Newsletter trimestral

CÁTEDRA

iDANAE

INTELIGENCIA · DATOS · ANÁLISIS · ESTRATEGIA

1T20

Data democratization

POLITÉCNICA

UNIVERSIDAD POLITÉCNICA DE MADRID

ManagementSolutions
Making things happen

Análisis de metatendencias

La Cátedra iDanae (inteligencia, datos, análisis y estrategia) en *Big Data* y *Analytics,* que surge en el marco de la colaboración de la Universidad Politécnica de Madrid (UPM) y Management Solutions, tiene el objetivo de promover la generación de conocimiento, su difusión y la transferencia de tecnología y el fomento de la I+D+i en el área de *Analytics*.

En este contexto, una de las líneas de trabajo que desarrolla la Cátedra es el análisis de las metatendencias en el ámbito de *Analytics*. Una metatendencia se puede definir como un ámbito de interés en un campo de conocimiento determinado que requerirá de inversión y desarrollo por parte de gobiernos, empresas y la sociedad en general en un futuro próximo[1].

En este informe trimestral se ha abordado la corriente de la democratización de los datos, tanto desde el ámbito público como del privado, abordando también la necesidad de la existencia de un gobierno de datos sólido.

[1]Para más información, véase la primera publicación de iDanae

Introducción

En los últimos años se ha producido un aumento masivo en la generación y recolección de datos en todos los ámbitos, tanto en el empresarial, como en la sociedad civil o la administración, al mismo tiempo que una mayor concienciación sobre la importancia de los mismos. Esta oportunidad para generar valor, junto con la aparición de herramientas que hacen posible la utilización y comprensión de los datos por usuarios no técnicos, ha provocado que exista una corriente cada vez más relevante a favor de la democratización de los datos.

Desde un punto de vista general, la democratización de los datos puede definirse como la generación y puesta a disposición de los datos a través de la participación y colaboración desde y entre distintos agentes, como instituciones públicas, empresas y ciudadanos, para su reutilización, bajo la premisa de que la disponibilidad y el análisis de datos permiten tomar decisiones más fundamentadas[2]. Esta democratización pasa necesariamente por hacer accesibles tanto los datos como las herramientas a usuarios con un perfil que podría ser no técnico, y sin requerir la involucración de usuarios técnicos o departamentos de IT que realicen funciones de intermediarios.

No obstante, tanto el propio concepto como el objetivo de la democratización presenta diferencias conceptuales, que deben analizarse, dependiendo de si el ámbito de tratamiento de los datos es el sector público o privado. Estas diferencias llevarán a distintas reflexiones, en función del ámbito analizado.

- En el ámbito del sector público, la democratización de los datos conlleva que el Estado tenga acceso a la información de sus ciudadanos, con el objetivo de mejorar su función de servicio, lo cual puede implicar grandes beneficios a la sociedad. No obstante, se han de equilibrar estos beneficios con el derecho a la privacidad de los datos, en especial los de carácter personal. Dentro de la perspectiva gubernamental resulta más interesante el proceso inverso: la democratización de los datos del gobierno para volverlos accesibles a sus ciudadanos. Algunas de las motivaciones que pueden justificar la democratización de los datos se sustentan en el aumento de la transparencia de sus procesos y órganos de gobierno, en la lucha contra la corrupción, o en el incentivo para que el ciudadano aumente su colaboración y participación, así como servir de base para las decisiones de los distintos agentes sobre el uso de recursos y espacios públicos.

- En el ámbito del sector privado, desde una perspectiva empresarial, se pueden plantear dos aproximaciones donde la democratización de los datos, su objetivo y su implementación son diferentes:

  - Una dentro de la propia empresa, en la que los datos recolectados por la compañía se comparten de manera transversal entre las áreas, evitando silos de la información. Esto permite que las áreas donde existen equipos de *analytics* puedan contar con datos adicionales a los que habitualmente tiene a su alcance, así como que otras áreas puedan incorporar en sus decisiones datos relevantes, enriqueciendo en ambos casos el conjunto de datos utilizado en la toma de decisiones.

