Big Data in Catalonia

March 2021

Technology snapshot

Catalonia Trade & Investment

Generalitat de Catalunya
Government of Catalonia

Big Data in Catalonia: Technology snapshot

ACCIÓ
Regional Government of Catalonia (Generalitat de Catalunya)

The contents of this document are subject to a Creative Commons licence. Unless otherwise indicated, reproduction, distribution and public communication are allowed, provided the author is quoted, no commercial use is made thereof and no derivative work is distributed. Please refer to a summary of the licence terms on:
https://creativecommons.org/licenses/by-nc-nd/4.0/

The use of trademarks and logos in this report is for information purposes only. The aforementioned trademarks and logos belong to their respective owners and are under no circumstances the property of ACCIÓ. This is a partial illustrative representation of the companies, organisations and entities forming part of the Big Data ecosystem. Some companies, organisations and entities may have not been included in the study.

Execution
ACCIÓ Strategic and Competitive Intelligence Unit
8Wires

Collaboration
Secretariat of Digital Policies, Regional Government of Catalonia (Generalitat de Catalunya)
CIDAI

Barcelona, March 2021

Technology trends target

This study is part of a collection of reports on the different disruptive technologies detected in the ACCIÓ Technology Trends Target (2018), which are expected to have a great impact on Catalan society and its productive fabric in the years to come.

- Virtual reality and augmented reality
- Artificial intelligence
- Internet of Things
- DLT and blockchain

DIGITAL

- 3D printing
- Nanotechnology and new materials
- Robotics
- Quantum science and photonics

PHYSICAL

- Neuroscience
- Advanced drug design
- Omics
- Biotechnology and synthetic biology

BIOLOGICAL

- Big data
- Connectivity – 5G
- Cybersecurity

CROSS-CUTTING

- Drones
- Connected vehicle
- *New space*

TECHNOLOGY HYBRIDISATION

* The reports published as of 1 March 2021 appear in blue. Those pending publication appear in red.

Table of contents

1. Big Data – Definitions 5

2. Open Data and data ethics 13

3. The global Big Data Market 18

4. Big Data – Its important to industry 29

5. Main applications by demand sector and SDGs 33

6. Big Data in Catalonia 38

7. Business cases 52

8. Big Data and COVID-19 58

1. Big Data - Definitions

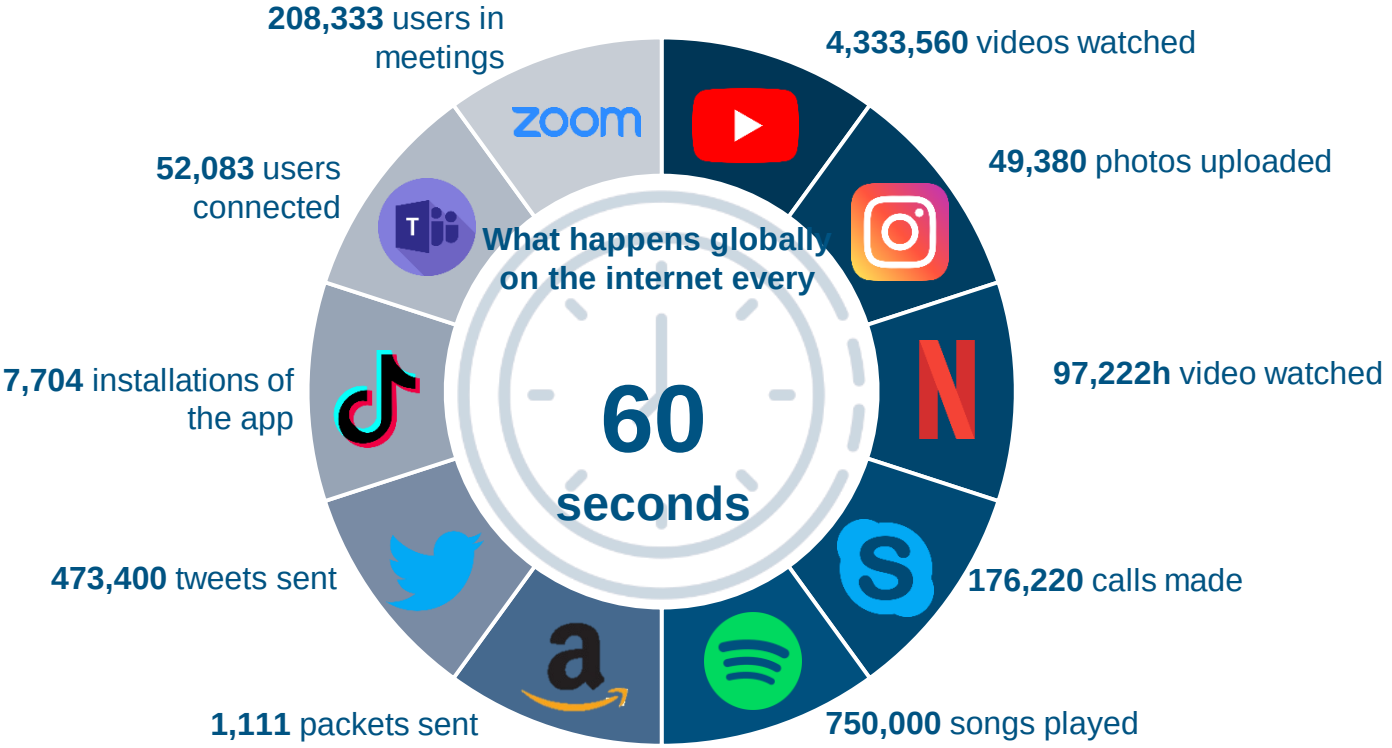
"Data is the new oil"

3.8 B active users in social media

6h 43min average daily internet use per user

4.54 M internet users

Global e-commerce consumption value **$3.43 B**



Source: Orange, marketing 4 eCommerce and Domo

Definitions

Big Data can be defined as:

“Volumes of data that are so big and complex that they require untraditional technologies to process data and extract their value”

The arrival of these problems means that new technological tools are also required to tackle these new needs.

In business, **Big Data** is used to define the large databases that major companies use. Whenever the company is smaller, normally SMEs, the concept used is **Small Data**.

What are its features?

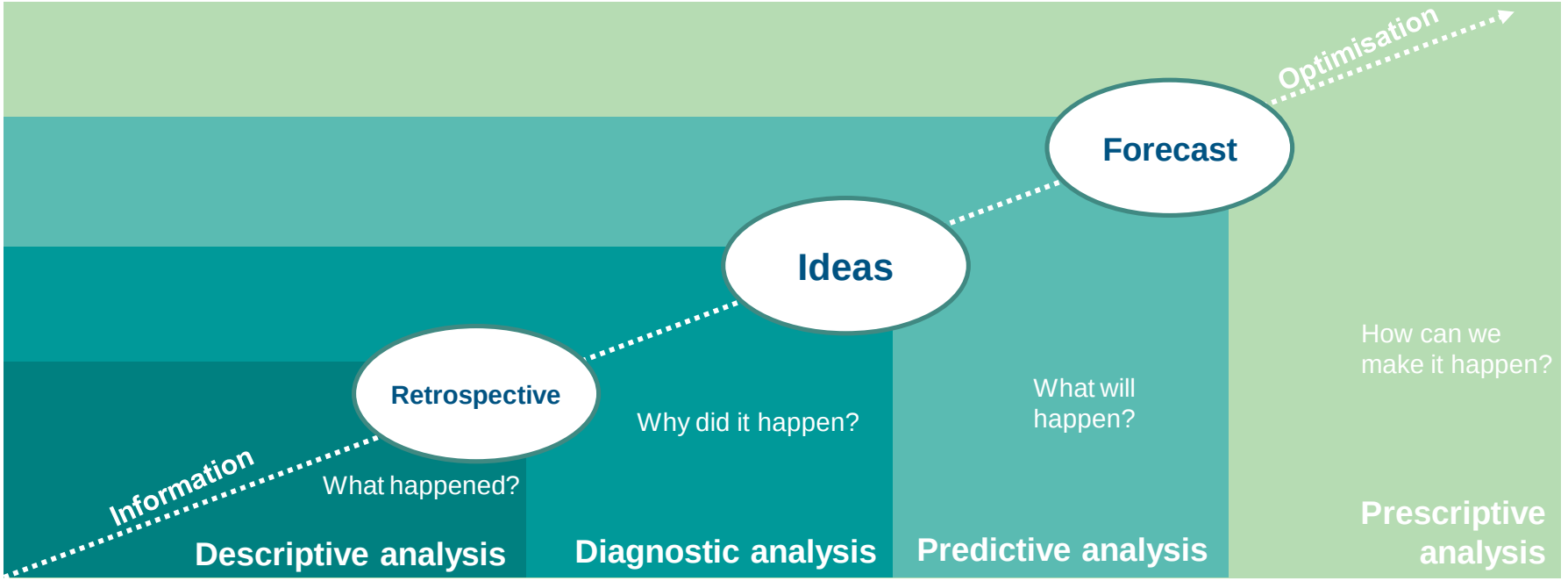
Velocity: data must be processed in real time, with the challenges that this involves

Volume: large volumes of data that require special technology to be stored and processed

Variety: different types and structures of data from a wide range of sources

Origin

The possibilities of analysis provided through the use of correctly processed information lead to the different ways available to obtain value from the information.



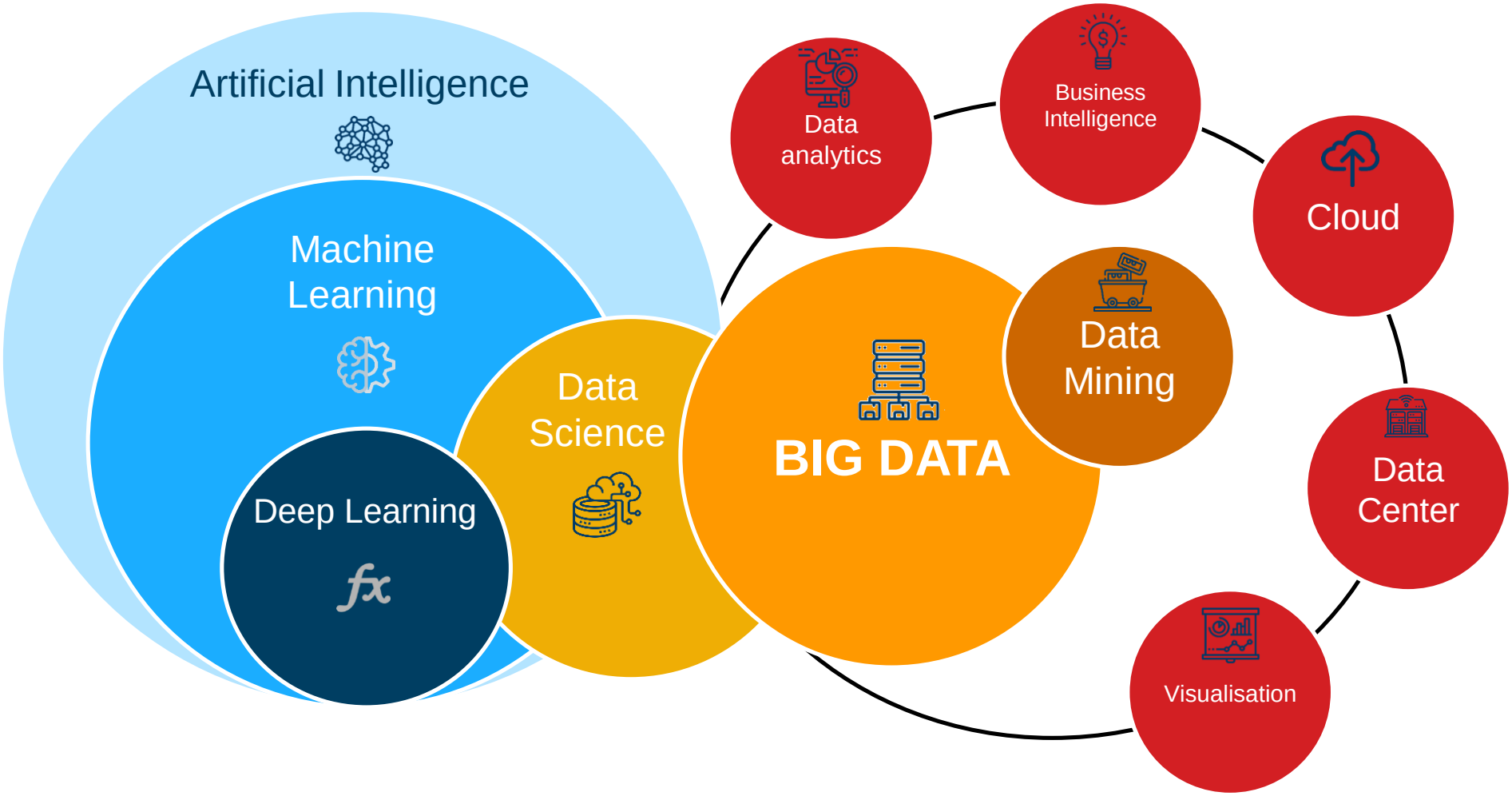
Extracting value from data through different modelling techniques to receive knowledge on what happened

