

“All things data”: a skilled workforce for a data-driven future?



THIS PAPER IS PART OF OUR DIGITAL BRIEF SERIES: A SET OF EASY-TO-GRASP DEEP DIVES TACKLING THE LATEST KEY TOPICS AROUND DIGITAL SKILLS AND JOBS. PRODUCED IN COLLABORATION WITH SOME OF EUROPE'S BEST EXPERTS IN THE FIELD.



Table of Contents

Summary	2
Keywords	2
About the author	2
Byte by byte: some context	3
Data and companies	3
Data, data, data... Which data?	3
How do companies use data - and for what purposes?	4
How do companies collect data?	4
How do companies manage data securely?	5
How do we interact with data-based IT products?	5
Mirroring Tools	6
Alerting Tools	6
Advising Tools	6
Automated decision tools	6
Which skills do we need to use these tools effectively?	7
Data in the past, and in the future	7
Skills for a data-driven future	7
Wrapping data up: some takeaways	8
References	10



Summary

As the future of work is increasingly looking more and more data-driven, some questions naturally emerge. How can we prepare our workforce to be more skilled at working with data? EU policy has followed market trends, by pooling investment for a data-driven economy. This approach foresees more data sharing within organisational ecosystems that consolidate and create new value from data. With its legal framework, the European Commission has tried to make sure all this happens in a fair, open, and secure way. But how will this affect our current jobs? Which skills will we need to participate in this data-driven economy? This article gives insight in how data is being used to define our future work.

Keywords

► Data, Data Spaces, Digital Markets Act, Digital Services Act, Data Act, Data Literacy, Sector

About the author

[Dr. Kamakshi Rajagopal](#) is an interdisciplinary researcher and freelance consultant in educational design and technology, with extensive experience in networked learning and social learning formats, supported by innovative technologies. She holds a Masters in Linguistics (2003) and Artificial Intelligence (2004) from KU Leuven (BE). She completed her doctoral research at the Open Universiteit (NL) in 2013, investigating personal learning networks and their value for continuous professional development.

Her current research is on studying the complexity of learning environments and more specifically on how teachers and learners can be supported in dealing with this complexity. Dr. Rajagopal has developed multiple (nationally-funded and European) collaborative research projects in primary, secondary and higher education with partners from the public sector, industry and civil society. Some examples of her projects are about the role of teacher networks in educational innovation, thesis circles in higher education, the multimodal measurement in collaborative hybrid learning spaces, and mainstreaming Virtual Mobility at higher educational institutions. Since 2023, she has been working on Learning and Development in IT & business consultancy.

Byte by byte: some context

Whenever we interact with our digital devices, we create data – little digital traces of numeric information that say something about our interaction with the device. This can be information about what we have clicked on, which webpages we have opened, what text we have written, which locations we have visited, or where we have made a telephone call from.

All of these digital traces together can give us comprehensive insight into our behaviour online. Data is increasingly being used to gain more insight into human behaviour for a variety of reasons. And more and more of it is being collected. [Estimates point](#) that global data volume will increase by more than 500% by 2025. Companies use data to understand their customers better and improve their products and services. Governments are using it to improve the effects of their policy decisions. Individuals can also use data to help them manage their daily tasks better. Example? Take a walk round the block and see the number of people with smart watches around you. If you use one to monitor how active you are during the day, to monitor your vital stats, or use any apps that track how often you are online, you are using data for insights.

With data storage moving to the Cloud, technically it becomes easier to bring together different types of data, as well as data coming from different sources – potentially resulting into a much more comprehensive insight. This reduces challenges on the part of companies to use this data to create more value – whether this is for themselves, others, or together with their business ecosystem partners.

Data and companies

Data is thus increasingly seen as an essential source of revenue by enterprises. But which data is useful, and how do they manage to collect this data in a secure and safe way? In this section, we dive deeper into the nitty-gritty of data becoming an asset and a commodity for companies, and how this is being regulated by policymakers in the EU.

Data, data, data... Which data?

There are different ways in which we can categorise data. Here, we discuss two categorisations that are of interest considering the broader [European Strategy for Data](#).

