

## Ciencia de datos para la innovación

---

**Autores:** Alvear Leyton, Alexis\*; Pérez, Nancy V.

**Contacto:** \*[aalvear@mat.uc.cl](mailto:aalvear@mat.uc.cl)

**País:** Chile

### Resumen

El desarrollo científico y tecnológico ha favorecido el acceso a más y mejor información. La disponibilidad de dispositivos tecnológicos hace posible medir múltiples variables en tiempo real y procesar grandes volúmenes de información mediante algoritmos computacionales para facilitar la toma de decisiones. Todo este fenómeno se conoce como “*Big Data*”.

Por su parte, la ciencia de datos o “*data science*”, se ocupa de analizar, simular, crear modelos y patrones para comprender el comportamiento de fenómenos complejos, mediante el uso de diversas herramientas matemáticas y estadísticas. De esta forma, el Big Data transforma los datos en información y Data Science transforma la información en conocimiento.

La aplicación combinada de estas técnicas favorece el desarrollo de esta disciplina en diversos campos: permite comprender el comportamiento de diversos fenómenos y de esta forma generar conocimiento relevante. Este nuevo conocimiento, a su vez, permite explorar nuevas ideas y soluciones, y mediante la interdisciplina, desarrollar innovaciones y con la ayuda de la tecnología, tomar mejores decisiones.

En este contexto, a través de la experiencia de una institución de educación superior latinoamericana, se ha podido observar cómo abordar este desafío a través del desarrollo proyectos de innovación y transferencia tecnológica que, mediante análisis científico de datos y su puesta en valor, ha creado diversas plataformas que le han permitido al Estado, a la academia, a las empresas y a la población en general, acceder a estudios especializados, proyecciones, análisis y seguimiento de indicadores para apoyar la descentralización, mejorar la eficiencia de las empresas, aumentar el crecimiento económico, impulsar políticas públicas focalizadas y mejorar la calidad de vida de los habitantes.

**Palabras clave:** inteligencia competitiva; vigilancia tecnológica; datos abiertos; descentralización; big data; data science; democratización de los datos.

## 1. Introducción

### 1.1. La nueva era del big data

Desde los inicios de nuestra civilización, la humanidad ha consignado los datos a través de diversos registros con el fin de almacenar la información y preservar el conocimiento. Diversas investigaciones han encontrado evidencias de estos registros desde la prehistoria, con los cuales las diferentes culturas transmitieron la información a través de registros pictográficos, como los papiros encontrados en el antiguo Egipto, o en la cultura maya, que, a través de códigos tallados en piedra, plasmaron antecedentes sobre su cultura, creencias y conocimientos que hoy son investigados por la arqueología.

Con el desarrollo de la escritura, la humanidad logró aumentar la capacidad de registrar y almacenar la información, y más recientemente, en el siglo XX, el registro de los datos se fue masificando, primero en libros, como los de actas que se utilizaban en las oficinas del Registro Civil de Chile en los años 60, en los

que se consignaban los nacimientos y defunciones, para luego, con la llegada masiva de la computación en la década de los 80, facilitar los registros digitalizándolos en bases de datos.

Actualmente, gracias al desarrollo tecnológico, los datos se registran, almacenan, procesan y comunican minuto a minuto. Por ejemplo, cada vez que se realiza una transacción comercial con una tarjeta de crédito o débito, la información inmediatamente se registra en la cuenta y se comunica al cliente a su teléfono móvil a través de un mensaje de texto o de la aplicación bancaria, indicando todos los datos de dicha compra: lugar, monto, fecha y horario en que se efectuó. Y así como las transacciones comerciales, hay millones de datos registrándose y procesándose al instante, a través de los más diversos dispositivos y aparatos tecnológicos, sensores y señales que circulan en todo el planeta.

Toda esta capacidad de almacenamiento y procesamiento, consecuencia del desarrollo científico y tecnológico, ha dado origen a un fenómeno conocido como “big data”, o grandes volúmenes de datos en español, que está multiplicando, tanto en términos de volumen como en velocidad, los registros de la información.