Having the information obtained from the data available enables us to establish relationship and understand the whys and wherefores of this situation

Generating value from the data through techniques and mathematical algorithms in order to ascertain what will happen

The prescriptive analysis enables us to establish action to take to meet the goals. Causal relationships are not as well defined in terms of prediction

Relationships among Big Data disciplines



Source: EIC (DGI-ACCIO) based on UPF

Definitions of disciplines

Artificial Intelligence	**Computer systems** that can perceive their environment, think, learn and act according to and as a result of their objectives. They are systems of algorithms designed to assist people in their work, which can analyse and process right data that may involve the use of Big Data.
Machine Learning	**Machine learning** is the method that automates the construction of data-based models of reality. This enables systems to learn the behaviour of real systems, identify patterns and make decisions with the minimal amount of human involvement.
Deep Learning	**Deep learning** is a specific case of an artificial neuronal network capable of processing large volumes of data in a short amount of time to solve automatic learning problems: classification, clustering, or regression. Its behaviour has proven to be excellent in image identification and tagging, among others.
Data Science	**Data Science** is a multi-disciplinary field that aims to generate knowledge by applying mathematical and statistical techniques on a large amount of data to extract information, understand the reality, and discover patterns with which to make decisions.
Big Data	**Big Data.** This refers to a large number of data, which may or may not be structured.
Data Mining	**Data mining.** This is used for the automated analysing of large databases, identifying patterns to explain the behaviour of the data.
Business Intelligence	**Business Intelligence** refers to the processing of data thanks to analytical processes that turn it into information to be able to make decisions.
Cloud	**Cloud** is the global network of remotely interconnected severs that work as a single system, allowing for access to its files from any device.
Data Center	A **Data Center** is a data processing centre where the associated components forming the necessary storage systems are located. It is where the necessary part of the hardware is to be found for the storage of Big Data.
Visualisation	**Visualisation** refers to the mechanisms required to transform data into visual elements to visualise reality through the data collected.
Data Analytics	**Data Analytics** involves data processing for interpreting, processing and decision-making thanks to the data collected, allowing for the resources obtained thanks to the data to be optimised.

Change in paradigm

Throughout this context and under the umbrella concept of the methods and strategies for management and automation to obtain value from information, certain technological factors are arising that change the panorama of data mining:

IoT and sensorisation

Being able to measure and collect data simply from different devices opens up a whole new world of possibilities.

Hyperconnectivity

The possibilities of communication to be able to collect, send and store this data favours its use.

Smart devices

Devices are appearing such as smartphones and wearables that are sensorised and connected and that generate a large amount of data.

The advance and integration with other technologies such as connectivity - 5G, sensors, computing and supercomputing - can significantly help transfer data into knowledge and expand the fields and models of application.

2. Open Data and data ethics

Open Data: Definition

Open Data is data that can be used, reused and distributed freely by anyone. This data has the following characteristics:

- **Available and accessible:** the information is entirely available at a reasonable reproduction cost. The format is accessible and editable.
- **Reusable and redistributable:** the data is in a format with conditions that allow for and encourage reuse and redistribution, and it can even be integrated into other data sets.
- **Universal participation:** everyone must be able to access, use, reuse and redistribute the information. There is no discrimination regarding effort, people or groups. The main feature of open data is that its access is free to everyone.

Open data is one of the main symptoms of democracy in regions thanks to the accountability, access to knowledge, increased transparency and the generating of economic and social value.

The **tendency** to open data is a phenomenon in **global expansion** that is being increasingly considered by more and more different countries, communities, authorities, organisations and institutions.

At European level, the **Open Data Portal** is the portal of reference for **open data**. The nature of the data is made up of fields such as geography, finance, election results, health, etc.

Open data fosters **inter-departmental** and **inter-institutional** collaboration among the different **authorities**, at both local and **international level**. The connection of this data may become useful for **individuals** and for **businesses**, exponentially multiplying its value thanks to interaction among authorities and with the end users themselves.

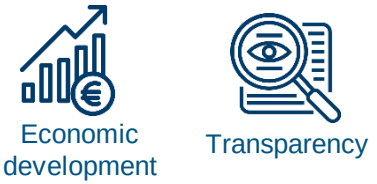
During the **COVID-19** crisis, open data has been open of the cornerstones of **transparency** through the process, and has played a key role in the total count of cases and the monitoring of the general public.

Open Data in Europe

The EU **Open Data** portal (data.europa.eu) was created in **2012** as a result of the desire of the **European Commission** to urge all European institutions to make their data public wherever possible. The goal is to provide free, easy access to data in order so that the general public can make the most of its economic potential.

The nature of the data in the portal includes the following fields:

All the data is in a **free-of-charge** format. It can be used in databases, projects or reports. The main **goals** of the European open data portal are:



The open data portals are distributed around all European Union countries. Also included are the portals of international European bodies.

INTERNATIONAL PORTALS

- eurostat — ec.europa.eu/eurostat
- Food and Agriculture Organization of the United Nations — fao.org/statistics/
- OECD BETTER POLICIES FOR BETTER LIVES — data.oecd.org/
- UNdata A world of information — data.un.org/
- WORLD BANK GROUP — data.worldbank.org/

EU MEMBER STATE PORTALS

- data.gov.be
- opendata.riik.ee
- data.gov.hr
- data.public.lu
- dados.gov.pt
- data.gov.ro
- data.egov.bg
- data.gov.ie
- dati.gov.it
- data.gov.mt
- podatki.gov.si
- data.gov.uk
- data.gov.cz
- data.gov.gr
- data.gov.cy
- data.overheid.nl
- data.gov.sk
- opendata.swiss
- opendata.dk
- datos.gob.es
- opendata.gov.lt
- data.gv.at
- avoindata.fi
- opingogn.is
- govdata.de
- data.gouv.fr
- data.gov.lv
- dane.gov.pl
- dataportal.se
- data.norge.no

Open Data in Catalonia

The **Open data portal** of the Generalitat de Catalunya was 10 years old in 2020. The current COVID-19 crisis has changed the situation of open data, and has marked a turning point:

- **More cross-cutting:** it was the first time data had been published in a cross-cutting manner on a single subject. Different authorities and external players came together to update and channel the data into a single platform.
- **Greater interest:** there has been a very significant increase in monthly activity on the open data portals during the pandemic.

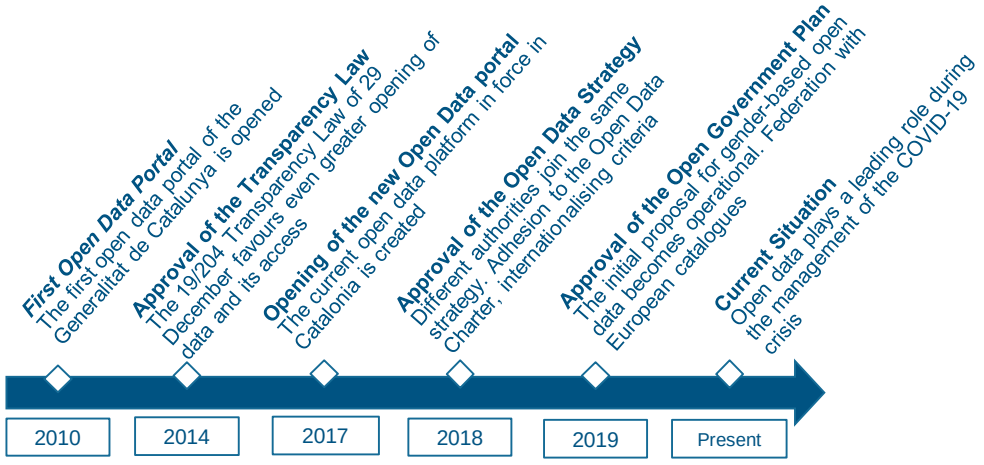
Current situation* of Open Data in Catalonia

654 Datasets	574K Visits	698K Downloads
	+ 500% Visits/month	+ 623% Dwnld./month

Variation during COVID-19

Related solely to the pandemic, there has been a total of **270K** visits and **450K** downloads.

Chronology of Open Data in Catalonia



Main future challenges:

To continue ensuring the **principle** of open data

Unifying international criteria to compare and use data **globally**

Prioritising opening data in **subjects** considered **essential** (gender, environment, culture, etc.)

Fighting for the **security** and **veracity** of data and encouraging the **media** to ensure the fight against and interpreting of fake news

*Historical data since the portal was opened

Ethics in the world of data

The world became aware of the importance of protecting personal data over three decades ago. In 1980, the OECD, in collaboration with the European Council, published a guide on the principles of privacy and the transborder flows of data based on the following principles:

I. Establish clear **limits** in obtaining the data

II. Determine the **relevance** of data for its foreseen use

III. Clearly define the **use** to be made of the data before it is requested

IV. Abstain from using data for **uses other** that those determined primarily by data subjects

V. Ensure the **protection** of data in light of illegal access or piracy

VI. Ensure the individuals on whom the data has been collected have **access** to it and can request modifications or deletions

VII. Ensure the advances, practices and policies on the use of data remain **open and transparent**

These principles are to be found in all guides and constitute the most long-lasting framework of reference on the protection of personal data, the **GDPR**.

The transversal implementation of data ethics is on **the rise**, and an increasing number of institutions of different kinds are supporting this kind of management. Some examples of this are:

Governments are the main players developing ethical frameworks in line with the international ethical frameworks

The IEEE produced a document regarding the importance of ethics in Big Data

Cities have also developed initiatives along these lines. Amsterdam, New York or Barcelona are just some examples

Universities have produced guides and frameworks of reference for the players working with their data

Companies as relevant as Microsoft, IBM and Google have formulated their ethical principles. This is a growing trend in business.

