Quarterly newsletter

CHAIR
iDANAE
INTELLIGENCE · DATA · ANALYSIS · STRATEGY

1Q20

Data democratization

UNIVERSIDAD
POLITÉCNICA
DE MADRID

POLITÉCNICA

MS **Management**Solutions
Making things happen

Analysis of meta-trends

The iDanae Chair for Big Data and Analytics (where iDanae stands for intelligence, data, analysis and strategy in Spanish), created within the framework of a collaboration between Universidad Politécnica de Madrid (UPM) and Management Solutions, aims to promote the generation and dissemination of knowledge, the transfer of technology, and the furthering of R&D in the Analytics field.

One of the lines of work developed by the iDanae Chair is the analysis of meta-trends in the field of Analytics. A meta-trend can be defined as a value-generating concept or area of interest within a particular field that will require investment and development from governments, companies and society in the near future[1].

This quarterly report is focused on the trend of data democratization, both from the public and private fields, addressing also the need for a strong data governance.

[1]For further information, please refer to the first iDanae newsletter.

Introduction

Recently, there has been a massive increase in the generation and collection of data in all areas, both in business, civil society and administration, as well as a greater awareness of the importance of data. This opportunity to generate value, along with the appearance of tools to use and understand the data for non-technical users, has led to an increasingly relevant trend in favour of the data democratization.

From a general point of view, data democratization can be defined as the generation and availability of data for its reuse through the participation and collaboration amongst different agents, such as public institutions, companies and citizens, under the premise that the availability and analysis of data allow for more informed decisions[2]. This democratization necessarily involves making both the data and the tools accessible to users with a profile that could be non-technical, and without requiring the involvement of technical users or IT departments acting as intermediaries.

However, both the concept itself and the objective of democratization present conceptual differences, which should be analysed, depending on whether the scope of data processing is the public or the private sector. These differences will lead to different considerations, depending on the area analysed:

▶ In the field of the public sector, the democratization of data implies that the State has access to the information of its citizens, with the aim of improving its service function, which can imply great benefits to society. However, these benefits have to be balanced with the right to privacy, especially for personal data. From a government perspective, the reverse process is more interesting: the democratization of government data to make them accessible to its citizens. Some of the motivations that justify the data democratization are the increase in the transparency of data processes and governance, the fight against corruption, or the incentive for the citizen to increase her collaboration and participation, as well as to serve as a basis for the decisions taken by the different agents on the use of resources and public spaces.

▶ In the field of the private sector, from a business perspective, two approaches can be proposed, where the data democratization, its objective and its implementation are different:

- One within the company itself, in which the data collected by the company is made available across areas, avoiding silos of information. This allows areas with analytic teams to have additional data to those usually available, as well as other areas to incorporate relevant data into their decisions, in both cases enriching the data set used in decision-making.

- One external to the company itself, where the data democratization is focused on specific objectives that lead to the collaboration of groups of companies, by sharing their data among themselves or by deciding to make their data public with the aim of solving sectorial challenges. This democratization, which can be intra-sectoral or global, should imply a benefit for society. An example of this case can be seen in the existing data sharing agreements in the automotive industry with the aim of improving autonomous driving systems[3]. Another example is the data democratization in the area of cybersecurity, with the aim of strengthening existing frameworks.

---

[2]Espinosa, Garriga, Zubcoff, & Mazón, 2014.
[3]Waymo, 2020.

However, the success in achieving the objectives of data democratization lies in the trust between the different parties involved: both those who share the data and those who process it. In the field of the public sector, it is necessary to promote a two-way trust between citizens and their State; as well as in the field of the private sector, where this same concept of trust has to be extended between the different areas of the company or between the different companies.

Once the trust pillar has been achieved, an efficient implementation of democratization will then be required, where sound management, access to appropriate technological tools, and a solid governance of both processes and data need to be ensured.

From a technical point of view, in both the public and private sectors, the needs are limited to the existence of adequate tools that allow the interpretation of the data by users of different backgrounds, as well as the standardization of the information, its access and storage by third parties. The concerns that companies may face lie in different elements: in the possibility of data misinterpretation by non-specialized teams; in the security risks; in the integrity and confidentiality of the data; in the possible redundancies in the analyses carried out by different teams (as opposed to a centralized approach); or in the use of algorithms and models whose application may be executed incorrectly.

