

HOW AND WHY DO WE ACCUMULATE SO MUCH DATA?

Authors and date

- Submitted on: April 12th 2021
- Laurent Devernay ; Referent Trainers Occitania ; simplon.co

SUMMARY

By definition, data is what is known, what can be used as a basis for thinking. During the evolution of the human being, it quickly appeared essential to manage data in order to gather, compare and transmit them. This is also why the media used have evolved.

DATA EVERYWHERE

One of the initial promises of the web was to allow everyone to access information more easily and especially for free. However, the growth of data requires processing, formatting, centralization and sharing, which is both time-consuming and resource-intensive. The latter have multiplied and, according to IBM, 90% of the data produced by humanity has been produced in the last 2 years¹. And this statement has been true for more than 30 years. Except that, today, 90% of the stored data would be useless².

Imagine for example that each autonomous car collects 4TB of data per day for about 1.5 hours of driving (even if only a part of it goes through the network)³! With tens of billions of connected objects in the world also collecting and transferring data, the problem becomes significant. The environmental impact of the transfer and storage of all this information is not insignificant⁴.

THE EXAMPLE OF PHOTOGRAPHY

Photography illustrates well the evolution of our uses linked to digital technologies. The transition from argentic to digital format has been a revolution. Cost reduction (no more need for film or development), new possibilities and new needs (print your own pictures but with specific paper and sometimes printers designed for that, easy storage). With smartphones, all this has become even easier in appearance. It's no big deal to have multiple tries to get a picture right and filters can easily be added. It has never been so easy to share photos via messaging or social networks. There is little risk of losing them as they are automatically sent to the cloud by default. All this is free, but only if you omit the price of the equipment used and the actual (environmental) cost of storage. Until recently, cloud storage was mostly free of charge. Until Google realized that 4 billion photos and videos are sent to its servers every day⁵. Beyond a certain limit, the service became chargeable. More generally, dematerialization and free access have impacted our vision of the digital world. It is

necessary to keep in mind that the cloud remains a set of physical equipment and that everything we store has a cost. Not necessarily a financial cost but a cost for the planet (equipments to manufacture, to supply, to cool down and which will be regularly replaced). Except that we rely more and more on digital services in our daily life. For administrative procedures but not only. The dematerialization of sales receipts is proposed (or even imposed) as a good practice for the environment, totally ignoring the digital pollution generated (and the issue of personal data) ⁶. The trend is also spreading to invoices and other bank statements. After thorough analysis, La Poste realized that some of its promotional communications were more impactful via digital than paper ⁷.

THE CASE OF PERSONAL DATA

With the web, one type of data in particular has become more valuable: personal data. Personal data is information about a physical person. It may concern your first name or last name but also what relates to your religion, your political opinions, the programs you watch on TV, etc. On the web ⁸, this information is recovered mainly in three ways:

1. Information that you willingly give, for example when you fill out a form to make an online purchase.
2. Cookies, which are text files used by websites to store data.
3. Trackers, which are bits of code designed to capture information about the user, such as what pages they visit, what they search for on the web, etc.

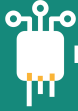
Thanks to all this, it is possible to build a profile for each Internet user. Some large companies built their economic model on the collection and resale of this data as well as on algorithms that allow, from personal data, to deduce new information about someone (if you liked this book, you will probably like this other book). Among these companies, data brokers made personal data a business in its own right ⁹.

Except that this collection of personal data is problematic. The GDPR ¹⁰ appeared in 2018 to protect european citizens. Thus the collection of data from European Internet users or by European structures can only be done with:

- the explicit consent of Internet users
- full transparency (on the nature of the data collected and the use made of it, especially if it is sold to third parties)
- a right to be forgotten (making it possible to ask a company to delete the data collected on oneself)

In addition, tools to protect oneself from trackers are emerging (Ghostery, Blacklight, etc.). In any case, it is important today to ask ourselves what happens to our data.

SOURCES



1. Bernard Marr. How Much Data Do We Create Every Day? The Mind-Blowing Stats Everyone Should Read [online], 2018. Available at [Forbes](#) [04/06/21] ←
2. Karlin Lillington. Digital pollution: 'We have a pandemic of data and 90% of it is crap' [online], 2020. Available at [irishtimes.com](#) [04/06/2021] ←
3. Brian Krzanich. Data is the New Oil in the Future of Automated Driving [online], 11/2016. Available at [newsroom.intel.com](#) [04/06/2021] ←
4. Gerry McGovern. The data pollution problem [online], 12/2020. Available from [gerrymcgovern.com](#) [04/06/2021] ←
5. Shimrit Ben-Yair. Updating Google Photos' storage policy to build for the future [online], 11/2020. Available at [Google](#) [04/06/2021] ←
6. Environmental impacts of shopping tickets. Available on the [Zéro déchet Strasbourg website](#) [04/06/2021] ←
7. Communicating responsibly, a study available on the [La Poste website](#) [04/06/2021] ←
8. Do Not Track (personalized documentary on privacy for internet users). Directed by Brett Gaylor, 2015. Available at [donottrack-doc.com](#) [04/06/2021] ←
9. Wolfie Christl. Corporate Surveillance in Everyday Life, [online], 2017. Available at [crackedlabs.org](#) [04/06/2021] ←
10. [Complete guide to GDPR compliance](#) ←