Generalitat de Catalunya

The **Generalitat de Catalunya** has promoted the creation of the **Data Ethics Committee** to guarantee responsible digital transformation by the Authorities. This body will be responsible for ensuring the ethics of the data through advice, reflection, and the generating of good practice in the deployment of advanced uses of data in the provision of public services. This committee is another addition to the ethical processing of data being developed by the **Artificial Intelligence Ethics Observatory of Catalonia** (OEIAC)

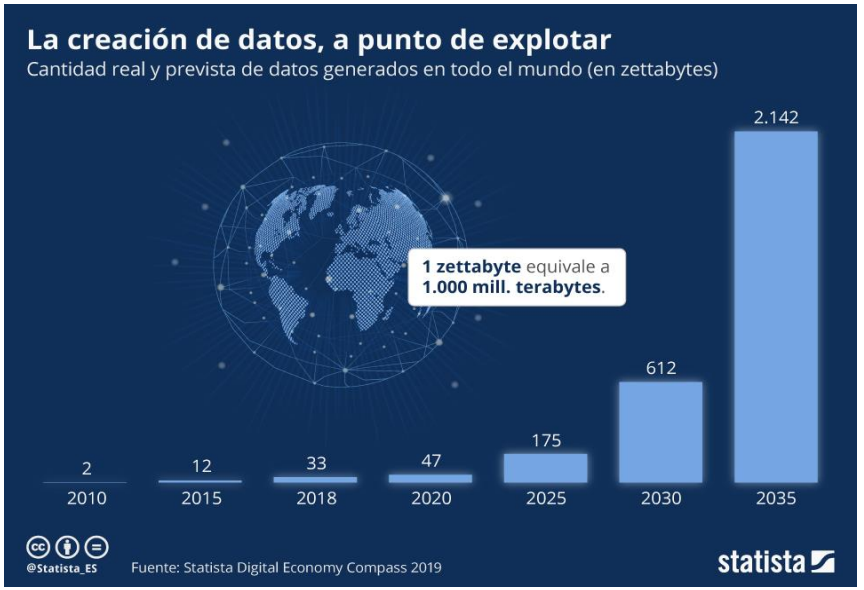
3. The global Big Data Market

Global market

In Europe in 2020, Big Data accounted for **4% of the European GDP** and involve 10.5 million jobs

Over the first decade, Big Data and Artificial Intelligence increased the **world GDP** by 1.2%. It is calculated that around 13 billion dollars will be generated in extra economic activity by 2030

Evolution in the volume of data generated:



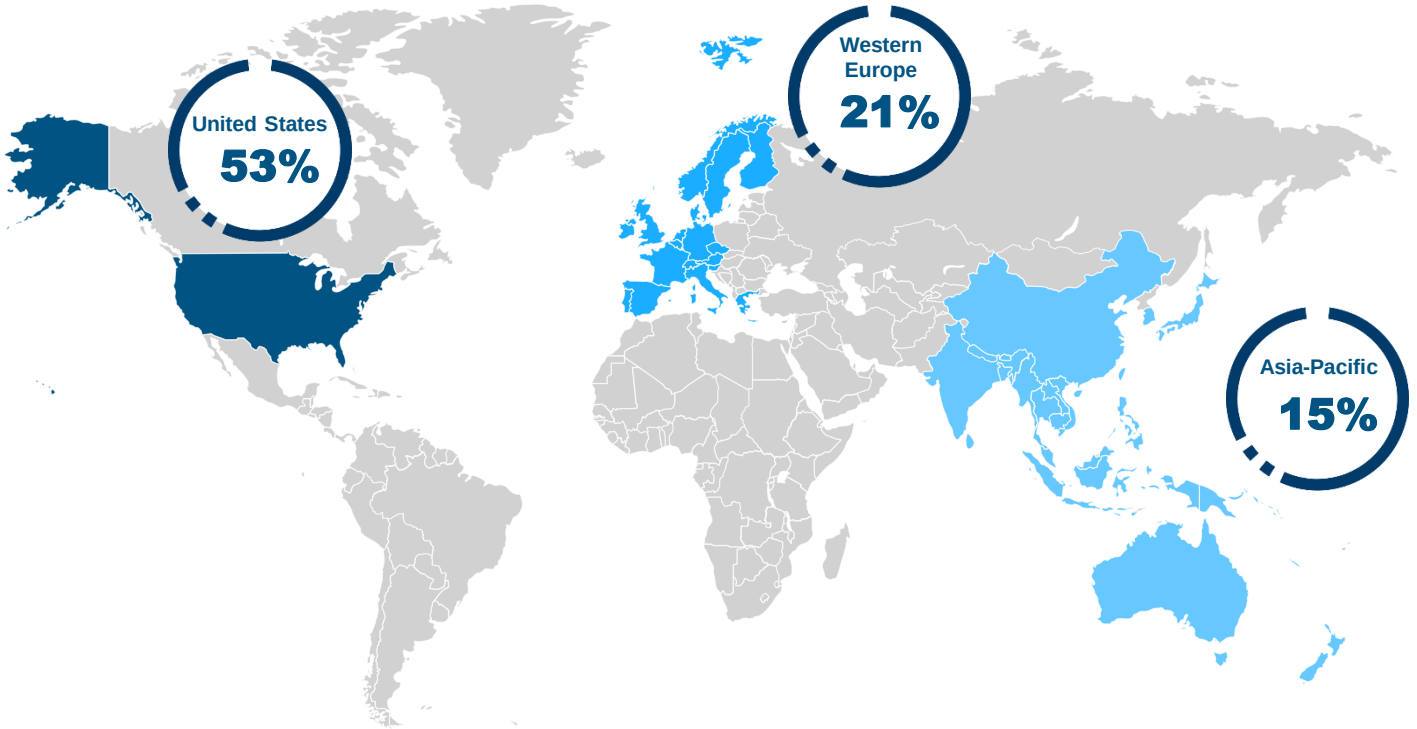
Big Data experts have become one of the most sought-after professional profiles and also one of the most difficult vacancies to cover

Sources: Kippel, McKinsey, IEBS and Statista

Turnover in Big Data

It is foreseen that global revenues from the Big Data market will rise from 138.9 billion dollars in 2020 to **229.4 billion dollars in 2025**. The United States is the country where the Big Data section will experience most growth.

Distribution of the Big Data market by the main regions (% of the total)

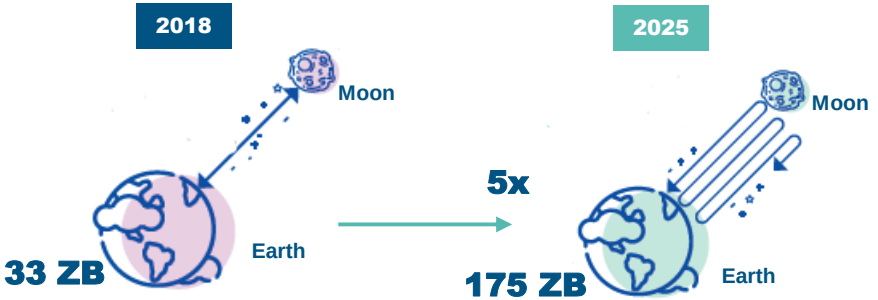


This market will have a **compound annual growth rate** (CAGR) of **10.6%** until 2025. The COVID-19 crisis, which favoured the digitisation of companies and of society in general, may boost the data sector even further.

Source: Markets and Markets, 2020; IDC, 2018

Evolution of the global and European data market

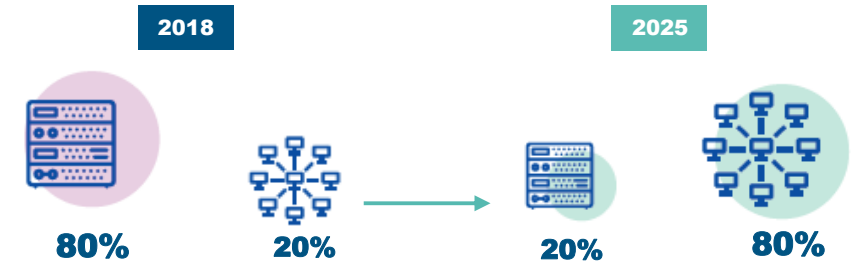
The volume of data will experience exponential growth



Stored in 512 GB tablets, it would form a tower to be sent to the Moon.

It would be possible to travel to the Moon and back five times.

Data storage will change



Centralised computer installations

Connected smart devices

Centralised computer installations

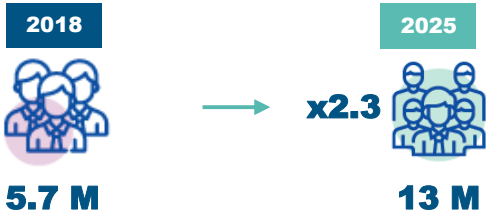
Connected smart devices

Value of the data market
(EU-27)

(2.4% European GDP)

(5.8% European GDP)

No. of professionals
(EU-27)



Population with basic digital knowledge (UE-27)

Sources: ACCIÓ, based on "A European Strategy for Data" by the European Commission, and OECD (2019) "Measuring the Digital Transformation: A Roadmap for the Future"

European Data Strategy

With an investment of between **€4 B and €6 B** up to 2025, the European Data Strategy seeks to make the EU a leader in a data-driven society.

The EU is to create a **single data market** to ensure a free flow in the **EU** and among sectors to benefit companies, researchers, and the public authorities.

As part of the European Recovery Plant, the Commission is to adapt its employment programme for 2020 in response to COVID-19 and is to present **legislative action on the exchange of data and governance**, which will be followed by a Data Law.

Data-based innovation will provide significant, specific benefits to European citizens, such as in transport:

The use of navigation to avoid real-time traffic can save up to **730 million hours**. This represents **€20 B** in labour costs.

Real-time notification of train delays can save **27 million hours** of work. This is equivalent to **€740 M** in labour costs.

GAIA-X

- **Gaia X is a public-private data infrastructure project** which the EU hopes will become the starting point of its cloud ecosystem and reach digital sovereignty so that it does not have to depend on non-European providers such as Amazon Web Services, Microsoft Azure or Google Cloud
- Headed by France and Germany, **it operates as a hub to which companies** (such as Deutsche Telekom, Orange, Atos or SAP) **and more than 300 organisations are connected**
- The common goal is to ensure all companies comply with the **European standards of digital sovereignty, data availability, interoperability, portability, transparency and fair participation**

Source: ACCIÓ, based on "A European Strategy for Data" by the European Commission

Main Big Data providers

amazon

The US giant is valued at over $427B. The services DynamoBD, Amazon EMR, Atenea and AWS stand out.

CLAIRVOYANT

Specialising in the end-to-end development of Artificial Intelligence (AI) and Machine Learning (ML) solutions.

EMC²

Consulting. Its technologies include Isilon, ECS, Boomi and PowerEdge for Hadoop.

Google

Google offers Cloud services for Big Data. A pioneering, benchmark company in this area.

IBM

IBM Big Data solutions provide features such as store data, manage data and analyse data.

MICRO FOCUS

Handoop Consulting and SQL Analytics, storage and processing of big data.

Microsoft

Microsoft provides HDInsigth tools for the analysis of structured and non-structured data.

ORACLE

Cloud services. Provide business solutions in Big Data Analytics, applications and infrastructures.

Prolifics

A digital transformation company using data, Analytics, Cloud, DevOps and Quality Assurance.

SAP

It provides a wide variety of Analytics tools, primarily HANA. Data storage and calculation.

ScienceSoft
Professional Software Development

Big Data Consultant. Focuses primarily on IT Consulting and software development.

splunk>

Machinery data analysis. Transaction and security monitoring. Integrating platform.

TERADATA

Analysis, marketing applications and consulting services.

vmware

Cloud platform and Big Data. Vmware vShpere Big Data as the main supply tool.

xplenty

ELT data integration platform and ELT based on the data processing cloud service.