A first distinction concerns the connection between **data and an individual person**: personal data vs non-personal data. Personal data is any data that is (or can be) related to an individual person: a name, identity details (e.g. card numbers, etc.), health-related data, location data, etc. Some personal data can be sensitive. This is a separate category data that must be treated with extra conditions, as misuse of this data can result in hurtful, undesirable or even dangerous situations. The [General Data Protection Regulation \(GDPR\)](#) specifies sensitive data (European Commission, 2023a). Non-personal data is all other data that can be gathered. Examples include data on daily use of public transport, energy use within a home or the number of school teachers in a specified area.

A second distinction concerns those organisations that collecting data: industrial data and public data. By industrial data, we mean all types of data collected by private enterprises. As more companies create digital products, enterprises have the opportunity to collect personal and non-personal data easily which they can later use to create more value for their consumers and themselves, which can give competitive advantage. Examples of non-personal industrial data that you may know and use include the daily use of public transport (e.g. on Google Maps), or data on the energy usage in your home gathered by your energy provider. Governments also collect data to inform their policy decisions and provide digital services. Sharing of this type of data has several benefits. For example, public data could

be used by companies to create better and more relevant services and products. Individual users could improve their decision-making based on real data: e.g. deciding when to switch on their washing machine to optimise the energy consumption at home.

In other words, different types of data could be useful for different purposes to different stakeholders. This means by creating more opportunities to share data, there is potential to create more value. However, sharing comes with risks:

- ▶ Misuse of personal data
- ▶ Misuse of sensitive data
- ▶ Misuse of position by platform companies who have a large role on the Internet.
- ▶ Inequality between individual users and bigger companies

These are some of the risks that the European policy regulation is trying to manage and prevent. The most widely known regulation in effect is the [General Data Protection Regulation \(GDPR\)](#), which specifies different type of data and regulates the rights and obligations of individuals and businesses in dealing with data. The [Digital Markets Act \(DMA\)](#), adopted by the European Commission in 2023, specifies the characteristics of “gatekeepers” (online platforms owned by private enterprises that have huge and widespread access to data due to their role on the Internet) and regulates how they are allowed to use the unique data they have access to. The [Digital Services Act](#) (European Commission, 2023c) helps to safeguard all users of digital services and specifies how gatekeepers need to address any misuses of their platform (e.g. when illegal goods are being sold through their platform, or if disinformation/misinformation is being shared). The [European Data Governance Act](#) (European Commission, 2023f) and the [EU Data Act](#) (European Commission, 2023d) – currently being formulated – aim to remove barriers in data sharing, but in a way that preserves control for individuals and smaller organisations. Another priority is that data sharing can support the creation of more incentives for people to invest in data generation.

In short, European policy aims to support organisations (and especially SMEs) to invest and capitalise on their data strategies. It facilitates more fair and secure data sharing between individuals and organisations of various sizes and roles, to create more value for themselves and European citizens. Companies are therefore increasingly re-assessing their business activities to see how the data they generate can potentially create new value for their business.

How do companies use data - and for what purposes?

Companies are using data in different ways to create business value. Largely, we can distinguish two avenues:

- ▶ Using data to gain business insights into how their customers engage with their products or services. This results in sentiment analysis, understanding when and how products are used and possibly with which other products they use them. [Klee, Janson & Leimeister \(2021\)](#) term this the supra-organisational level where business value is generated by realising external data benefits.
- ▶ Using data for operational excellence to improve their internal working or supply chain management, in order to provide value to their clients efficiently and effectively. [Klee, Janson & Leimeister \(2021\)](#) nuance this under the organisational level (developing data-driven organisational models) and the work-practice level (working with data in daily business processes).

When data can be shared with others safely and securely, companies are also able to explore a myriad of opportunities related to the creation of new products or service – in an ecosystem they share with like-minded partners.

How do companies collect data?

Increasing online activities means the technical threshold of collecting data is small: data created in interactions with digital devices, or measures picked up by sensors for example, are logged and stored in cloud-based services for further analysis and interpretation. Analytics can be generated from these data to facilitate decision-making.

As mentioned, some companies act as “gatekeepers” on the Internet, where they form and manage the online platforms through which other activities take place. These gatekeepers naturally have many sources of data which they store to further commercialise and analyse for advertising purposes (think of Meta, X, and more recently, Tik Tok). This situation has created a seemingly unlimited amount of data storage, where companies end up holding onto data that could be useful for commercial use. However, this situation is likely to change – partially driven by the policy regulations mentioned above - as there are other voices coming up.

