales (privacidad, propiedad, exposición)

Tradicionalmente, el acercamiento a los datos por parte del ciudadano común es el de un consumidor, especialmente de datos procesados en formato de información (o conocimiento en menor medida). Con el advenimiento de aparatos digitales diversos, comienza a recibir datos directamente (sin pasar por el sistema de significación) y hoy también se ha convertido en un productor de datos. El almacenamiento, que tradicionalmente era local (y luego subentendía la propiedad de los datos) comienza a ser un tema. Lo mismo el análisis, debido al uso de herramientas y software externos.

Entonces aparece la responsabilidad de gobernar, esto es, de dirigir ese flujo de recepción y envío de datos. Algunas preguntas que ayudan a conformar esas necesarias decisiones: ¿Qué datos produzco? ¿Qué datos consumo? ¿Qué y cómo interpreto datos que recibo? ¿Qué datos comunico? ¿A quién? ¿Qué nivel de exposición y privacidad quisiera? ¿Dónde, cuándo y cómo almaceno mis datos? ¿Qué propiedad tengo sobre ellos?, etc.

B. Datos en el espacio público

La creación de un espacio público (una plaza pública) de datos es algo que comienza a instalarse con la Web. Nuestra interacción a distancia con otras personas, con instituciones, con aplicaciones diversas, levanta los temas de privacidad, de propiedad, y de otros derechos. Ya no son sólo mis datos, sino que es mi "contribución" a estadísticas, a facetas del mundo.

Surgen las preguntas: ¿Cuáles son los temas de salud mental y conformación de identidad para los jóvenes? (tanto de la vigilancia del sistema como de su involucramiento en redes sociales). ¿Cómo proteger a las personas más vulnerables? ¿Cómo compatibilizar la necesaria vigilancia —por ejemplo contra el crimen— con la libertad de las personas? ¿Cómo regular mensajes políticos y la propaganda? Aquí es donde

La aproximación a los datos ha sido hasta hoy muy similar a la conquista de las riquezas de América, el rush del oro en California, o el petróleo en el medio oriente. Una carrera por un bien "natural" que está ahí y hay que "explotar".

se ajusta de muy buena manera el enfoque de los comunes (*commons*) para estos datos. Son espacios que necesitamos cuidar. Como escriben Heiss y Ostrom:

"Las preguntas esenciales de cualquier análisis de comunes son inevitablemente sobre equidad, eficiencia y sustentabilidad. La equidad refiere a asuntos de apropiación justa o igual, y contribución a, el mantenimiento de un recurso. La eficiencia habla de la producción, gestión y uso óptimo de los recursos. La sustentabilidad mira los resultados sobre el largo plazo." [4]

C. Gobierno de datos en empresas y organizaciones (gestión, gobernanza, socialización, ética)

Éste es un tema que ha generado mucha literatura en la última década. En las empresas es donde el nuevo rol que juegan los datos aparece más ligado a lo inmediato. En esta área es donde se acuña también la expresión "gobernanza" por oposición a "gobierno" (que sugiere algo más jerárquico), indicando con esto que las partes interesadas e involucradas son muchas, pues los datos están en todas partes (ésta es su novedad, que establece su complejidad, valor y genera los riesgos).

La gobernanza de datos en la empresa abarca decisiones sobre: la gente que es responsable por la calidad de los datos; las políticas y procesos asociados con la recolección, gestión, almacenamiento y reporte de los datos; y por supuesto, los sistemas tecnológicos que apoyan y proveen infraestructura para ellos. Aparecen en esta lógica cinco grandes dominios de toma de decisión en TI: principios; arquitectura; infraestructura; necesidades de aplicaciones comerciales; e inversión y

priorización. Éstos llevan a cinco dominios interrelacionados de toma de decisión: principios de datos, calidad de datos, metadatos, acceso de datos y el ciclo tradicional de datos [5, 6].