Presence in Catalonia

Source: Software Testing Help 2021

Main global investors (I)

The **world's most active investors** in Big Data during the 2016 – 2020 period were the following:

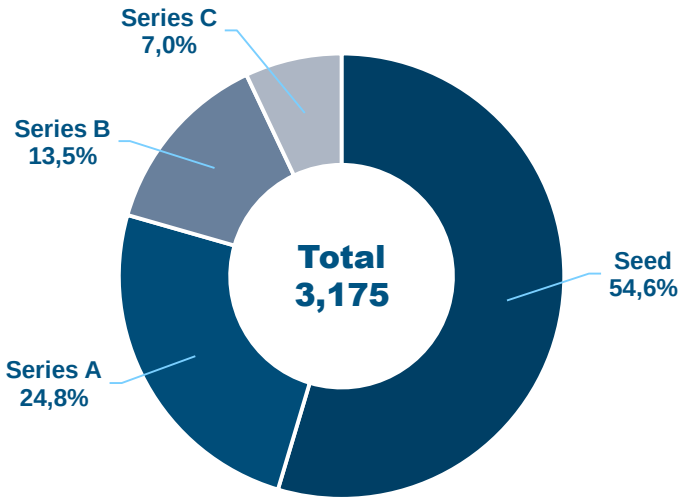
Source: Crunchbase. Projects by order of investment received

Main global investors (II)

3,175 investment rounds in Big Data were completed between 2016 and 2020, for a value of over **€22 B**

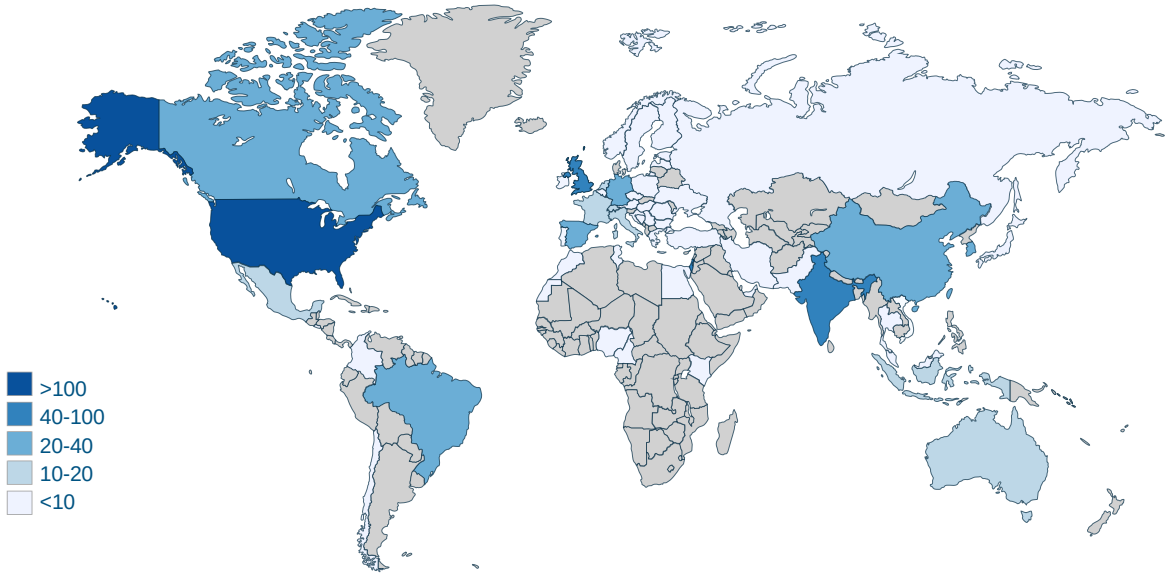
- Between 2016 and 2020, **3,175 funding rounds were completed worldwide in the area of Big Data**.
- The rounds **focused on seed investments**, followed by investments in more consolidated phases (Series A, B and C, in order).
- The investment rounds amounted to a total of **22.6 billion euros**, and the Series C rounds concentrated the highest quantity of money (36.2% of the total).
- The **United States concentrates a large part of the funding rounds**, with a total of 1,385 (43.6% of the total) and €13.4 B. Behind it are **China** (347 funding rounds and €3.4 B) and the **United Kingdom** (223 and €1.1 B).

Funding rounds in Big Data in phases
(% of the total, 2016-2020)



Source: Crunchbase
*Investment round intervals according to Crunchbase

Concentration of startups per country



Country	Number
United States	242
Israel	58
United Kingdom	51
India	43
Germany	37
China	27
Canada	24
Singapore	23
South Korea	23
Spain	23
Remainder	349
Total	**900**

- Of the **900 startups*** that include Big Data in their processes, 242 are from the United States (26.8%). Four out of every ten of these are located in California.
- **India and China** in Asia, along with **Israel, the United Kingdom and Germany**, concentrate most of the remaining startups.
- Of the 23 startups established in Spain, almost half (**10 startups**) **are located in Catalonia**.

Source: ACCIÓ, based on Crunchbase

*Startups established during 2018-2020

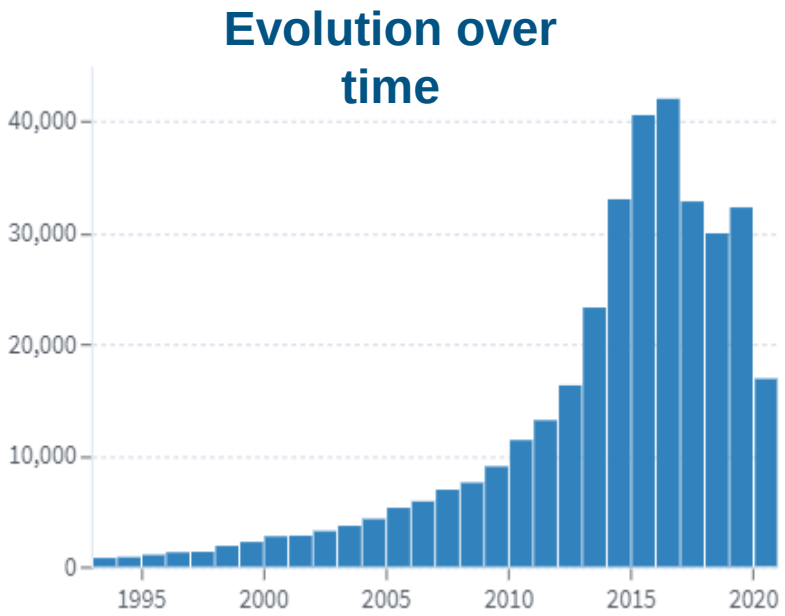
Patents in Big Data (I)

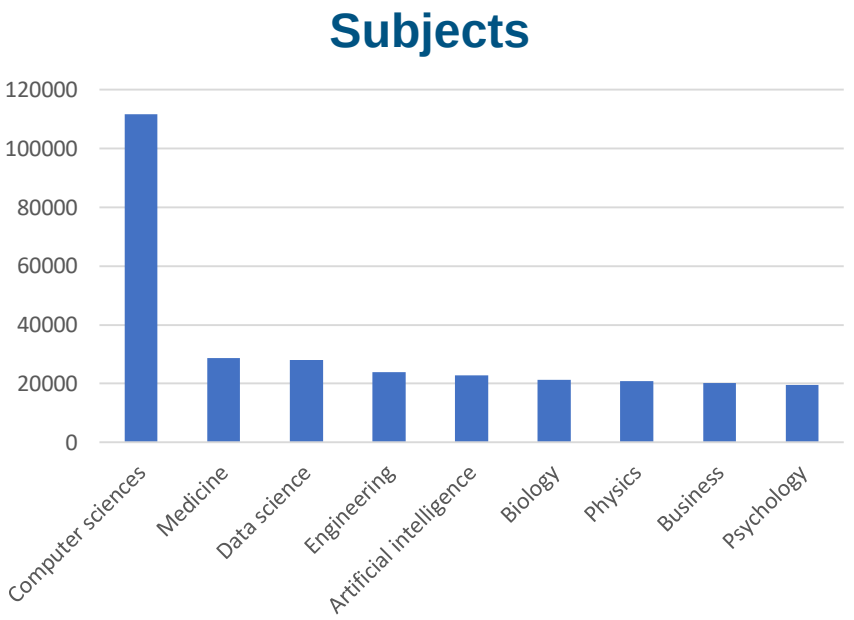
According to the portal Lens.org, which specialises in patents around the world, there are currently **356,500 patents registered** that involve Big Data. The following map shows the sites where most have been registered and the most noteworthy geographic areas.

Source: lens.org (21-09-2020)

Patents in Big Data (II)

The growth in the “no. of patents / year” of Big Data reached its peak a few years ago but remains stable.

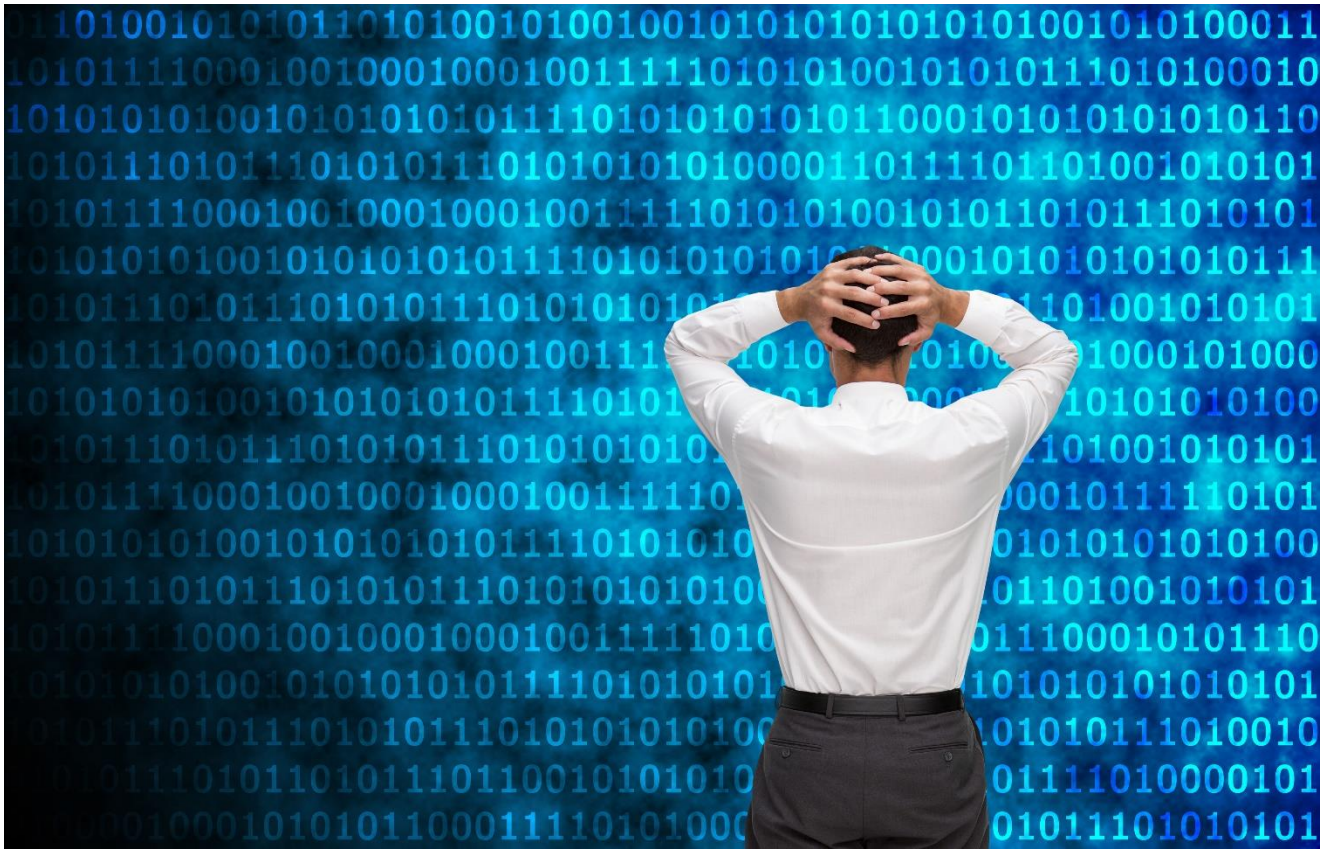




The area of knowledge in which most patents are registered is, without a doubt, computer sciences.

Source: lens.org (21-09-2020)

4. Big Data - Its importance to industry



The importance of Big Data to industry

It allows for the analysis of critical aspects to effectively reduce costs without affecting quality. The evolution of technologies also means that Hadoop costs only 10% of traditional systems.

Reducing costs

Innovation and Business opportunity

Big Data generates new business models. Companies, measuring the satisfaction and needs of their customers, have the power to give them what they want. They can create new products based on their preferences.

Big Data requires quality standards to assure the precision, completeness, integrity, consistency, relevance and reliability of the data and of the processes involved.

Control and reliability

Cross-cutting approach

Beyond a cross-cutting approach involving all economic sectors, Big Data can play a fundamental role in companies, it being possible to extract data and information of value from any area or process.

