



GENERATION DATA

USING DATA FOR PROFIT

Notes 1.0



Erasmus+

This project has been funded with support from the European Commission



EXPOSITION

1. A BRIEF HISTORY OF DATA

Data is not new. But it is getting so big and so omnipresent that it has implications for everyone, whether we want it or not. It is completely transforming the way we do business and is impacting most other parts of our lives.

2. WHAT IS BIG DATA?

So now we know what has changed, what IS big data?

- The official definition is *“Data of a very large size, typically to the extent that its manipulation and management present significant logistical challenges.” Oxford English Dictionary, 2013.* But this definition has limitations because focuses only on technical aspects and presents data as problematic.
- An alternative definition: *“Big data is high-volume, high-velocity and high-variety information assets that demand cost-effective, innovative forms of information processing for enhanced insight and decision-making.” Gartner, 2014¹* This gives us a bigger picture. The basic idea behind the phrase “Big Data” is therefore that everything we do is increasingly leaving a digital trace, which we can use and analyse. Big Data therefore refers to the data itself, but also to our ability to make use of ever increasing volumes of data.
- Some people go further: *“Big data is a new attitude by businesses, non-profits, government agencies, and individuals that combining data from multiple sources could lead to better decisions.” Gil Press, contributor to Forbes.²*

2. 1 The 5 Vs of Big Data

What’s changed about data in last 10 years making it a hot topic are the 5Vs:

VOLUME – the magnitude of data being generated.

VELOCITY – The speed at which data is being generated and aggregated.

VARIETY - The types of data available to us.

VERACITY – The accuracy or trustworthiness of the data.

VALUE – The extent to which data generates economically valuable insights or benefits.

Volume

Where just a few years ago, Gigabytes (GB) were as large as it got, the Terabyte (TB) and Petabyte(PB) have now become the common currency of Data Managers’ lives.

- More data cross the internet every second than were stored in the entire internet just 20 years ago.
- According to IBM, 90 per cent of the data that exists today was generated in the previous two years.⁴

The vast amounts of data have become so large in fact that we can no longer store and analyze data using traditional database technology. We now use distributed systems, where parts of the data is stored in different locations and brought together by software.

Velocity

- Every second of every day data is increasing. Not only must it be analyzed, but the speed of transmission and access to the data must also remain instantaneous to allow for real-time access to data, ej for website, credit card verification and instant messaging.
- For many applications, the speed of data creation is even more important than the volume because by analyzing real-time or nearly real-time information makes it possible for a company to be much more agile than its competitors.

¹ Retrieved 18 July 2019 <http://www.gartner.com/itglossary/big-data/>

² <https://www.forbes.com/sites/gilpress/2014/09/03/12-big-data-definitions-whats-yours/#21c1558d13ae>



Variety

Big data takes the form of messages, updates, and images posted to social networks; readings from sensors; GPS signals from cell phones, and more. As more and more business activity is digitized, new sources of information and ever-cheaper equipment combine to bring us large amounts of hugely diverse types of data. The data available are often unstructured— not organized in a database—and unwieldy, but there's a huge amount simply waiting to be released.

Veracity

In today's era, we must also question the reliability or trustworthiness of the data. Most data is not likely to be "fake" but it can be inaccurate or incomplete. For example, often a GPS will "drift" as you travel through an urban area, meaning it is not 100% accurate to exactly where you are. This is because satellite signals are lost as they bounce off tall buildings or other structures. When this happens, location data has to be fused with another data source like road data, or data from an accelerometer to provide accurate data.

Value

Having access to big data is no good unless we can turn it into value. Examples of value deriving from data include:

- Data which helps companies to better understand and serve customers. Ej. The data collected by Netflix regarding our viewing habits helps them recommend more programmes for us.
- Data which helps optimize processes. Ej. Uber is able to predict demand, dynamically price journeys and send the closest driver to the customers thanks to the data it collects.
- Data which enables innovation. Ej Sports companies that place sensors in balls, and GPS trackers on clothes to increase value for athletes who can then analyze their own data streams and improve upon what they do.

2.2 UNDERSTANDING DATA

Data can be understood and classified according to how (and by whom) it is generated and/or by how organized or structured it is.

There are infinite types of data available to us, but just having MORE data is not always better. The quality of the data, and how relevant or valuable it is to your purposes, is just as important as how much data you have.

3. FROM BIG DATA TO SMART DATA.

We need to be realistic, pragmatic and even sceptical about what can be achieved and knowing what value can be derived and how to maximise value obtained.

- Data refers to any raw and unstructured information (text, number, media).
- Big data refers to the inflation of data quantity, sometimes impairing quality as well.
- Smart data is about creating create cost-effective and innovative ways to process large amounts of data for better insights, decision-making and process automation.³
- It can also refer to data that requires little human intervention/coordination to be useful.

(In the past, most analytics was done with batch processing. Data was collected according to schedule, converted to a desired state, put into a database and processed on an hourly, overnight or weekly basis. A drawback of this approach is that by the time the data is analyzed, it's already old. In contrast, smart data monitors data at the source, assesses them, makes a decision and shares the output -- all within a specific window of time consisting of seconds or fractions of a second.)

How you can measure if your data will be smart?

Ask questions to ensure that it will quickly and effectively lead to a better, more profitable service or product.

How to become smart? Don't rely on HiPPOs!

³ <https://securities.bnpparibas.com/insights/what-is-smart-data-2.html>



One of the most critical aspects of big data is its impact on how decisions are made and who gets to make them. When data are scarce, expensive to obtain, or not available in digital form, it makes sense to let well-placed people make decisions, which they do on the basis of experience they've built up and patterns and relationships they've observed and internalized.

However, many in the big data community maintain that companies often make most of their important decisions by relying on "HiPPO"—the highest-paid person's opinion.⁴ To be sure, a number of senior executives are genuinely data-driven and willing to override their own intuition when the data don't agree with it. But throughout the business world today, people rely too much on experience and intuition and not enough on data.

Is your business smart?

Data driven businesses are proven to be more productive and profitable. The NESTA study reveals that Datavores and Data Builders are over 10% more productive than the Dataphobes after controlling for other factors.⁵

4. BENEFITS OF DATA FOR BUSINESS

The importance of big data does not revolve around how much data a company has but how a company utilizes the collected data. Every company uses data in its own way; the more efficiently a company uses its data, the more potential it has to grow.

We see this manifested in innovative data-driven products and services, improvements in processes, and more informed decision-making across the economy and society.

Benefits can be separated into two main categories:

- **Process.** The company can take data from any source and analyze it to find answers which will enable it to make improvements in internal processes.
- **Product / innovation.** Data doesn't just create value by allowing us to do the same things better, but also by allowing us to do completely new things.

⁴ <https://hbr.org/2012/10/big-data-the-management-revolution>

⁵ <https://www.nesta.org.uk/report/skills-of-the-datavores-talent-and-the-data-revolution/>