D. Gobierno de datos a nivel país (economía, política, soberanía, ética)

Éste es un tema muy reciente. Seguimos el enfoque del Centro Internacional para la Innovación en la Gobernanza de Canadá [7]. Allí se plantea que cualquier estrategia nacional de datos debe abordar tanto las dimensiones económicas como las no-económicas del aprovechamiento de los datos masivos (*big data*). Esto involucra numerosos objetivos que deben balancearse. Entre ellos: cosechar las ganancias de las potencialidades económicas de los datos; respetar, y realzar, sus elementos de privacidad fundamentales; preservar una sociedad abierta y la democracia; mantener la seguridad pública; desarrollar las instituciones (como redes de información y procesos de gobernanza) que mantienen o realzan la identidad del país; y preservar la soberanía del país (ver Figura 2).

E. Gobierno de datos internacional

Éste es un tema de alta relevancia que comienza a surgir con los conflictos entre Estados y grandes compañías. Pero eso es sólo la punta del iceberg de este nuevo espacio que se está instalando, donde realmente se juega el fondo de lo que estamos tratando. Los datos tienen la capacidad de sortear fronteras materiales, y de desestructurar el tiempo tradicional. Eso los hace muy apetitosos para algunos (las compañías transnacionales y el capital global) y muy complejos para los Estados tradicionales.

Cualquier estrategia [de gobierno de datos a nivel país] debe abordar tanto las dimensiones económicas como las no-económicas del aprovechamiento de los datos masivos.



Figura 2. Algunas de las principales dimensiones de un marco de datos país.

Fuente: Jim Balsillie.

A manera de conclusión

La metáfora que los datos son el nuevo petróleo es muy limitada. El petróleo incide sobre el medio ambiente; los datos son parte del ecosistema humano, luego permiten incidir sobre el humano mismo directamente. Como escribe R. Medora, los datos “ejercen una única, perdurable y potente influencia sobre las vidas individuales, sobre las relaciones sociales, sobre la sociedad y el país” [7].

Por ello la responsabilidad sobre los datos no termina nunca. Comienza con las definiciones sobre su recolección; sigue con las maneras, lugares y tiempos de almacenamiento; continúa con los alcances de su agregación, procesamiento y análisis; y vuelve al ciclo con sus usos, transferencias y exposición. Así los datos redefinen el ambiente, nuestro hacer, crean un mundo nuevo.

Un mundo nuevo, creado por nosotros, que hay que aprender a gobernar si no queremos que quienes tienen hoy poder sobre el ciclo de los datos lo hagan a su manera, tomando sobre nuestras vidas decisiones que corresponden determinarlas soberanamente a cada país, a cada comunidad, a cada persona. ■

REFERENCIAS

- [1] Dany-Robert Dufour. *Gobernanza Versus Gobierno*. Cuadernos de Administración, Universidad del Valle, no. 41, enero-junio, 2009.
- [2] Renate Mayntz. “From Government to Governance: Political Steering in Modern Societies”. *International Summer Academy, From Government to Governance: The Case of Integrated Product Policy*. Alemania. Septiembre, 2003.
- [3] Anne Young y Kevin McConkey. *Data governance and data quality: Is it on your agenda?* Journal of Institutional Research, vol. 17, no. 1, p. 69–77. 2012.
- [4] Charlotte Hess y Elinor Ostrom. *Understanding Knowledge as a Commons From Theory to Practice*. Introduction, p. 6. MIT Press. 2006.
- [5] Vijay Khatri y Carol V. Brown. “Designing data Governance”. *Communications of the ACM*, vol. 53, no. 1. Enero, 2010.
- [6] Ibrahim Alhassan, David Sammon y Mary Daly. “Data Governance Activities: An Analysis of the Literature”. *Journal of Decision Systems*. vol. 25, no. 1, p. 64–75. 2016.
- [7] Rohinton P. Medhora. “Data Governance in the Digital Age”. Centre for International Governance Innovation. *Data Governance in the Digital Age*. www.cigionline.org. p. 2–5. 2018.
- [8] Rene Abraham, Johannes Schneider y Jan vom Brocke. “Data Governance: A Conceptual Framework, Structured Review, and Research Agenda”. *International Journal of Information Management*. vol. 49, p. 424–438. 2019.