Meanwhile, the society as a whole, sometimes with government support, has taken an active role in this trend by assuming the supervision of companies and governments, with initiatives such as the Open Data Barometer[4], sponsored by the World Wide Web Foundation; or the Open Data Charter[5], formed by 73 countries and 53 organizations, whose objective is to establish principles that serve as a guide for sharing information, as well as to evaluate the degree of access to information presented by different organizations.

[4]Open Data Barometer, 2013.
[5]Open Data Charter, 2013.

Data democratization in the public sector

There are several examples of data democratization in the public sector. One of the most relevant is the establishment in 2014 by the G20 of the principles of an open data government[6], whose main objective is to fight against corruption, mainly through the following points:

▸ **Open data by default:** promotes the publication of data without the need for a prior request for information. This facilitates equal access for all users.

▸ **Timely and comprehensive:** promotes the publication of complete data sets, i.e. that they are accurate, recent and up-to-date. It also encourages the publication of data at a disaggregated level, properly documented and following internationally agreed standards.

▸ **Accessible and usable:** promotes the reduction of unnecessary entry barriers and the publication of data on central portals, facilitating the use of the data by a greater number of citizens and organizations.

▸ **Comparable and interoperable:** promotes the comparison and traceability of data from many sectors related to the fight against corruption

These principles allow, among others:

▸ Improved governance and citizen engagement: citizens are empowered and allowed to ask government institutions for explanations. It also helps them to understand, influence and participate directly in decision-making processes and the development of public policies.

▸ Inclusive development and innovation: which promote greater social and economic benefits by providing quality information to build effective and accountable institutions. Open data, in addition to preventing corruption, facilitates

the development of new knowledge, business models and digital innovation strategies. For example, it is estimated that U.S. government published data on weather across the country generates more than $30 billion annually, while GPS data generates more than $90 billion annually[7].

In Spain, the initiative to open the data to the citizens arises in 2009 with the Aporta project[8], whose basic objectives are:

▸ To promote a culture favourable to the opening of public data.

▸ Facilitate the opening up of public data by public authorities.

▸ Promote the market for the reuse of public information.

In 2011, and thanks to these previous initiatives, the Government of Spain launched a website[9] with open data collected by the Administration, with around 300 initiatives and tens of thousands of data sets. At the same time, a similar initiative was launched at the European level[10].

---

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

Data democratization in the private sector

The areas in charge of carrying out analytic processes have traditionally performed this function separately from the rest of the areas, including those with analytical functions, counting on the IT support to ensure quality and time-to-market. However, nowadays, numerous tools that enable data democratization have been developed, making it easier to share and analyse data by different areas of an entity or among several companies.

There exist two types of tools that make these analyses possible: on the one hand, those focused on improving data storage, collection and processing; on the other hand, those focused on the visualization and comparison of data, the generation of models and hypothesis testing.

Regarding the former, the software for cloud storage stands out, since it allows data from different areas to be stored together, as well as the virtualization software, which allows the user to collect and manipulate data through functional tools in order to eliminate inconsistent data, carry out analyses on the quality and completeness of the data and execute transformations on the variables. Currently, around 94% of companies use cloud-based services[11], which is expected to represent a global market of over 620 billion dollars by 2023[12].

About the latter, there are various business intelligence products that allow data to be analyzed and interpreted without any training in data science, thanks to the use of standardized procedures that are easy to implement. The use of business intelligence tools is expected to result in a market of over 26 billion dollars by 2021[13].

Case study: Airbnb y Uber

Data democratization has emerged first in those companies whose value proposal is based on data collection and analysis, both because of the competitive advantages it represents as well as because of the business culture based on data exchange. Airbnb is an example of this, as they are carrying out numerous initiatives, both externally and internally. Within these initiatives, they have arrived at some solutions to solve the problems mentioned above, based on the development of the following points:

▸ Accessibility to data through a stable and scalable data infrastructure.

▸ Processing tools designed both to be more powerful, for the data scientists to obtain more information from the data, and to be exploited by skilled workers.

▸ Training for employees, so that all workers have sufficient knowledge of the infrastructure and tools to be able to make decisions based on the data. To this end, they have created a data university[14], which functions internally and to which any employee of the company can join.