Convergence with other technologies

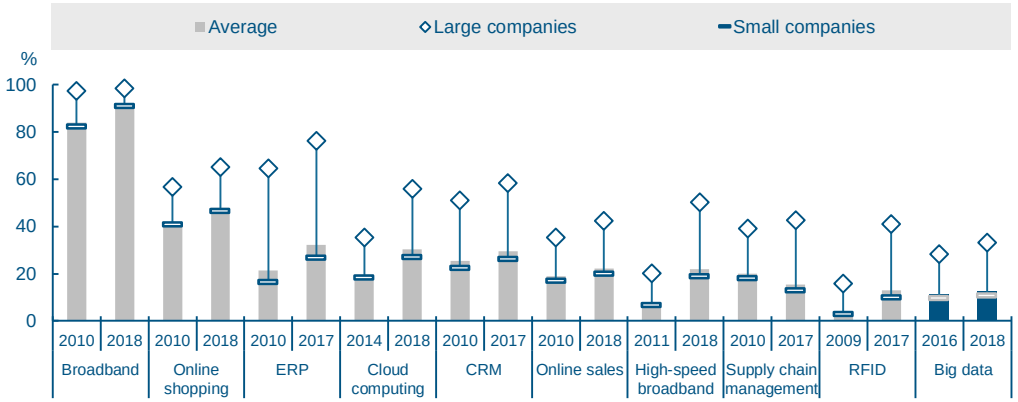
Cloud, IoT (Internet of Things), 5G, Artificial Intelligence (AI) and Big Data technologies are and shall remain the key players in the growth and development of the industry.

Source: ACCIÓ

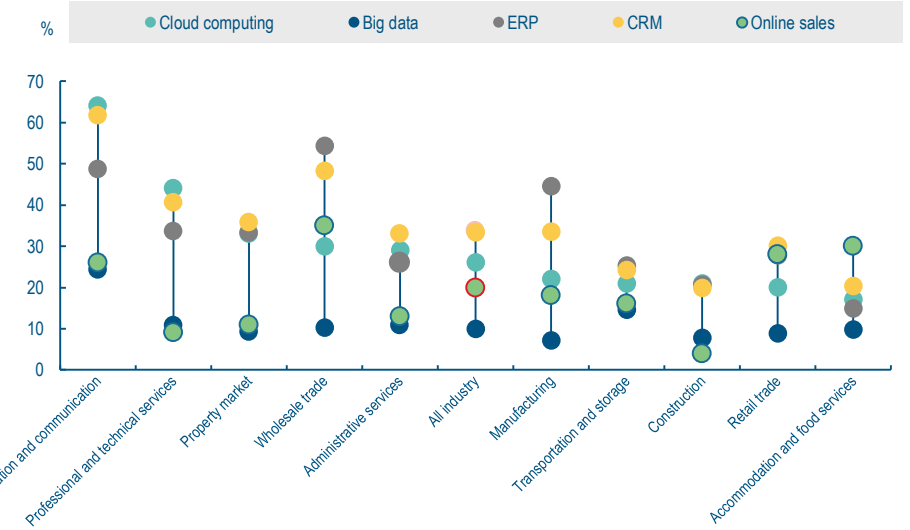
Big Data to businesses and industry

Big Data is recognised as one of the ICTs that is to play a leading role over the coming decade, along with high-speed broadband (5G) and cloud computing.

Adoption of ICTs in businesses (OECD countries, 2010-18)



Adoption of ICTs per industry (EU-28, 2018)

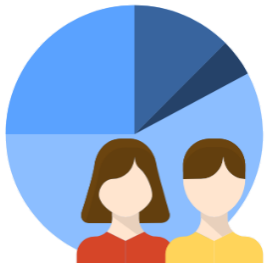


- Big Data has even lower implementation levels in companies (12.6% in 2018, 1.3 p.p. up on 2016).
- The level of implementation of digital tools is very high among large companies and not so high among SMEs, despite the fact that these technologies are increasingly accessible.
- By industry, Big Data is more heavily implemented in information and communication (24%), and transportation and storage (15%), and is at the rear particularly in the manufacturing sector (7%).

Source: OECD (2019) "Measuring the Digital Transformation: A Roadmap for the Future"

Industrial applications of Big Data

In any industrial area where data is generated by machines and humans, there is an opportunity to extract very valuable information for the business.



Customer knowledge

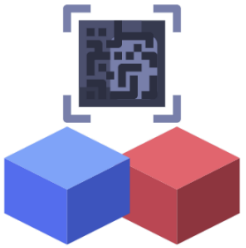
Useful to correctly monitor the market, analyse the market feel, customer support, recommendations, dropout prevention, customer lifecycle management, etc.

Operations and production

Optimisation of business processes, production flows, quality assurance, dynamic scenario simulation, demand forecasting, pattern detection, predictive and preventive maintenance of equipment.

Risk management

SLA (service level agreement) negotiations, with knowledge of the actual behaviour of equipment, detection of cyberfraud and cybercrime, perimeterisation of facilities with automatic image recognition.

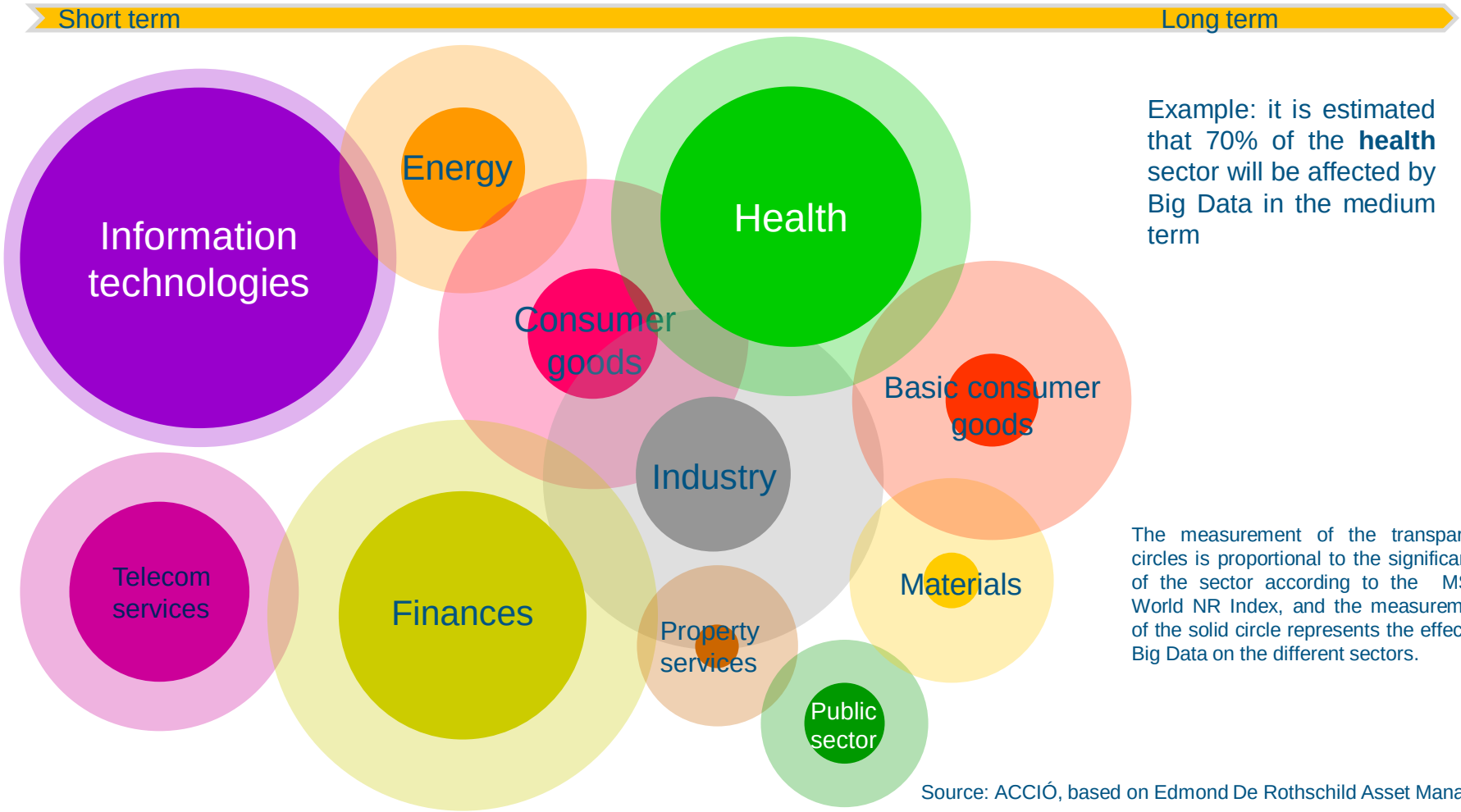


Digital products and services

Data as a product and as a service, technical knowledge management.

5. Main applications by demand sector and SDGs

Big Data by sector (I)



Example: it is estimated that 70% of the **health** sector will be affected by Big Data in the medium term

The measurement of the transparent circles is proportional to the significance of the sector according to the MSCI World NR Index, and the measurement of the solid circle represents the effect of Big Data on the different sectors.

Source: ACCIÓ, based on Edmond De Rothschild Asset Management

Big Data by sector (II)

The sectors that invest most heavily and in which Big Data is expected to have most impact are:

Information technologies: the arrival of Big Data is related to the development and the advances in information technologies, such as storage systems, cloud, and data security and integrity.

Telecom: offering consumers better contracts, fine-tuning marketing campaigns, calculating many more variables (social media) when weighing up customer satisfaction, designing new products and services.

Energy: anticipating grid errors, improving maintenance, adjusting demand (predictive models), identifying faults, discovering new consumer trends or making the most of sunlight or wind.

Trade: personalising supply, recovering inactive customers and responding to their demands in a fast, effective manner, optimising stock.

Finance: control of risk of fraud or defaults, real-time reaction to market events, cash management or advanced customer segmentation.

Health: predictive analysis of diagnoses, patient monitoring, epidemiological studies, pharmacological interaction.

Industry: optimisation of logistics, inventory or production, predictive diagnoses, variability in manufacturing processes or traceability.

Property services: optimising, ensuring the flexibility of and improving services, and identifying opportunities and patterns of behaviour.

Public sector: smart cities, knowledge management, cybersecurity, infrastructure performance, optimisation of citizen requests.

Materials: design, selection of materials, and decrease in waste.

Use Cases: Data-based business model

The use of Big Data is essential in major corporations. Different business models that promote their business model through the analysis of data. Some examples of this are:

FOOD

Big Data combined with artificial intelligence (AI) has enabled them to improve on a wide range of issues, from the bottling plant to the operations of them vending machines. **They analyse**, for example, the data generated by their 100 million-plus followers on social media for the continuous improvement of their **marketing** campaigns.

SCIENCE

Discovering the secrets of the universe with Big Data. In 2013, the CERN announced that it had found the **Higgs boson**. The sensors on the Large Hadron Collider (LHC) record hundreds of millions of collisions between particles, some of which reach speeds that are just a fraction under the speed of light. This generates a **large quantity of data** and requires very sensitive and precise equipment to **measure** and **record** the results.

COMMUNICATIONS

Thanks to its extensive range of applications launched for banking, insurance, travel and entertainment and to its portable devices and wearables, Apple is **collecting** more data than ever. This provides **knowledge** on its users and the experience of each client is increasingly **personalised**.

DISTRIBUTION

Big Data improves the **efficiency** of supermarkets. Supermarkets compete not only through price but also through **customer service** and, essentially, through **convenience.** Having the right products in the right place at the right time so that people can buy them involves huge **logistics** problems. They must always have competitive prices to remain competitive themselves.

Big Data and OBJECTIUS DE DESENVOLUPAMENT SOSTENIBLE (II)

End of poverty: To know the exact situation of poor people (location, name). Along with other tools, to pass on knowledge of agriculture and medicine.

Health and well-being: Tracking mobile phone user movements could help predict the spread of infectious diseases. The rapid analysis of large amounts of cases allows for better conclusions to be drawn.

Quality education: Reports from citizens could expose the reasons for the school dropout rates.

Gender equality: Analysis of financial transactions could disclose expenditure patterns and the different impact of financial crises on men and women.

Affordable and clean energy: Smart computers enable companies to increase or restrict the flow of electricity, gas or water to reduce wastage and guarantee an adequate supply.

Decent work and economic growth: Trends in global postal traffic could provide indicators on economic growth, deliveries, trade and GDP.

Clean water and sanitation: Sensors connected to water pumps could detect clean water and prevent disease.

Reduction of inequalities: Analysis of the rhetoric of the contents of local radio could reveal problems of discrimination and provide support in adopting inclusive policies.