One aspect concerns **rational use and collection of data**, i.e. a more operational approach to navigating huge data volumes, with companies considering where the potential value of said data lies, and which data exactly is useful to keep track of ([Mazzei & Noble, 2017](#)). Concerns related to the environmental impact of data storage are also increasingly coming into play. According to [Lucivero, 2019](#), awareness of how much energy it takes to store data for an unlimited time frame is rising – which in turn, forces companies to consider a more rational approach to data use and storage. Finally, there is the cost impact: data storage is becoming more and more expensive. With companies moving business activities to the cloud, data storage has to go hand in hand with business acumen (Gartner, 2022).

How do companies manage data securely?

For data to become something that creates value in an ecosystem, there needs to be trust in the good and secure management of data. This requires organisations to step up to the plate in defining a vision and strategy for data collection and analytics, but also align their data architectures (IBM, 2023) to those strategy and vision. The overall strategy of European data policy is to create a true European single market for data ([European Commission, 2023e](#)), where value can be created from data within an ecosystem of organisations. There is an ambition to create sectoral data spaces, where partnerships around shared data can form and result in the co-creation and development of new and innovative applications. This requires many [practical considerations](#) for organisations in their technical architecture, such as:

- ▶ Traceability of data: when we create data spaces, the opportunity to track who collected what data and under which conditions, has to exist.
- ▶ Managing secure access to authorised partners and people: data spaces also need to foresee governance in terms of who has access to shared data and to what extent this access reaches.
- ▶ Legal framework to agree access and management of access: all parties in a data ecosystem need to have a legal cooperation structure through which they can share data in a trusted context.

If data is becoming so prevalent, this means more and more of us will come in touch with data-based IT products. Gartner predicts that this year, “data literacy will become an explicit and necessary driver of business value,” with nearly 80% of companies indicating data as a key factor in its strategic plans (Gartner, 2023a). This article gives more insight into how data can be used to guide human decision-making and which skills are demanded from us, to employ these new tools effectively to gain more insight and choose wisely.

How do we interact with data-based IT products?

We already interact with data every day through various data-based applications. We may not always recognise them as such, as they present to us this data in non-obvious ways.

Applications based on data can be categorised based on how they present this data to us.

Mirroring Tools

Mirroring tools offer the user visualisations of data on outcomes, processes, etc., through graphical methods, without interpretation of the meaning of the data. Often, these visualisations are grouped into a dashboard. A prototypical example of this are time-series data such as population growth over time or the physiological data such as ECG, where the graph shows you a visualization of the physical activity. Let us take a running example of a pollen meter. A mirroring tool for visualizing pollen levels, will present a graph where pollen levels are plotted over time (e.g. daily level of pollen in the atmosphere). It is up to the Human to read interpret the visualisation, give meaning to it (*is it a high level? Is it a low level? Is it increasing or decreasing? What is the desired level?*) and define potential interventions to act upon this information (*I will stay indoors because of high pollen levels as I suffer from allergies; I can do intensive outdoor sports because of acceptable levels of pollen*). ([van Leeuwen & Rummel, 2019](#))

Alerting Tools

In alerting tools, the visualised data will mirror real activity, but also includes minimal interpretation as remarkable elements in the data are highlighted. For example, in our pollen example, desired levels of the daily pollen levels can be made concrete in the tool itself by including a specification of what normal/high/low range of pollen levels are. The tool can then alert a human to say that a particular level has been reached (e.g. *“The pollen levels are high today”*). These alerts can be used by humans as a starting point to design interventions. The step for an intervention is taken by the human, but they are alerted by the tool to do so. ([van Leeuwen & Rummel, 2019](#))

Advising Tools

Advising tools go one step further by also recommending a particular intervention to take. For example, an advising pollen meter might say: *“don’t do intensive outdoor sports today, as the pollen levels are high and you are highly prone to allergies”*. Here, the human can choose to implement the proposed intervention, but does not need to independently interpret the data nor define an intervention based on it. The tool recommends an intervention based on its analysis of the data. ([van Leeuwen & Rummel, 2019](#))

Automated decision tools

In automated decision tools, the tool controls all decision-making and implementation of intervention itself, based on its own data analysis. In this context, the human would not be involved directly in the decision-making nor in the implementation of the intervention. In our example of the pollen meter, this could be a fictitious situation where the automation tool could say: *“the pollen levels are high today. As you have allergies, you are advised to stay home. To facilitate this, all your meetings today have been changed to web-conferencing meetings.”*

Potentially worrying (but completely fictitious) behaviour of such a tool might be *“To make sure that you do not leave the house, the doors have been automatically locked till pollen levels are at an acceptable level.”* In such a case, the human only participates in the decision-making, intervention definition and implementation process as an observer.