## 1.2. La dimensión del big data

Ante lo descrito surge de inmediato la pregunta: ¿cuál es la magnitud del volumen de datos cuando hablamos de big data? Un antecedente que sirve para dimensionar el impacto del big data lo aportó Eric Schmidt, ex Director Ejecutivo de Google, en la conferencia “Techonomy” efectuada en Lake Tahoe, California, Estados Unidos el año 2010. En la conferencia, Schmidt indicó que, desde los inicios de nuestra civilización, hasta el año 2003, la humanidad había generado, aproximadamente, un total de 5 exabytes de información (equivalentes a un millón de terabytes), mientras que hoy, generamos esa cantidad de información cada dos días y esta capacidad se multiplica cada vez más (Schmidt, 2010).

Lo cierto es que, gracias a todo este desarrollo, hoy podemos acceder a información que nunca imaginamos tener, y nos permite comprender de mejor forma los fenómenos que nos rodean. Sin embargo, también nos ha generado un nuevo problema. Si antes el desafío era acceder a datos, hoy nos enfrentamos a un verdadero mar de información, exceso de datos que superan nuestras capacidades de discreción y selección, y que nos expone a problemas como el uso de datos no fiables (fake news), desactualizados, incompletos, de difícil comprensión, entre otros aspectos, para lo cual necesitamos de técnicas avanzadas capaces de analizar estos datos en toda su complejidad para así obtener conocimiento relevante.

## 2. Desarrollo

### 2.1. Definición del problema u oportunidad

#### 2.1.1. El big data y la complejidad en la toma de decisiones

Por su definición en inglés, se tiende a relacionar el big data con el desafío de trabajar con grandes volúmenes de información. No obstante, el big data no sólo trata de volumen, sino también de complejidad, la que se expresa en 5 características que se conocen como las “5 V” del big data.

FIGURA 1. Las “5 V” del big data



Fuente: Elaboración propia.

La primera “V” es la de volumen, que como hemos señalado, es la que refiere a la dimensión y carga de los datos con los que estamos trabajando. Esta característica implica un desafío técnico y metodológico. Técnico, porque requiere de capacidades de infraestructura y arquitectura tecnológica que den soporte a la captura, almacenamiento y procesamiento de grandes volúmenes de información, y metodológica para abordar los procesos de análisis de información en grandes magnitudes.