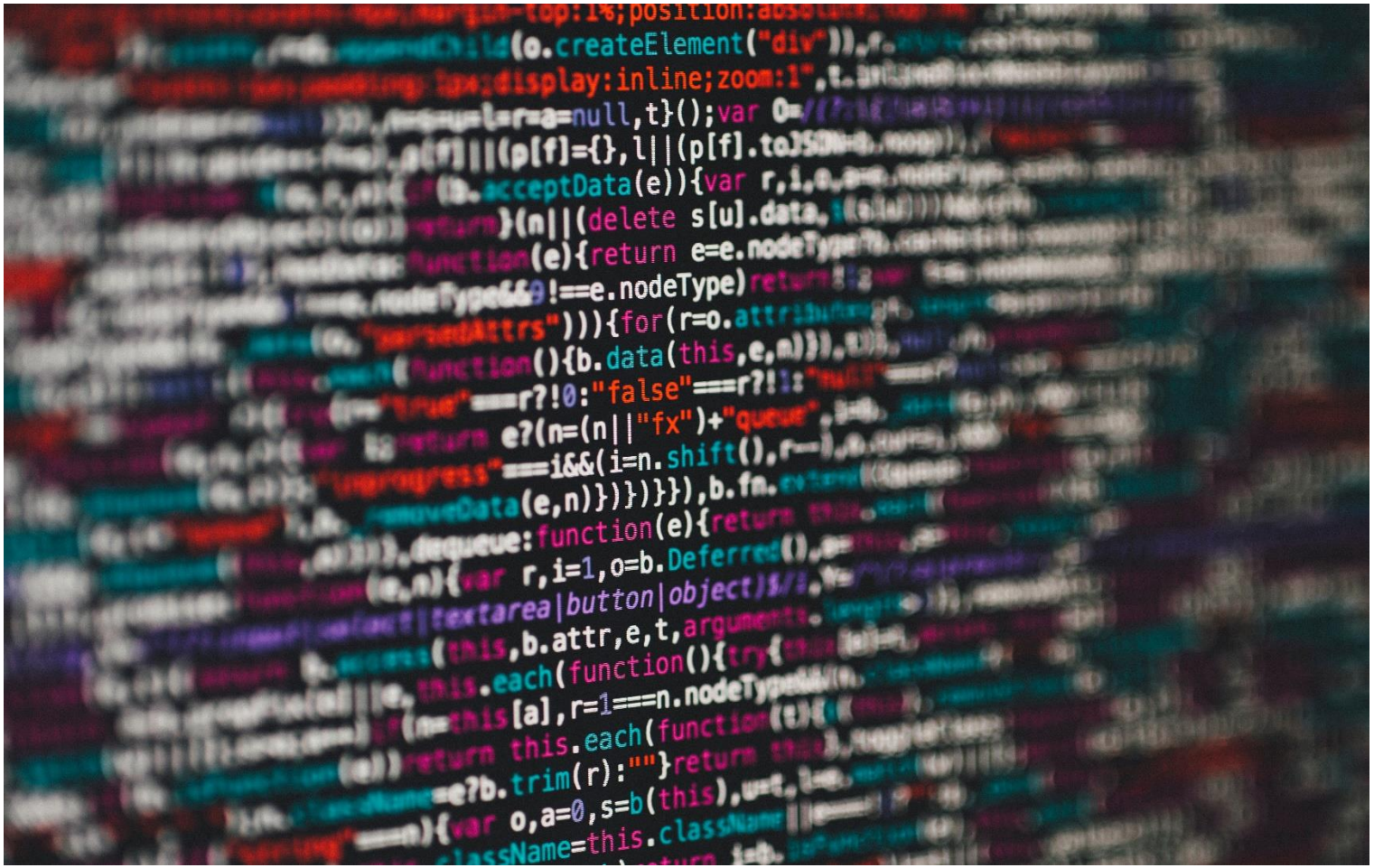
Industry, innovation and infrastructure: Data from GPS devices could be used to control traffic and improve public transport.

Climate action: monitoring of pollutant emissions, and the analysis and optimisation of energy consumption. The combination of satellite images, the testimonies of people and free-access data could help track deforestation

Peace, justice and strong institutions: Analysis of emotions on social media could show public opinion on issues such as effective governance, the provision of public services, or human rights.

Sustainable cities and communities: Remote detection via satellite to track trespassing on land or public spaces, such as parks and forests. Helps improve security in cities, reducing crime.

6. Big Data in Catalonia



Value chain segmentation in Catalonia

Within the Big Data market, the players can be classified in line with the following value chain:

Owners of the data

Owners of different types of data, from public sector information (governmental data) to scientific or commercial data, and that supplied by sensors. In some cases, the data is owned by specific bodies, such as the government or research centres.

Science

Public Authorities

Vertical applications

Data applications in specific sectors.
Service providers and analytics tools devoted to specific sectors and verticals (e.g. Retail, health, etc.). This includes analytical and AI services and tools devoted to specific verticals such as marketing, legal, governance, science, health or finance. Always with a role as an analytics service provider for third parties.

Data analytics

Companies that structure and convert the data into useful information:
Cross-cutting providers of analytics and AI services and tools. General data analytics and AI services, tools and products.

Technology providers. Data Technologies.

Companies that provide data management tools, platforms, services and knowledge. Providers of services, tools and technologies related to the data cycle (collection, storage, management and visualisation). All services, support tools, technologies and infrastructures are covered, whether they are commercial and/or open code, acting on a more infrastructure-based level.

Marketplaces

Data, algorithm, model and service markets. Supply-demand connection.

Facilitators

Facilitators, promoters and/or enhancers of the Data Analytics and AI ecosystem. This includes venture capital companies that provide risk funds for the creation of data, incubators that provide support for growth, research institutes that provide innovation, training organisations that provide adequate skills, public regulators such as the European Commission that can have a great impact on the development of the sector through, for example, its data protection directives, etc.

Active users

Active Data Technology and AI users. Users with assets and resources devoted to data analytics and AI to improve their own business.

Industry

Public Authorities

Civil Society

Source: ACCIÓ in collaboration with CIDAI

Focal point of the study. These segments must be considering in mapping and quantification.

Big Data in Catalonia: main mapping conclusions

220 companies

€250 million

2,493 jobs

89% of the total are **SMEs**

35.5% of the companies have a turnover of **more than one million euros** and **12.7%** have a turnover of **more than ten million euros**

Relationship with Artificial Intelligence:

11.8% of the companies also appear in Artificial Intelligence mapping

61.4% of the companies are less than ten years old

38.2% of these companies are startups

Gender diversity in management

16.4% of the companies employ women in management

Low degree of internationalisation:

25.5% of the companies are exporters

Most noteworthy in the value chain are companies that deal in data analytics (**64.1%**); and technology providers (**38.2%**).

The Big Data ecosystem in Catalonia is supplemented by companies with sector-based vertical applications, including those dealing in e-commerce and in health.

Source: ACCIÓ 2019 data

Value chain segmentation in Catalonia: *Pure Players*

Partial illustration chart*

Data analytics

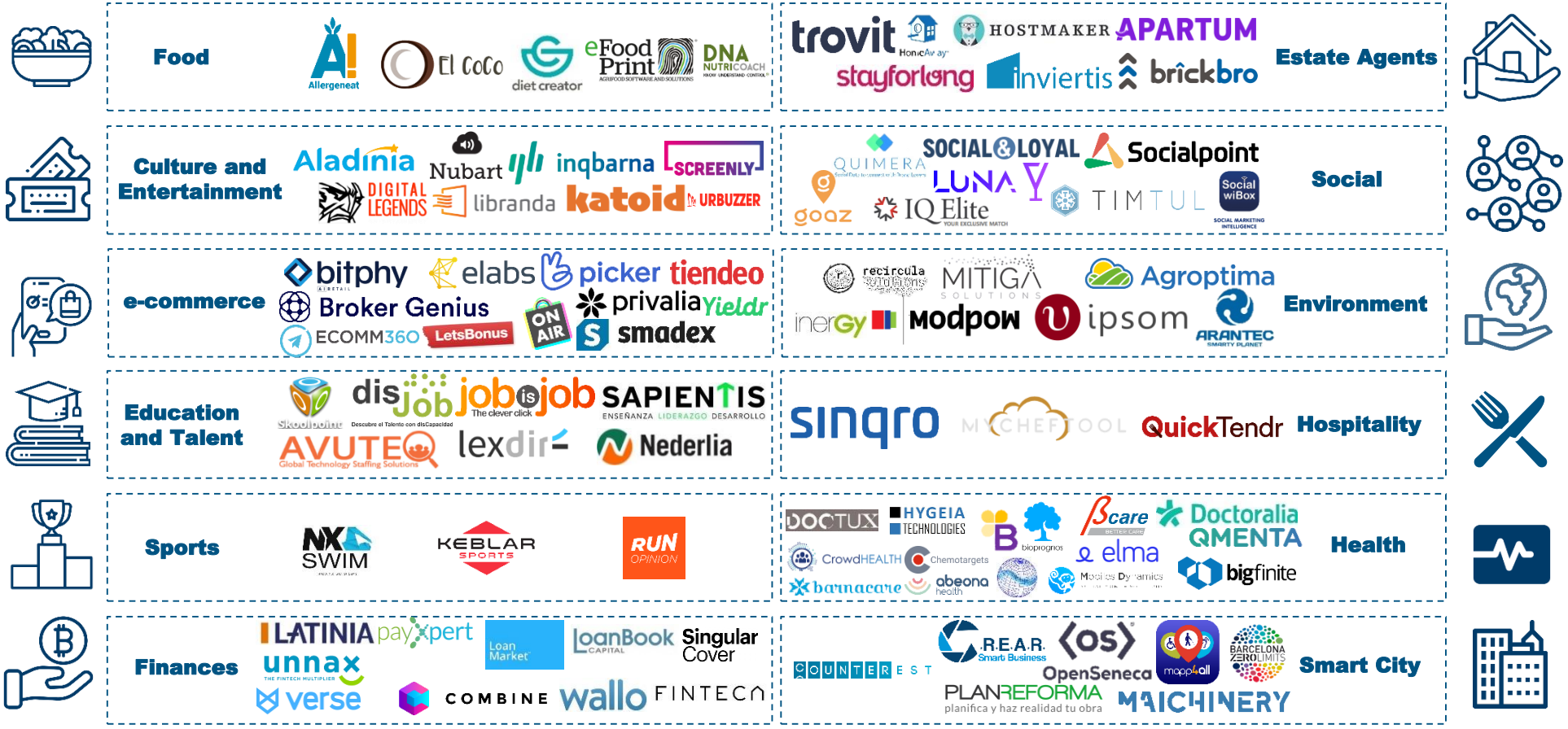
Technology providers

Marketplace

* Source: ACCIÓ. The main companies in each segment are represented, according to turnover

Value chain segmentation in Catalonia: Vertical Applications

Partial illustration chart*



* Source: ACCIÓ. The main companies in each segment are represented, according to turnover

Value chain segmentation in Catalonia: Owners and active Users of data

Partial illustration chart*

Owners of the data

Agència Catalana de l'Aigua
AISTECH
ARANTEC
camerdata
Diputació Barcelona
DNA NUTRICOACH
IMAN
ONEDATA
SALUS.COOP
STORY DATA
sue2

Vertical applications

Data analytics

Technology providers. Data Technologies.

Marketplace

Active users

aeca group
ALTOPLAST
arquia banca
cabify
Corporació Catalana de Mitjans Audiovisuals, SA
DIGITAL LEGENDS
DKV
Doctoralia
FLUIDRA
im! in movement
MAICHINERY
OgilvyOne worldwide
privalia
re
Socialpoint
SJD Sant Joan de Déu Barcelona · Hospital
santasusana
SAPIENTIS
Signaturit
TEXTINGBIGDATA
trovit

Source: ACCIÓ

Big Data ecosystem agents in Catalonia

Source: ACCIÓ

Big Data Centres in Catalonia (I)

The Barcelona Supercomputing Centre is the national supercomputing centre. It is a consortium formed by the Generalitat de Catalunya Department of Business and Knowledge, the UPC, and the Ministry of Science, Innovation and Universities. The BSC manages the Spanish Supercomputing Network (RES) and is a top-level member of the Partnership for Advanced Computing in Europe (PRACE) and manages MareNostrum, one of the most powerful supercomputers in Europe, which is located in the deconsecrated chapel Torre Girona.