  - Una externa a la propia empresa, donde la democratización de los datos ha de focalizarse en objetivos concretos que lleven a que un conjunto de empresas colabore compartiendo sus datos entre ellas o decidan hacer públicos sus datos con el objetivo de resolver retos sectoriales. Esta democratización, que puede ser intra-sectorial o global, debería implicar un beneficio para la sociedad. Un ejemplo de este caso puede verse en los acuerdos existentes para compartir

[2] Espinosa, Garriga, Zubcoff, & Mazón, 2014.

datos en la industria del automóvil con el objetivo de mejorar los sistemas de conducción autónoma[3]. Otro caso ilustrativo es la democratización de los datos en el ámbito de la ciberseguridad, con el objetivo de robustecer los marcos existentes.

No obstante, el éxito en la consecución de los objetivos de la democratización de los datos reside en la confianza entre los distintos agentes, tanto los que comparten los datos como quienes los procesan. En el ámbito del sector público, es necesario potenciar la confianza bidireccional entre los ciudadanos y el Estado; también en el ámbito del sector privado, donde este mismo concepto de confianza se ha de extender entre las distintas áreas de la empresa, o entre las distintas empresas.

Una vez superado el pilar de la confianza, será entonces precisa una implantanción eficiente de la democratización, donde se garantice una buena gestión, el acceso a herramientas tecnológicas adecuadas, y una gobernanza sólida, tanto de los procesos como de los datos.

Desde un punto de vista técnico, tanto en el ámbito del sector público como del privado, las necesidades se circunscriben a la existencia de herramientas adecuadas que permitan la interpretación de los datos por parte de usuarios de diferentes perfiles, así como a la estandarización de la información, su acceso y el almacenamiento por parte de terceros. Las preocupaciones a las que se enfrenta una compañía residen en la posibilidad de interpretaciones erróneas de los datos por parte de equipos no especializados, los riesgos en la seguridad, integridad y confidencialidad de los datos o en las duplicidades y redundancias en los análisis realizados por diferentes equipos frente a un enfoque centralizado, así como en el uso de algoritmos y modelos cuya aplicación puede ejecutarse de forma errónea.

Por su parte, la sociedad en su conjunto, en ocasiones con apoyo gubernamental, ha tomado un rol activo dentro de esta corriente asumiendo un cometido supervisor frente a empresas y gobiernos, con iniciativas como el *Open Data Barometer*[4], auspiciado por la fundación World Wide Web; y el *Open Data Charter*[5], formada por 73 países y 53 organizaciones, cuyo objetivo es establecer principios que sirvan como guía para compartir información, así como evaluar el grado de acceso a la información que presentan distintos organismos.

[3]Waymo, 2020.
[4]Open Data Barometer, 2013.
[5]Open Data Charter, 2013.

Democratización de los datos en el sector público

Existen distintos ejemplos de democratización de datos en el sector público. Uno de los más relevantes es el establecimiento en el año 2014 por parte del G20 de unos principios de un gobierno open data[6], cuyo objetivo principal es el de luchar contra la corrupción, fundamentalmente a través de los puntos siguientes:

- ***Open data* por defecto:** promueve la publicación de datos sin ser necesario una solicitud previa de información. Esto facilita el acceso en condiciones de igualdad para todos los usuarios.

- **A tiempo y completo:** promueve la publicación de conjuntos de datos completos, es decir, que sean precisos, recientes y actualizados. Fomenta también la publicación de los datos a un nivel desagregado, adecuadamente documentado y que siga los estándares acordados internacionalmente.

- **Accesibles y utilizables:** promueve la reducción de barreras de entrada innecesarias y la publicación de los datos en portales centrales, facilitando la utilización de los datos por un mayor número de ciudadanos y organizaciones.

- **Comparables e interoperables:** promueve la comparación y la trazabilidad de los datos de numerosos sectores relacionados con la lucha contra la corrupción.

Estos principios permiten, entre otros:

- La mejora de la gobernanza y aumento de la participación ciudadana: se empodera a los ciudadanos y se les permite pedir explicaciones a las instituciones gubernamentales. También les ayuda a entender, influir y participar directamente en los procesos de toma de decisiones y en el desarrollo de políticas públicas.