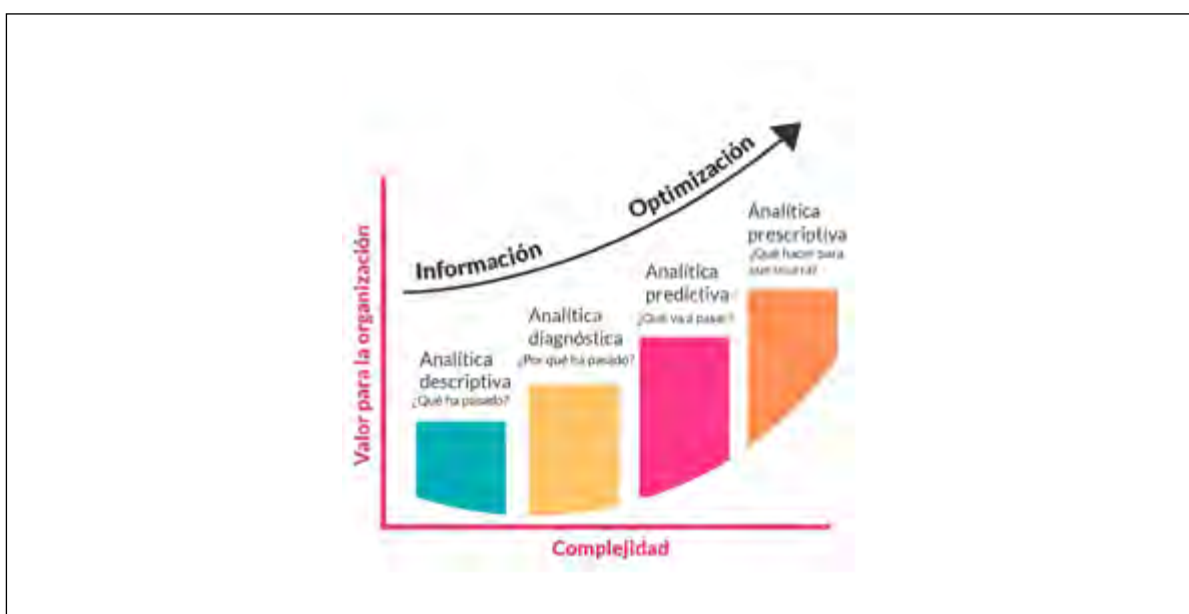
La segunda “V” es la de variedad, que tiene relación a que en big data se trabaja con diferentes tipos y estructuras de datos. Los datos pueden provenir de fuentes estructuradas, que tienen un marco definido. Por ejemplo, una ficha médica de un paciente tiene una estructura definida en el momento que fue diseñada, y que recoge datos del paciente en campos predeterminados, como el nombre, que tiene una configuración alfabética con una determinada extensión (apellido paterno, materno y nombres), una dirección que se compondrá de campos alfanuméricos, un número telefónico de contacto o el DNI, con campos numéricos, etc. Así también, podemos trabajar con datos no estructurados, que no cuentan con ese estándar predefinido. Por ejemplo, las imágenes constituyen datos no estructurados. Si queremos contabilizar cuántas personas transitan por una calle céntrica de la ciudad, recurriremos a los registros obtenidos por las cámaras de video de esa arteria. Las cámaras son capaces de obtener imágenes de un cuadro en donde circulan personas, las que no siguen un patrón predeterminado, ya que pueden hacerlo por diferentes partes de ese cuadro. Algunas personas transitarán caminando despacio, otros más rápido, otros se detendrán a mirar una vitrina, quizás niños transiten corriendo o jugando. Además, las personas que aparezcan en cuadro probablemente tendrán diferentes características fisonómicas. El desafío aquí es que la computadora logre identificar adecuadamente cuando se trate de una persona, y que contabilice de forma única su aparición de forma adecuada.

La tercera “V” del big data es la de velocidad, que tiene relación con los tiempos asociados a la captura y procesamiento de los datos. En esta era en donde los dispositivos tecnológicos como sensores y las redes disponibles nos permiten registrar datos en tiempo real, es muy importante desarrollar la capacidad de trabajar con esos datos en el momento en que se generan, ya que el uso de datos desactualizados ensucia los modelos de análisis y nos podría llevar a establecer conclusiones inadecuadas de un determinado fenómeno. Así como un médico para proceder a una cirugía requerirá de exámenes recientes, un científico de datos también debe contar con información lo más reciente posible para configurar buenos modelos.

La cuarta “V” es la que refiere a la veracidad de los datos. Con esto nos referimos a la calidad de los datos con los que trabajamos. En términos concretos, esto se refiere a 3 aspectos: 1) fuentes de origen: determinar que los datos que se utilicen provengan de una fuente que garantice fiabilidad en la información que se va a utilizar, que éstos sean actualizados, que la fuente sea oficial y que tenga un adecuado procedimiento de captura; 2) estructura de los datos de origen: muchas veces nos encontramos con datos mal digitados desde su origen, o con registros incompletos porque los formularios de captura están mal diseñados. En los proyectos de ciencia de datos, la mayor parte del tiempo de trabajo se destina a la limpieza y depuración de las bases de datos, por lo que apoyar los procesos de diseño de los mecanismos que se usarán para la obtención de información siempre será una buena inversión de tiempo. Y finalmente, 3) representatividad de los datos: el uso de datos que en hayan considerado técnicas de muestreo representativo que permitan establecer conclusiones relevantes del fenómeno que se está analizando. En estadística existen diversas técnicas metodológicas que favorecen el trabajo de muestreo para garantizar una adecuada representatividad y técnica de análisis de datos faltantes o en áreas pequeñas que permiten reducir los sesgos de selección.