The CVC is a non-profit research centre with its own legal status, established in 1995 by the Generalitat de Catalunya and the Autonomous University of Barcelona (UAB). The mission of the CVC is to carry out leading research in the field of Computer Vision and achieve maximum local and international impact. The CVC is a CERCA centre, which actively promotes the transfer of knowledge generated in industry and society, and prepares and trains researchers and engineers to the highest standards.

The Technological Telecommunications Centre of Catalonia is a public R&D+i centre created by the Generalitat de Catalunya. The research (basic and applied), innovation and technology transfer are based on technologies of a physical, linking and network level of communication systems, on network services and infrastructure, as well as on Geomatics. The activities are organised into 4 divisions: Systems, Networks, Communication Technologies, and Geomatics, and include the advice of an international third-party scientific committee.

The DAMA Centre specialises in the development of software for the management and analysis of big data for companies and the public authorities, and responds to the demand of different sectors through technology transfer projects, the provision of specialist services, technical advice, or training.

Big Data Centres in Catalonia (II)

DATASCIENCE@UB is a research and technology transfer centre involved in Big Data technologies and advanced data analytics. It conducts industrial research projects in Big Data technologies, such as the recommending of products and trends, revenue management to maximise profits, or demand forecasting models. It also provides consultancy services to help companies monetise and value data in order to improve their internal operations or to introduce new lines of business.

The Easy Centre has been performing basic research in the field of artificial intelligence, combined with technology transfer in many applied research projects, for more than 15 years. It aims to create leaders in technological innovation and to develop research projects linked to the real world.

Eurecat, technology centre of Catalonia, has a long history that dates back to 2008, in the application and innovation of data technologies and artificial intelligence in different sectors, particularly industry, energy and resources, health, agrifood, robotics, cybersecurity, and cultural and creative industries. Eurecat promoted the creation of the Centre of Excellence in Big Data of Barcelona (Big Data CoE Barcelona) in 2015 and is an active member of the Big Data Value Association. In 2020, the BIG Data CoE was awarded the "Gold i-Space" qualification by the organisation of the same name and, as the party responsible for the CIDAI (Centre of innovation in Data technologies and Artificial Intelligence) Technical Office, Eurecat continues to promote the adoption of the data-controlled economy and the transfer and innovation in AI in the Catalan production network.

The Grup de Tecnologies Interactives focuses on the human side of technology and explores new uses and technologies, with research components in person-computer interaction, learning supported by 3D technologies and graphics. The GTI specialises in projects with companies seeking to work together to develop value-added lines of innovation with an important visual component. Examples are: 3D graphic editors on the web, animated avatars, mining, Big Data narratives and visualisation, gamification strategies, GPS mining, spaces and platforms for future learning, etc.

Big Data Centres in Catalonia (III)

The i2CAT Foundation is an applied research body in the field of the Internet and advanced digital technologies. Backed by more than 10 years of experience in multiple national and international R&D&I projects, it leads research lines in fixed and mobile network architectures, wireless sensor networks, big data and content-based multimedia technologies. The aim of the research carried out at i2CAT is to develop new products, services and applications in the fields of eHealth, Smart Cities & Smart Regions, Advanced Manufacturing and Digital Social Innovation.

The primary mission of the Institut d'Investigació en Intel·ligència Artificial is research into Artificial Intelligence, development and innovation through the transfer of results from society-based research. The main lines of research are learning systems, multi-player systems, and reasoning and logic.

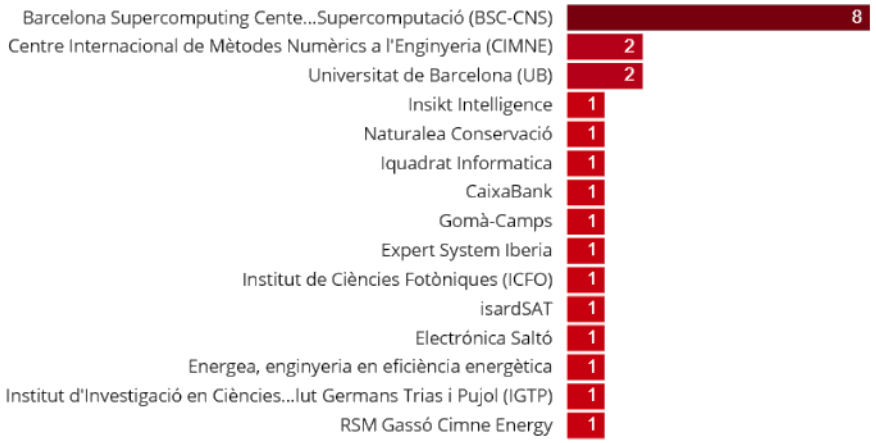
The inLab FIB UPC is the innovation and research laboratory of the Faculty of Computer Science at the UPC with a track record of over 40 years of collaboration with organisations and companies. Its mission is to transfer knowledge to society and develop human talent through R&D+i projects based on the latest ICT, Data Science and Big Data technologies, Simulation, Smart Mobility, and Cybersecurity.

The Institut de Robòtica i Informàtica Industrial is one of the leaders in research into robotics in Spain, given the volume of its scientific production and its major presence in R&D projects in Europe. The centre offers expertise and technological capacities in different areas of robotics, computer-based vision, artificial intelligence, automatic control, etc. The sectors to which its technology can be transferred are mobility and automotive, advanced manufacturing, industrial robotics, logistics and automation, healthcare robotics, augmented reality, smart cities, energy efficiency, environmental management, etc.

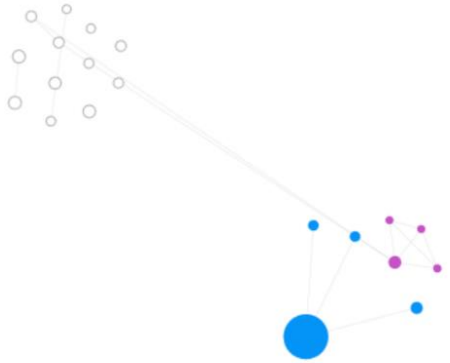
La Salle R&D aims to be a landmark technology centre in the field of smart cities and the health sector, driving technology transfer to the business fabric, renowned and recognised on a national and international level for its excellence in research and development. The mission of La Salle R&D is to promote the use of ICTs in every-day life, providing added value and competitiveness to companies through applied research and the development of new innovative and unique solutions.

Big Data research activity

Ranking of organisations according to number of projects



Organisation relationship diagram



19 projects

20 organisations

€10.1 M of investment

236 foreign partners

Project keyword cloud

Source: RIS3-MCAT platform

CIDAI Centre of Innovation for Data tech and Artificial Intelligence

A key element in the CATALONIA.AI strategy

Mission:

To demonstrate the advantages and streamline the adoption of innovative data mining and artificial intelligence technologies.

Vision:

To make Catalonia an international benchmark in innovation and adoption of an efficient, ethical AI in business and social spheres.

The CIDAI is inspired by an operational model of the European *Digital Innovation Hubs*. Founded by 9 partners with the coordination of Eurecat, the CIDAI is the instrument used to implement the innovation tool for the CATALONIA.AI strategy.

Proximity

In constant dialogue with companies and organisations throughout the region to assess their main challenges and opportunities.

Utility and impact

To prioritise and promote the most relevant initiatives to inspire and streamline access to state-of-the-art technologies.

Excellence, neutrality, and ethical principles

Activities and projects led by experts in the sectors, with continuous technology mining and application of ethics and good practice.

Generalitat de Catalunya Departament de Polítiques Digitals i Administració Pública · Ajuntament de Barcelona · Barcelona Supercomputing Center Centro Nacional de Supercomputación · CVC · eurecat Centre Tecnològic de Catalunya · everis NTT DATA Company · i2cat · Microsoft · sdg group · UNIVERSITAT POLITÈCNICA DE CATALUNYA BARCELONATECH

ACCIÓ · Generalitat de Catalunya

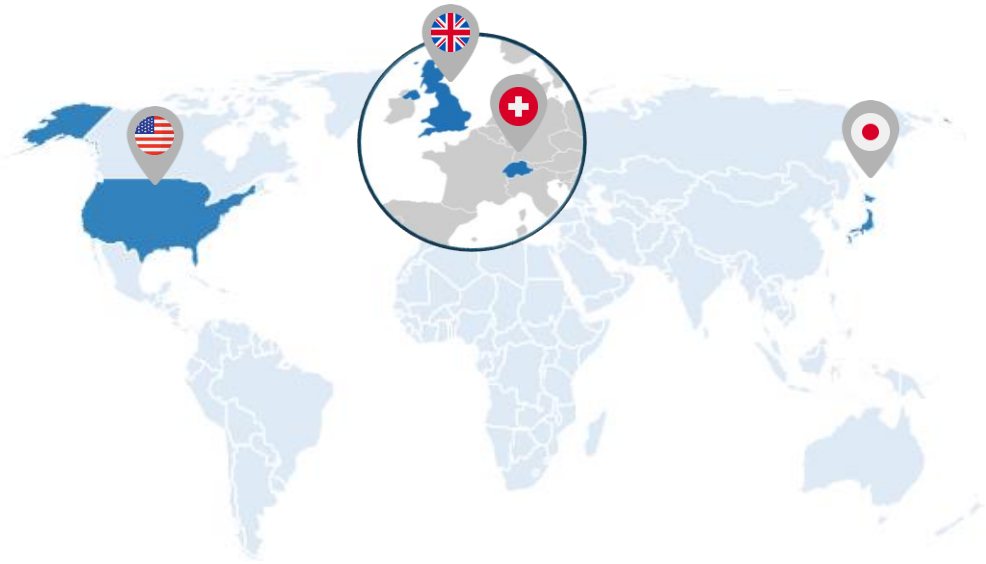
FDI in Big Data in Catalonia (2016-2020)

13
Big Data Projects

1,831
Jobs created

593.45
Capital invested (€M)

Countries of origin of the investments



13
Big Data Projects
▲**x1.8** in relation to 2011-2015

593.45
Capital invested (€M)
▲**x15.5** in relation to 2011-2015

1,831
Jobs created
▲**x5.2** in relation to 2011-2015

- Between 2016 and 2020, **Catalonia became a national point of investment, combining:**
  - **33.3%** of FDI projects
  - **70.7%** of capital invested
  - **43.7%** of jobs created

Companies completing FDI projects in Catalonia (2016-2020)

RED SIFT, Cien, everis an NTT DATA Company, ZURICH, Cytel, mixpanel, ocado, dynatrace, KOCHAVA, NetApp

Source: ACCIÓ based on fDi Markets

Talent in Catalonia

Despite still being a very young technology, which explains why there are no specific degrees on the subject, Catalonia has a wide range of **master's degrees, courses and post-graduate studies** on **Big Data** to train professionals in one of today's most sought-after areas with highest growth in the near future. Below are the **main Big Data courses in Catalonia**:

INSA Business, Marketing & Communication School

- Global Executive Master's Degree in Business Intelligence
- Master's Degree in Big Data & Intelligence

Universitat de Barcelona

- Master's Degree in Data Science Fundamentals
- Advanced Course on an Introduction to Data Science & Big Data

Institut de Formació Contínua-IL3 Universitat de Barcelona

- Post-graduate Degree in Marketing Intelligence and Big Data applications
- Master's Degree in Big Data Engineering

UOC Universitat Oberta de Catalunya

- Post-graduate Degree in Data Engineering and Big Data.
- Master's Degree in Business Intelligence and Big Data: Data Analytics.
- Post-graduate Degree in Business Intelligence and data analytics.
- Master's Degree in Business Intelligence and Big Data: Information Systems.
- University Master's Degree in Data Science

upf. Universitat Pompeu Fabra Barcelona

- Official Master's Degree in Data Science
- Master's Degree in Intelligent Interactive Systems (track Big Data)

laSalle Universitat Ramon Llull

- University Master's Degree in Big Data Engineering

UPC Universitat Politècnica de Catalunya BarcelonaTech

- Master's Degree in Big Data Management, Technologies and Analytics
- Erasmus Mundus Master's Degree in Big Data Management and Analytics
- Master's Degree in Data Science

upf. Universitat Pompeu Fabra Barcelona / **barcelona school of management**

- Master of Science in Management (specialization in Business Analytics)

eurecat BTS Barcelona Technology School

- Master's Degree in Big Data Solutions

ulc barcelona

- Executive Master's Degree in Big Data Science

UVIC

- Official Master's Degree in Omic Data Analytics

esade

- Executive Master's Degree in Business Analytics
- Diploma in Business Analytics
- MSc in Business Analytics

ubiqum

* Specialist Big Data course

- Bootcamp in Data Analytics & Big Data

7. Business cases

Business cases in Catalonia: Holaluz and the use of data

Company established in Barcelona in 2010, which deals in the **resale of 100% renewable electricity**, and of gas. It also represents producers of renewable energy in the electricity market, and provides services relating to photovoltaic self-consumption and electric mobility. It currently has around **300 thousand customers**.