- El desarrollo e innovación inclusivos, que promueven mayores beneficios sociales y económicos al proporcionar información de calidad para crear instituciones eficaces y responsables. Los datos abiertos, además de prevenir la corrupción, facilitan el desarrollo de nuevos conocimientos, modelos empresariales y estrategias de innovación digital. Por ejemplo, se estima que los datos publicados por el gobierno de los Estados Unidos sobre la meteorología a lo largo del país genera más de 30 000 millones de dólares anuales, mientras que los datos de GPS generan más de 90.000 millones de dólares anuales[7].

En España, la iniciativa de abrir los datos a los ciudadanos surge en 2009[8] con el proyecto Aporta, cuyos objetivos básicos son:

- Fomentar una cultura favorable a la apertura de datos públicos.

- Facilitar que las administraciones acometan dicha apertura.

- Impulsar el mercado de la reutilización de la información pública.

En el año 2011, y gracias a estas iniciativas previas, el Gobierno de España lanzó una web[9] con datos abiertos recolectados por la Administración, con alrededor de 300 iniciativas y decenas de miles de conjuntos de datos. En paralelo también se lanzó una iniciativa análoga a nivel europeo[10].

[6] G20, 2015.
[7] Gurin, 2014.
[8] Gobierno de España, 2011a.
[9] Gobierno de España, 2011b.
[10] European Union, 2012.

Democratización de los datos en el sector privado

Tradicionalmente, las áreas que han llevado a cabo procesos de *analytics* eran las que contaban con mandatos específicos y realizaban esta función de manera aislada al resto de áreas, incluso de otras con funciones analíticas, contando con el soporte de áreas de IT para garantizar cuestiones como la calidad y el *time-to-market.* Sin embargo, en la actualidad se han desarrollado numerosas herramientas que posibilitan una democratización de los datos, facilitando que se puedan compartir y analizar los datos por diferentes áreas de una entidad o entre varias compañías.

Existen dos tipos de herramientas que posibilitan estos análisis: por un lado, aquellas centradas en la mejora del almacenamiento, la recolección y el tratamiento de datos; por otro, aquellas centradas en la visualización y comparación de los datos, la generación de modelos y el contraste de hipótesis.

Sobre las primeras, destacan los softwares de almacenamiento en *cloud*, ya que permite almacenar los datos de distintas áreas conjuntamente, así como los softwares de virtualización, que permiten al usuario recolectar y manipular datos a través de herramientas funcionales con el objetivo de eliminar datos inconsistentes, realizar análisis sobre la calidad y completitud de estos, y ejecutar transformaciones sobre las variables. Actualmente, el 94% de las empresas utilizan servicios basados en *cloud*[12], llegando a suponer un mercado global de más de 620.000 millones de dólares para 2023[12].

Respecto a las segundas, existen diversos productos de *business intelligence,* que permiten realizar análisis sobre los datos e interpretarlos sin tener formación en ámbitos *data science,* gracias a la utilización de procedimientos estandarizados y sencillos de ejecutar. El uso de herramientas de *business intelligence* supondrá un mercado de más de 26.000 millones de dólares para 2021[13].

Caso de estudio: Airbnb y Uber

Las empresas cuya propuesta de valor se apalanca en la recolección y análisis de datos son aquellas donde las corrientes de data democratization han surgido antes, tanto por las ventajas competitivas que suponen como por la cultura empresarial basada en el intercambio de datos. Airbnb es un ejemplo de lo anterior, ya que están llevando a cabo numerosas iniciativas tanto de forma externa como interna. Dentro de estas iniciativas destacan las soluciones a las que han llegado para solventar los problemas mencionados anteriormente, basadas en el desarrollo de los siguientes puntos:

1. Accesibilidad a los datos a través de una infraestructura de datos estable y escalable.

2. Herramientas para tratamiento diseñadas tanto para tener una mayor potencia y obtener mayor información de los datos por parte de los *data scientists* como para ser explotadas por parte de los trabajadores con suficientes conocimientos.