Finalmente, la última “V” del big data es la de valor. Y la forma de obtener valor de los datos es a través de la ciencia de datos, herramienta que, mediante la aplicación de técnicas de analítica avanzada, en diferentes niveles de complejidad, permite desarrollar procesos de análisis para sacar el máximo provecho de los datos.

FIGURA 2. Jerarquía de complejidad analítica



Fuente: Elaboración propia.

Es así como en data science encontramos 4 niveles de complejidad analítica, los que van avanzando de forma gradual. Mientras mayor complejo sea el modelo analítico, mayor valor entregará a la organización. Los niveles son:

1. **Analítica descriptiva:** El primer nivel de complejidad es el descriptivo, que permite explicar un fenómeno a partir de la composición y caracterización de sus variables de origen. De esta forma, las técnicas estadísticas de análisis exploratorio y descriptivo de datos nos permiten “contar una historia”, y describir un hecho acontecido. Por ejemplo, discutir respecto de las matrículas de una carrera de pregrado en una universidad, los datos nos van a revelar cuántos alumnos matriculados hay por año, la distribución de ellos por género, edades, colegios de procedencia, niveles socioeconómicos, entre otros aspectos que describen el fenómeno que queremos estudiar.

2. **Analítica de diagnóstico:** Un segundo nivel de complejidad analítica es la de diagnóstico, que permite explicar el fenómeno bajo estudio a partir de la identificación de las variables que indican en su comportamiento. En este caso, el uso de herramientas de inferencia estadística permitirá establecer correlaciones entre diferentes variables. Por ejemplo, en el caso de una carrera de pregrado, las tasas de deserción de alumnos en primer año de carrera podrían presentar correlación con los puntajes de ingreso o los establecimientos educacionales de los cuales provienen dichos estudiantes.

3. **Analítica predictiva:** El tercer nivel de complejidad analítica es la predictiva, que se basa en el uso de técnicas estadísticas que permitan pronosticar el comportamiento futuro de determinadas variables a partir del análisis de su desempeño histórico. De esta manera, el uso de métodos estadísticos como los modelos de regresión o las series de tiempo, permitirán identificar patrones en las variables, detectar la frecuencia con que se presentan dichos comportamientos, identificar factores de estacionalidad o incidencia de factores aleatorios que permiten proyectar este fenómeno hacia el futuro. En el ejemplo del análisis del comportamiento de las matrículas de una carrera de pregrado, se podría proyectar en el futuro la composición de las cohortes, estimar matrícula de nuevos alumnos o anticipar indicadores de retención dado los comportamientos históricos.

4. **Analítica prescriptiva:** Finalmente, el cuarto nivel de complejidad analítica es la prescriptiva, que consiste en anticipar el comportamiento de un fenómeno dado el funcionamiento de determinadas variables que inciden en éste. De esta forma, la analítica prescriptiva, que viene de “pre-escribir” la historia, permite simular diversos escenarios, estableciendo la probabilidad de ocurrencia de cada uno de ellos. Por ejemplo, si determinamos acciones para favorecer la matrícula de estudiantes de género femenino en carreras científicas, y habiendo ya establecido previamente la correlación de dicha variable con ese resultado, se podrá simular el resultado de esa acción en admisiones futuras. Para los procesos de analítica prescriptiva, se utilizan algoritmos y modelos matemáticos de nivel avanzado, también conocidos como técnicas de machine learning (aprendizaje automático de máquinas), que son la base de la inteligencia artificial, avance tecnológico muy discutido en la actualidad.

### 2.1.2. Big data y data science: disciplinas que ayudan a reducir la complejidad

Pero, concretamente, ¿para qué nos sirven estas disciplinas? Anteriormente habíamos señalado el problema que representa el estar expuestos a un mar de datos, un verdadero tsunami de información que más que ayudarnos a comprender las cosas, las complica. De todos los datos que tengo, ¿cuál me sirve?, ¿cuál considero fidedigno?, ¿cómo lo interpreto? El principal desafío del big data es ayudarnos a tomar decisiones

con el apoyo de sistemas computacionales que, siguiendo lógicas humanas, nos ayudan a reducir la complejidad de esos grandes volúmenes de información que superan nuestras capacidades de procesamiento.