It uses **data** primarily to **forecast the daily energy consumption of its customers** (standard use in the sector) based on historic consumption, meteorological data, consumption patterns and different mathematical models, and for more exclusive uses such as the calculation of flat rates or power optimisers.

Data processing at Holaluz currently cuts across all departments (purchasing, sales, pricing, forecasting, BI, etc.) and they jointly use the **Amazon Cloud** service (option selected particularly due to security). But this has not always been the case: initially they had physical servers in the offices and the data processing was outsourced and they worked with simple tools such as Excel.

holaluz

Information provided by Roger Lloret, Business Intelligence Lead at Holaluz

Business cases in Catalonia: the most influential woman in the world in the field of technology is Catalan

"Analytics Inside", the magazine specialising in information on the technology and Big Data analytics sector, has recognised the vice president of NetApp, Anna Navarro, for her "**career, her contribution to global engineering, and for making women visible in the field of technology**".

The executive from Olot is the **head of Women in Technology** along with several other managers, and specialises in content and globalisation strategies within the technology sector. She is currently the **vice president of NetApp, a leading company in data storage**, with headquarters in Silicon Valley, and leads a team of 200 professionals.

In 2020, NetApp opened a new Digital Sales Hub in Barcelona from where it provides support to its sales teams in the EMEA region.

This new project, supported by ACCIÓ, seeks to manage the business opportunities of the digital sales teams generated through agreements with partners such as Amazon Web Services, Google Cloud and Microsoft Azure, strategic alliances that **promote cloud innovation**.

NetApp

Business cases in Catalonia: collaboration of the BSC and SUEZ to manage the water networks

The **Barcelona Computing Center** and the company **SUEZ Environment** have reached an agreement to use supercomputing technologies, algorithms, artificial intelligence and Big Data to promote the company's expert knowledge on the **drinking water distribution** networks. The aim is to promote the sampling and development phases and validate new water management tools.

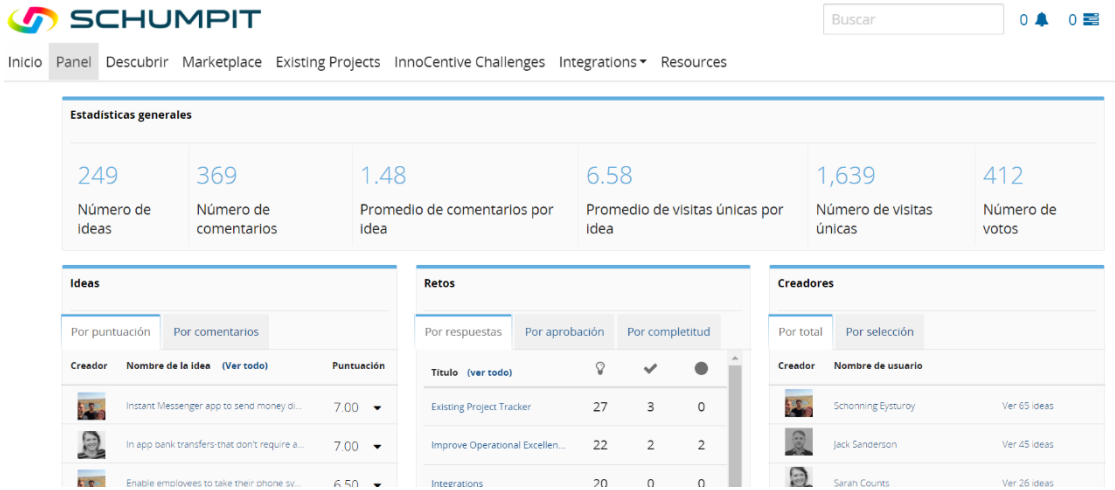
The challenges involved in this project must allow for the value of using Deep RL for the **optimisation** of the water networks to be assessed and identified. The BSC works with artificial intelligence algorithms capable of autonomous piloting industrial machinery using similar strategies to the DeepMind lines used by Google. This project will therefore test these techniques on an infrastructure of the size of the **Metropolitan Area of Barcelona**.

The goal is not only to develop solutions that improve the service, but to validate **autonomous learning** in considerable volume, such as the AMB, so that the algorithms can train with different types of network and in changing conditions regarding the water operationalisation system.

Business cases in Catalonia: Schumpit

Schumpit helps companies and public bodies promote the digital transformation of their business. This transformation takes place through a holistic vision of organisations in which the change comes from the combination of professional teams, the most appropriate technology, and the generating of a business model of true value, with a realistic return on investment.

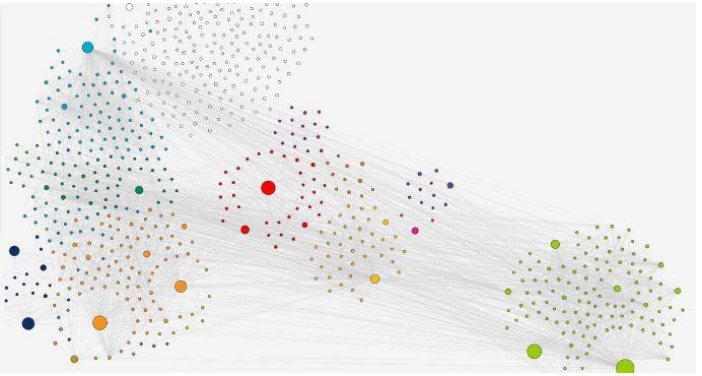
Schumpit is associated with technology platforms that work for governmental bodies such as the United Kingdom Ministry of Defence and major corporations such as Telefónica or Société Générale.



Schumpit uses Big Data to collect ideas on the design process for continuous improvement (Ideo Management). The platform collects all the ideas, extracts the relationships between them, and fosters participation of the entire organisation in developing those that generate the highest value and speed up the meeting of strategic goals.

RIS3-MCAT platform

Open government, artificial intelligence, Big Data and data visualisation project that integrates and relates open data from science and innovation projects to meet certain goals:



- To understand the impact of European funds on the specialisation of the research and innovation ecosystem in Catalonia
- To identify opportunities to maximise the collective impact of research and innovation in Catalonia based on synergies and coordinated efforts
- To provide new evidence to streamline the decision-making of players in the research and innovation ecosystem in Catalonia, fostering new collaboration dynamics and inspiring new public policies
- To give visibility to the bodies in Catalonia that take part in the European research and innovation networks
- To understand how the European funds contribute and to provide innovative responses to the sustainable development goals (SDG)
- The Platform allows for the activity of Catalan organisations in projects of the European Horizon 2020 programme and the RIS3CAT instruments to be mapped and characterised

8. Big Data and COVID-19

Big Data, data against the virus

- The Covid-19 pandemic is highlighting the importance of having data to analyse and anticipate trends so that decisions can be made.
- Since the pandemic began, a lot of information has been collected related to its spread and to how it affects the population.
- Using all this data, advanced analytics techniques are being applied to help predict how the virus will evolve based on different geographic and social factors, including the development of mathematical algorithms that strictly monitor its progress and advance.

AREAS OF RESPONSE BY TECHNOLOGY TO COVID-19

Source: ACCIÓ, ICT Sector Report 2020, BigData CoEBCN, Big Data Catalonia, EUDatalandscape

Artificial intelligence and Big Data enable the Hospital Clínic to reduce the COVID-19 death rate from 11.6% to 1.4%

The Hospital Clínic has developed an application that predicts the evolution of Covid patients so that early action can be taken. They have identified patterns that can be used along with artificial intelligence to apply personalised treatments to patients, and they have been able to reduce the death rate, even in the more elderly population and/or those at risk.

Artificial intelligence solution capable of analysing more than a trillion anonymised data on Covid-19 patients in real time, identify the different clinical patterns, and propose a personalised treatment to provide the most suitable for each patient

A project is being developed, funded by EIT Health, the European Institute of Innovation and Technology in Health, to extend this solution to other hospitals. To date, it has the participation of Mútua de Terrassa and Hospital Germans Trias i Pujol (Can Ruti) and, at European level, the ERASMUS MC hospitals in The Netherlands, and the UZ Leuven University Hospital in Belgium.

Source: Elperiódico.cat and ara.cat
The Lancet Regional Health Europe

Business cases in Catalonia COVID-19: *Big Data (I)*

Mocaplatform is a company that deals in the use of artificial intelligence for smart localisation and marketing, and which has created an app and a digital platform to respond to the needs of the public authorities and citizens during the phases of post-lockdown.

COVID-19 management and control platform based on Big Data and localisation

MOCA SDK to be integrated into any mobile app (native or Cordova).

- Detection of contacts of proximity among citizens through Bluetooth Low Energy (BLE) to know whether someone has been in contact with someone else who is infected (the data is only disclosed in positive cases).
- Automatic tracking of user mobility via GPS/WiFi to a precision of around 50-100 m, with low battery consumption and proactive management of the GDPR.
- Automatic processing of alerts to indicate when a user enters/exist a geobarrier (e.g. a place in lockdown, an area of risk, etc.)
- Segmentation of people according to their category and health (infected, risk, users who have not taken the test, etc.)
- Real-time Big Data Analytics based on GPS with heat maps to show areas of infection.

Source: ACCIÓ based on the press

Business cases in Catalonia COVID-19: *Big Data (II)*

beabloo

Interaction Care (IC) is a solution developed by **Beabloo** to **minimise the risk of COVID-19 in physical spaces**. IC controls access to spaces, prioritises assistance for vulnerable groups, measures interactions between people, and ensures social distancing is maintained.

It includes video analytics **technology**, WiFi, zenith video sensors, wide-angle IP cameras, heat cameras, and a people counter. IC reacts with **real-time** alerts and measures the different areas of risk in a space.

Valid for any kind of physical space, such as supermarkets, chemists, shops, public authorities, banks, or airports.

A team of mathematicians and statisticians from the **UAB**, from **Humboldt-Universität zu Berlin** and from the **UPC** are developing a **model** to estimate the **actual number of new cases of people infected by COVID-19 every day**, given that this data cannot be obtained directly from all the population because many cases are minor or asymptomatic.

The results provide a more realistic image of the pandemic in real time, and a more precise estimate of fundamental data such as real death rates or the basic reproduction number, which are required by professionals and politicians in order to make decisions. The analysis is designed to be easily reproduced using data from other countries.

UAB Universitat Autònoma de Barcelona

UPC

Humboldt-Universität zu Berlin

Source: ACCIÓ based on the press

ACCIÓ

Passeig de Gràcia, 129
08008 Barcelona
www.accio.gencat.cat
www.catalonia.com
@accio_cat
@catalonia_ti

See the report here:

http://catalonia.com/.content/documents/big-data-in-catalonia.pdf

More information on the sector, news and opportunities:

http://catalonia.com/industries-in-catalonia/sectors/industrial-systems-smart-industry-and-3d-printing/ICT-and-digital-transformation.jsp

For the full technological report, please send an e-mail to irodriguez@catalonia.com

Catalonia Trade & Investment

Generalitat de Catalunya
Government of Catalonia