3. Formación para los empleados, de forma que todos los trabajadores tengan conocimiento suficiente de la infraestructura y las herramientas como para ser capaces de tomar decisiones basadas en los datos. Para ello, han creado una universidad del dato[14], que funciona de forma interna y a la que cualquier empleado de la compañía puede inscribirse.

Un enfoque distinto es el que ha tomado Uber. Para ello, inicialmente estableció una organización centralizada de datos responsable de la plataforma de datos de la compañía, de forma que se garantizase la consistencia y la eficiencia, ya que anteriormente cada equipo realizaba sus propios análisis. Una vez conseguida esa centralización de los datos, Uber desarrolló dos enfoques, uno orientado a usuarios técnicos y otro a no técnicos. Para los usuarios técnicos se desarrolló una plataforma de autoservicio basada en SQL, mientras que para los usuarios no técnicos se decidió establecer un equipo de DataOps y dedicación a tiempo parcial por parte del equipo de datos de la compañía, de forma que pudiesen mejorar la plataforma de autoservicio y apoyar a los usuarios no técnicos en su uso[15].

De forma externa, Uber lanzó en 2017 el *Uber Movement*[16], donde ha publicado información anonimizada de más de diez mil millones de viajes. El propósito es el fomento de la colaboración investigadora, con objetivos que van desde la planificación urbana (para adaptar la infraestructura existente y realizar futuras inversiones) hasta la reducción de atascos y emisiones o aumentar la seguridad en la conducción.

[11] RightScale from Flexera, 2019.
[12] MarketsandMarkets, 2019.
[13] MarketsandMarkets, 2017.
[14] AirBnB, 2017.
[15] Sarma & Thusoo, 2017.
[16] Uber, 2017.

Gobierno del dato

Si bien existen diferencias en el enfoque seguido por las distintas organizaciones en sus iniciativas de democratización de datos, en el núcleo de todas las estrategias se encuentra la existencia de un gobierno de datos sólido, que garantice un control y una calidad lo suficientemente alta como para cumplir con las obligaciones, tanto regulatorias como las que permiten optimizar la eficiencia de los procesos internos.

Muchas compañías tienen una experiencia reducida a la hora de desarrollar un gobierno del dato a lo largo de toda la organización, ya sea por falta de coordinación o por la inexistencia de normativa interna que establezca los roles, las funciones y las responsabilidades, resultando en fracasos a la hora de implementarlo. La existencia de una elevada manualidad en la generación de información, la fragmentación de sistemas *front* o carencias en la completitud, actualización y coherencia de la información, también suponen obstáculos a la hora de establecer una cultura de datos en la compañía. Para poder realizar una buena implantación de un gobierno del dato junto con el establecimiento de una estrategia de datos, existen varios pilares que deben considerarse que sintetizan los pasos a seguir:

- Establecimiento de un marco y un gobierno que se encargue del desarrollo de la normativa interna, de la creación de la estructura organizativa y de las nuevas funciones y roles (CDO[17], DPO[18], DDM[19], *Data Owner*[20], *Data Steward*[21], etc.), que establezcan la estrategia del dato, así como la protección de los datos.

- Establecimiento de una estructura y de unos metadatos, mediante la creación de un diccionario de datos (definición de la estructura y su contenido), el establecimiento de una trazabilidad funcional y técnica, la garantía de consistencia y coherencia de los datos entre entornos, y la implantación de las herramientas que soporten lo anterior.

- Establecimiento de planes de calidad de datos, mediante la creación de KPIs, controles de calidad, planes de remediación y análisis de consistencia.

[17] *Chief Data Officer,* responsable del marco completo de gobierno de la información.
[18] *Data Protection Officer,* responsable de la protección de los datos, en particular ligado a reglamentos de protección de datos, como por ejemplo la regulación europea GDPR.
[19] *Data Domain Manager,* interlocutor de IT para ámbitos específicos de los datos.
[20] Responsable de la definición y validación conceptual de los datos y su caracterización.
[21] Responsable operativo de un determinado dominio con un conocimiento concreto del dato y sus usos.