FIGURA 3. Aplicaciones del big data



Fuente: Elaboración propia.

Para ilustrar de forma sencilla lo que la ciencia de datos puede hacer por nosotros, pondré de ejemplo un problema de big data al que nos enfrentamos cotidianamente. Es muy probable que muchos consumidores de televisión por cable realicen frecuentemente “zapping” con el fin de encontrar algo interesante para ver, y también es muy probable que, a pesar de tener una oferta muy diversa de contenido en más de 700 canales de televisión, muy pocas veces encontremos algo interesante para ver. Eso se debe a que nuestras capacidades humanas no nos permiten analizar, en tan poco tiempo, los datos sobre el contenido que están transmitiendo en los 700 canales y discernir si el contenido es interesante o no, por ende, tomamos malas decisiones. En cambio, es probable que muchas personas frente a esta disyuntiva prefieran plataformas como Netflix o Amazon, pues consideran que satisfacen de mejor forma sus intereses. Pues bien, este tipo de plataformas, mediante algoritmos de data science, recogen datos del consumo audiovisual de sus usuarios, sobre qué películas o series ven, qué actores, en qué días y en qué horarios son vistos, para así construir perfiles de comportamiento y hábitos de consumo de sus suscriptores, con el fin de construir una oferta personalizada de contenido que, generalmente, se transforma en maratones de horas frente al televisor.

Ejemplos como estos se replican en otras plataformas como Youtube, cuando nos sugiere ver videos que podrían ser de nuestro interés, Spotify, que crea playlists de canciones según lo que escuchamos frecuentemente, la publicidad en línea que nos “persigue” con ofertas de productos que alguna vez cotizamos o las tiendas del retail, que, en función del análisis de nuestras boletas de compra, nos envía e-mails con ofertas de productos y marcas que generalmente preferimos.

Es así como la ciencia de datos simplifica la complejidad del big data, procesando y analizando la información en función de lógicas humanas, ayudándonos en la toma de decisiones.

### 3. Resultados

#### 3.1. La democratización de los datos

Así como en los ejemplos anteriores, es probable que en muchas ocasiones de nuestras vidas nos enfrentemos a procesos de toma de decisiones complejas. ¿En qué colegio matricular a nuestros hijos?, ¿Qué carrera estudiar?, ¿En qué ciudad sería más conveniente vivir?, Si busco empleo, ¿qué cifra debo indicar como pretensiones de renta?, ¿Me conviene o no endeudarme en estos momentos?, ¿Por quién debo votar en las próximas elecciones?, entre otras muchas preguntas que surgen en nuestras vidas.

En todas estas interrogantes, hay diversas variables que influyen en la decisión, y que derivan de los fenómenos económicos, políticos y sociales que ocurren en el país, los cuales deben observarse de forma integral y con la perspectiva del tiempo, con el fin de proporcionar un contexto adecuado. Y es aquí donde los datos son muy relevantes, puesto que nos ayudan a comprender mejor el mundo en el que vivimos. Pero los datos no son estáticos, son dinámicos, las estadísticas nos muestran no solamente la fotografía del momento, sino que nos cuentan una película completa, los datos nos cuentan historias y nos permiten entender el por qué ocurren los fenómenos, cómo evolucionan e incluso anticipar el futuro.

Sin embargo, muchas veces todos estos datos no están siempre al alcance de la población, y los que están, son complejos de analizar e interpretar. Si bien los esfuerzos que se han implementado en Chile por transformar digitalmente al Estado y el fomento de políticas de transparencia de la información han contribuido en los últimos años a la mayor publicación de datos públicos, aún hay mucho por hacer en términos de políticas de datos abiertos, que garanticen el acceso a la información por parte de la población.