Uber has taken a different approach. Since each team performed its own analyses, it initially established a centralized data organization responsible for the company's data platform, so that consistency and efficiency were guaranteed. Once this centralization of data was achieved, Uber developed two approaches, one aimed at technical users and the other at non-technical users. For technical users, a self-service platform based on SQL was developed, while for non-technical users it was decided to establish a DataOps team and part-time dedication by the company's data team, so that they could improve the self-service platform and support non-technical users in its use[15].

Externally, Uber launched the Uber Movement[16] in 2017, where it has published anonymized information on more than ten billion trips. The purpose is to promote research collaboration, with objectives ranging from urban planning (to adapt existing infrastructure and make future investments) to reducing traffic jams and emissions or increasing driving safety.

---

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

Data governance

While there are differences in the approach taken by different organizations in their data democratization initiatives, at the core of all strategies is the existence of a strong data governance, ensuring control and quality high enough to meet both the regulatory obligations as well as those policies that allow optimizing the efficiency of internal processes.

Many companies have limited experience in developing organization-wide data governance, due to either a lack of coordination or the absence of internal policies establishing roles, functions and responsibilities, resulting in failures in its implementation. The existence of a high level of manual processes in the generation of information, the fragmentation of front systems, or deficiencies in the completeness, updating and coherence of the information, also represent obstacles when it comes to establishing a data culture in the company. To be able to develop a good implementation of a data governance framework together with the establishment of a data strategy, several points need to be considered, that synthesize the steps to be followed:

▸ Establishment of a framework and a governance responsible for the development of internal policies, the creation of the organisational structure and the new functions and roles (CDO[17], DP[18], DDM[19], Data Owner[20], Data Steward[21], etc.),

which establish the data strategy, as well as data protection.

▸ Establishment of a data structure and metadata, through the creation of a data dictionary (definition of the structure and its content), the establishment of functional and technical traceability, the guarantee of consistency and coherence of the data between environments, and the implementation of the tools that support the above.

▸ Establishment of data quality plans, through the creation of KPIs, quality controls, remediation plans and consistency analysis.

---

[17]Chief Data Officer, responsible for the entire information governance framework.
[18]Data Protection Officer, in particular linked to data protection regulations, such as the European GDPR.
[19]Data Domain Manager, the IT partner for specific areas of the data.
[20]Responsible for the definition and conceptual validation of the data and its characterization.
[21]Operational manager of a given domain with specific knowledge of the data and its uses.

Conclusions

Within the revolution that the generation and collection of data is bringing to the industry, with companies developing numerous strategies to become data-driven, the democratization of data is a step further towards generating value, new opportunities and competitive advantages. The achievement of these objectives is based on a principle of mutual trust between the actors involved in the democratization process. On this basis, the sharing of data, whether within the same company or between different companies with a common objective, increases the flexibility and immediacy of the analysis that allows for an improved decision-making process. However, it also involves various risks such as data security or duplication of analysis. That is why it is essential to have a solid and transversal data governance throughout the entity, which ensures both regulatory compliance and efficiency in the processes.

Bibliography

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

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

Government of Spain, 2011a. Data from Government of Spain. Retrieved from https://datos.gob.es/en

Government of Spain, 2011b. Open data initiative of the Government of Spain. Retrieved from Open data initiative of the Government of Spain: https://datos.gob.es/en

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.

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

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

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/

Authors

Ernestina Menasalvas (UPM)

Alejandro Rodríguez (UPM)

Manuel Ángel Guzmán (Management Solutions)

Daniel Ramos (Management Solutions)

Carlos Alonso (Management Solutions)

UNIVERSIDAD
POLITÉCNICA
DE MADRID

POLITÉCNICA

ManagementSolutions
Making things happen

The Universidad Politécnica de Madrid is a public-law organization of a multisectoral and multidisciplinary nature that is engaged in teaching, research, as well as science and technology development activities.

www.upm.es

Management Solutions is an international consulting firm whose core mission is to deliver business, risk, financial, organizational and process-related advisory services, with operations in more than 40 countries and a multidisciplinary team of 2,500 professionals working for over 900 clients worldwide.

www.managementsolutions.com

For more information, visit
blogs.upm.es/catedra-idanae/