Conclusiones

Dentro de la revolución que está suponiendo para la industria la generación y la recolección de los datos, con las compañías desarrollando numerosas estrategias para convertirse en empresas *data-driven,* la democratización de los datos supone un paso más a la hora de generar valor, nuevas oportunidades y ventajas competitivas. La consecución de estos objetivos se basa en un principio de confianza mutua entre los agentes que intervienen en el proceso de democratización. Sobre esta base, el hecho de compartir los datos, ya sea dentro de una misma compañía o entre diversas compañías con un objetivo común, supone un aumento en la flexibilidad y en la inmediatez de la realización de análisis que permitan mejorar el proceso de toma de decisiones. Sin embargo, también conlleva diversos riesgos como puede ser la seguridad de los datos o la duplicidad de los análisis realizados. Es por esto por lo que se vuelve imprescindible un gobierno de datos sólido y transversal en toda la entidad que garantice tanto el cumplimiento regulatorio como la eficiencia de los procesos.

Bibliografía

AirBnB. (2017). How Airbnb Democratizes Data Science With Data University. Retrieved from How Airbnb Democratizes Data Science With Data University: https://medium.com/airbnb-engineering/how-airbnb-democratizes-data-science-with-data-university-3eccc71e073a

Barometer, O. D. (2013). Open Data Barometer. Retrieved from Open Data Barometer: https://opendatabarometer.org/

Charter, O. D. (2013). Open Data Charter. Retrieved from Open Data Charter: https://opendatacharter.net/

Espinosa, R., Garriga, L., Zubcoff, J. J., & Mazón, J.-N. (2014). Linked Open Data mining for democratization of big data. IEEE.

European Union. (2012). EU Open Data Portal. Retrieved from EU Open Data Portal: https://data.europa.eu/euodp/en/about

G20. (2015). Introductory note to the G20 anti-corruption open data principles. Retrieved from Introductory note to the G20 anti-corruption open data principles: http://www.g20.utoronto.ca/2015/G20-Anti-Corruption-Open-Data-Principles.pdf

Gobierno de España. (2011a). Datos Gobierno de España. Retrieved from Datos Gobierno de España: https://datos.gob.es/es

Gobierno de España. (2011b). Iniciativa de datos abiertos del Gobierno de España. Retrieved from Iniciativa de datos abiertos del Gobierno de España: https://datos.gob.es/es

Gurin, J. (2014). Open Data Now: The Secret to Hot Startups, Smart Investing, Savy Marketing, and Fast Innovation. McGraw-Hill.

MarketsandMarkets. (2017). Business Intelligence Market by Type (Platform, Software, Service), Data type (Unstructured, Semi-Structured, Structured), Business Application, Organization Size, Deployment Model, Industry Vertical, and Region - Global Forecast to 2021.

MarketsandMarkets. (2019). Cloud Computing Market by Service Model (IaaS, PaaS, and SaaS), Deployment Model (Public, Private, and Hybrid), Organization Size, Workload, Vertical, and Region - Global Forecast to 2023.

RightScale from Flexera. (2019). 2019 state of the cloud report.

Sarma, J. S., & Thusoo, A. (2017). Creating a Data-Driven Enterprise with DataOps. O'Reilly.

Uber. (2017). Uber Movement. Retrieved from Uber Movement: https://movement.uber.com/

Waymo. (2020). Waymo Open Dataset. Retrieved from Waymo Open Dataset: https://waymo.com/open/

Autores

Ernestina Menasalvas (UPM)

Alejandro Rodríguez (UPM)

Manuel Ángel Guzmán (Management Solutions)

Daniel Ramos (Management Solutions)

Carlos Alonso (Management Solutions)

UNIVERSIDAD POLITÉCNICA DE MADRID

La Universidad Politécnica de Madrid es una Entidad de Derecho Público de carácter multisectorial y pluridisciplinar, que desarrolla actividades de docencia, investigación y desarrollo científico y tecnológico.

www.upm.es

Management Solutions es una firma internacional de consultoría, centrada en el asesoramiento de negocio, finanzas, riesgos, organización, tecnología y procesos, que opera en más de 40 países y con un equipo de más de 2.500 profesionales que trabajan para más de 900 clientes en el mundo.

www.managementsolutions.com

Para más información visita

blogs.upm.es/catedra-idanae/