Democratizar los datos implica facilitar no solo el acceso, sino también la comprensión de la información que sea relevante para la toma de decisiones de los ciudadanos. En este contexto, no basta con publicar planillas con datos en los sitios web, sino que se debe trabajar en estandarizar los parámetros bajo los cuales se publican los datos, acompañar estas publicaciones con documentos explicativos que los pongan en contexto y describan la metodología utilizada para su análisis o captura, presentar datos que cuenten con una frecuente actualización y que tengan niveles adecuados de desagregación a nivel regional o comunal, entre otras muchas consideraciones técnicas.

En este contexto es donde se presenta una gran oportunidad, la de aprovechar todo el conocimiento sobre big data y el desarrollo de la ciencia de datos y ponerlo a disposición de la ciudadanía con el fin de democratizar el acceso a ellos, para apoyar a las personas en sus procesos de toma de decisiones, mejorar la eficiencia de las pequeñas y medianas empresas, aumentar el crecimiento económico, impulsar políticas públicas focalizadas y mejorar la calidad de vida de los habitantes del país.

El acceso a los datos, en la sociedad de la información, garantiza la participación de la comunidad en las decisiones políticas, económicas y sociales, un aspecto muy relevante que contribuye también a la paz social.

#### 3.2. Innovación basada en datos para la descentralización del país

La experiencia demuestra que los países más desarrollados son países descentralizados. Un país descentralizado otorga a las regiones la autonomía necesaria para decidir sus estrategias competitivas, pero junto con dotar a las regiones de poder y autonomía, se le deben brindar herramientas para que tomen esas decisiones, ya que requieren de información estratégica para ello.

### 3.2.1. Proyecto “Tarapacá Intelligence”

A partir del contexto anteriormente descrito, y sobre la base que el fenómeno de la descentralización se debe abordar como un problema complejo, la Pontificia Universidad Católica de Chile a través de DATA UC, desarrolló “Tarapacá Intelligence”, una plataforma de inteligencia competitiva y vigilancia tecnológica que mediante el uso de datos abiertos genera indicadores dinámicos sobre aspectos sociales, económicos, políticos y culturales de la región de Tarapacá, una región “extrema”, ubicada en el norte de Chile, y que enfrenta constantes desafíos de diversificación de su matriz productiva, relaciones geopolíticas con países limítrofes y el impacto de la migración en su composición social.

La plataforma se sustenta en un modelo conceptual que a través de la integración de diversas variables plantea el desarrollo regional como un problema complejo y aborda su tratamiento desde una perspectiva sistémica. Su estructura se basa en el proceso de la ciencia de datos, que luego de procesos de análisis transforma éstos en información relevante para la ciudadanía, y la información en conocimiento para la toma de decisiones.

Es así como se configuró como primera etapa, un amplio repositorio de datos regionales, de las más diversas fuentes de información, presentados de forma organizada a través de 7 áreas temáticas que dan cuenta de la evolución y estado actual de la región. Luego, y a través del uso de modelos matemáticos y estadísticos, se desarrollaron procesos de análisis de datos regionales y de variables externas que influyen directamente en resultados económicos, generando proyección de indicadores claves y detectando patrones de comportamiento de fenómenos económicos y sociales que dan respuesta a los desafíos de la región. Todo esto plasmado en el apartado “Intelligence” del sitio, que transforma la información en conocimiento.

De esta manera, Tarapacá Intelligence se configura como una plataforma “big data”, dado que procesa importantes volúmenes de datos de diversas fuentes para modelar resultados. Para su desarrollo, se utilizaron más de 100 bases de datos de fuentes públicas y privadas. Finalmente, y con el objetivo de generar conocimiento, la información se estructuró en base a un formato de “storytelling”, que permite contar historias para que los habitantes de la región en su conjunto puedan comprender con facilidad la información. La construcción de estos relatos fue gracias a la participación de los diferentes perfiles de usuarios en mesas técnicas que se organizaron en la región, instancias en donde se levantó información relativa a sus intereses y medios a través de los cuales accedían a los datos para tomar sus decisiones.

### 3.2.2. Proyecto “CEGIR” para la Región Metropolitana de Santiago

El año 2017 se aprobó en Chile una reforma constitucional que contiene la “Ley de fortalecimiento de la regionalización del país”, que creó una nueva institucionalidad para la administración de las regiones estableciendo a las Gobernaciones regionales como organismos autónomos de la administración del Estado, y definiendo mecanismos para la transferencia de competencias desde la Presidencia y la anterior figura de “Intendente Regional”, que fue eliminada. Una de las innovaciones más valoradas de esta ley es la creación de la figura de “Gobernador/a Regional”, quien dirige este nuevo órgano ejecutivo regional, el cual, a diferencia del ex “Intendente”, es elegido por votación popular, y la creación de un Consejo Regional, órgano colegiado que tiene como misión la aprobación de proyectos de inversión local, en los ámbitos de fomento productivo e industria, desarrollo social y humano, infraestructura y transporte.

El año 2021 se realizó por primera vez en Chile la elección de gobernadores regionales, quedando instalado, a contar del 14 de julio de dicho año, la institución de los Gobiernos locales en cada región del país. Como todo proceso de instalación, en el ejercicio del cargo se han identificado la necesidad de contar con



mayores atribuciones para las tareas que les fueron encomendadas, facultades duplicadas con otros organismos, necesidad de mayor coordinación con municipios y otros actores, entre otros aspectos.

Con el fin de dotar a los nuevos Gobiernos regionales de información que les permita monitorear el funcionamiento de la ciudad, actualmente está en desarrollo el proyecto CEGIR: Centro de gestión integrada regional, que de forma piloto será implementado en la región Metropolitana de Santiago, la más poblada del país. El objetivo de esta iniciativa es contribuir con la reducción de las brechas y asimetrías de información y democratizar el acceso a los datos para que los diferentes actores de la región tomen decisiones inteligentes y definan estrategias de desarrollo a partir de sus propias capacidades y potencialidades, proyectando un desarrollo sustentable y con énfasis territorial, aportando de esta forma a la descentralización de los órganos decisionales.

La disposición de información estratégica en la interfaz es de carácter personalizado para cada tomador de decisión, lo que constituye la principal propuesta de valor: la democratización de los datos para reducir las asimetrías de información y la generación de conocimiento colaborativo, favoreciendo principalmente a los actores regionales que contarán con información relevante para definir estrategias competitivas, focalizar la inversión pública y promover sectores económicos que aprovechen la diversidad productiva y el capital humano local, contribuyendo así a que las políticas públicas orientadas a la descentralización tengan mayor impacto.

El proyecto también contempla el desarrollo de un módulo denominado “Smart data ciudadano”, una aplicación a través de la cual los ciudadanos podrán conocer el desempeño de la región en las diferentes áreas económicas, observar datos sobre el empleo, remuneraciones, conocer datos sobre la evolución demográfica y examinar la eficiencia del gasto público tanto del Gobierno central como de los municipios. A su vez, la *app* oficiará como un canal de comunicación bidireccional entre el Gobierno Regional y los ciudadanos, para incentivar la participación ciudadana en decisiones que les afecte. De esta forma, se contribuye con el empoderamiento de la ciudadanía que en el contexto actual de la sociedad de la información que vivimos, no son sujetos pasivos, sino que requieren de participar y fiscalizar de forma crítica las decisiones que adopta la autoridad.

## 4. Discusión y análisis

### 4.1. Los desafíos de la transformación digital

Los constantes cambios tecnológicos, la incipiente transformación digital de la economía y la llegada de nuevos fenómenos como el big data, están obligando a los distintos actores de la sociedad a impulsar adaptaciones que les permitan sumarse oportunamente a esta nueva era tecnológica.

El proyecto “Tarapacá Intelligence”, fue pionero en generar un desarrollo tecnológico basado en big data como un bien público, beneficiando a diversos actores de una comunidad amplia como fue la región de Tarapacá. Pero, además, este proyecto ha sido uno de los pocos en Chile que ha contemplado un plan tan masivo de formación de capital humano, entregando herramientas técnicas para que la comunidad en el futuro obtenga el máximo provecho en el uso de esta plataforma, asegurando así la sustentabilidad de la iniciativa. Por su parte, el proyecto también ha demostrado que, para los Gobiernos regionales, el uso de la tecnología es un factor habilitante para impulsar cambios culturales y potenciar la innovación. Democratizar los datos también es una forma de generar igualdad, pues con esto se pueden reducir las asimetrías de información, permitir a los más pequeños competir en igualdad de condiciones con los grandes y a la ciudadanía en general estar más informada de lo que ocurre en su comunidad.

#### 4.2. Responsabilidad con aspectos éticos

Como en toda actividad humana, en la ciencia de datos también existen sesgos que pueden vulnerar los derechos de las personas, transgredir principios éticos y afectar a grupos en su desarrollo. Riesgos como el diseño de algoritmos con sesgos de género, culturales o económicos, el uso de datos personales sin consentimiento o la divulgación, aprovechando el masivo uso de redes sociales, de noticias falsas para incidir en comportamientos sociales (como lo fue el caso de Cambridge Analytics), nos obliga a los desarrolladores de esta disciplina a alertar sobre el correcto uso de estas técnicas, y contemplar en todo momento el resguardo de la información, la protección de los datos personales sensibles y poner siempre en el centro de todo desarrollo a las personas, sin perder el sentido humano de lo que estamos haciendo.

#### 5. Conclusiones

El mayor acceso a datos e información sin duda tendrá como consecuencia la multiplicación del conocimiento, y con ello, la mayor interacción de disciplinas y áreas del saber estudiando fenómenos complejos. Por otra parte, la democratización de los datos ha demostrado ser un factor habilitante para la innovación. Es así como, en los proyectos descritos, esta disciplina ha favorecido la constitución de espacios colaborativos que facilitan la innovación pública y social.

A partir de los hallazgos encontrados en los datos, se han generado nuevos proyectos, de carácter participativo con diferentes actores sociales y con perspectiva multidisciplinaria, que desde las realidades y necesidades regionales han permitido dar respuesta a problemáticas locales. Esta apertura de información en el largo plazo favorecerá la constitución de ecosistemas de innovación, con resultados exitosos en experiencias similares en países desarrollados, en donde los ciudadanos se sienten partícipes de las soluciones, se abren posibilidades para que start-ups o pequeñas y medianas empresas también participen de éstas, y en donde la Academia también es protagonista de los procesos de construcción regional.

Finalmente, también es importante consignar que el desarrollo de estas soluciones tecnológicas favorece la participación ciudadana y contribuyen con acciones de transparencia y accountability de parte del Estado. Lo anterior aporta con un clima social armónico e incentiva el interés de la ciudadanía por incidir en decisiones relevantes y en la participación política.

#### Referencias bibliográficas

- Schmidt, E. (2010). Tech + Economy, how technology can drive economic growth? *Techonomy*.
- Instituto Nacional de Estadísticas (2017). Censo 2017. *Estadísticas de población*. Instituto Nacional de Estadísticas, Chile.
- Dirección de Presupuestos (2019). *Presupuestos Regionales*. Ministerio de Hacienda, Santiago.
- University of Columbia (2020). *Data Science Institute*. <https://datascience.columbia.edu>
- Tarapacá Smart (2020). *Tarapacá Smart*. [www.tarapacasmart.cl](http://www.tarapacasmart.cl)
- Tarapacá UC (2020). *Tarapacá UC*. <http://tarapaca.mat.uc.cl>
- Banco Central de Chile (2018). Informe de Cuentas Nacionales. *Series trimestrales de estadísticas económicas*.
- Facultad de Matemáticas UC (2020). *Facultad de Matemáticas UC*. [www.mat.uc.cl](http://www.mat.uc.cl)
- Biblioteca del Congreso Nacional de Chile (2004). *Ley 19.175: Orgánica constitucional sobre Gobierno y Administración Regional*.
- Biblioteca del Congreso Nacional de Chile (2018). *Ley 21.074: Fortalecimiento de la regionalización del país*. Ministerio del Interior y Seguridad Pública.
- Comisión Europea (2015). Towards a thriving data-driven economy. *Official Journal of the European Union*.