Which skills do we need to use these tools effectively?

With widespread use of data-based tools, it is clear that our skills and competencies in using these tools need to be up to the mark, to use them effectively, without unintentionally creating or facilitating harm to humans.

Data in the past, and in the future

In earlier days, when data was scarce and difficult to collect, big data sets were primarily created by academics at universities (or as collaborations of universities), governments or large organisations. As archiving data is intensive, longitudinal work, these were multiyear activities requiring significant investment in time, efforts and financial resources. Only companies that had the resources to invest in data archiving and saw an immediate economic interest were effectively able to do this.

Skilled people, who created these data sets specialised in bringing data together, structuring, analysing, and interpreting various data sets. This process is not that cumbersome anymore: instead, it is now supported with many sophisticated digital instrumentation and tooling. This means it is now much more accessible to organisations, whose initial investments can be lower. At the same, the situation gets more complex as a result. There is more and more up-to-date, easily accessible data: it is easy to get overwhelmed. To create value from data, we now need better visualisation techniques, and a strategic approach to information.

Since the process is no longer dependent on 1 person who gathers, analyses, and interprets the data, more coordination is needed to manage shared access to data. Increasingly, labour-intensive parts can be automated through available tooling, reducing pressure on businesses. At the same time, more automation leads to more standardised analyses. On the one hand, this creates more possibilities as more people can access the data in various ways. On the other hand, this can also be restrictive as it is likely to fall into default options and taking away deep consideration and decision-making from the user. In extreme cases, decision-making can be completely hidden.

As there is an increasing gap between data interpretation and data collection and analysis, there are possibilities for mistakes in interpretation if the context of data collection and analysis is not clearly documented and indicated.

Skills for a data-driven future

Below, I specify a number of skills that we need to build for this data-driven future.

1) Data literacy

The first essential skills are related to **data literacy**. [Gartner \(2023b\)](#) describes this as *“the ability to read, write and communicate data in context, including an understanding of data sources and constructs, analytical methods and techniques applied, and the ability to describe the use case, application and resulting value.”*

In other words, as humans engaging with data tools, we need to know and be able to assess which data is being used, where it comes from, if it is reliable, if the analytics that have been conducted on top of the data can be trusted, and if there are any human controls available, that allow me as a user to verify these issues. Especially in high level of automated decision-making and intervention by tools, these factors of control become extremely relevant for human oversight. To be able to sufficiently trust data-based tools, we as users need to have access to this contextual information and demand access to this contextual information. The policy requirements in formulation around the European Data Act (European Commission, 2023) reinforce this level of contextualisation of data.

Several programs exist to support the development of data literacy skills. For example, the [MOOC](#), funded under the Erasmus+ scheme and developed by University of Copenhagen, University of Warsaw, University of Milan, Sorbonne University, and Charles University in Prague (part of the [4EU+ Alliance](#)), gives a broad insight to how data works and the role of data in our daily lives. The [Data Spaces Support Centre](#) aims to support organisations working with data spaces in practice.

2) Awareness and deep understanding of the scope of data-based tools / working with data tools

A second skill is an awareness and deep understanding of the scope of the data-based tools by working extensively with them. This allows us as users to understand the limitations of tools, the possibilities of tools and also to assess the validity of interpretations based on the data – irrespective if they are human-made or automated. If you are more aware of what the tool can and cannot do, it will also allow you to define valid actionable insights, that lead to acceptable interventions. As a user, it is also imperative to consider data-based tools as just one of multiple data sources, and cross-checking any outcomes of these tools with other, more qualitative data.

3) Negotiation about interpretations of data

As data-based decision-making in business ecosystems becomes more prevalent, it is obvious that multiple data-based tools working around the same data will exist next to each other. This has a lot of consequences: different companies may create different tools with multiple analysis methods elicits those analytics that are important for them, in their context. This effectively means that different companies make their perspectives on the world tangible through their analytics of data.

What happens when these different multiple interpretations of the world are contrasted with each other? It is expected that negotiation on sense-making of data will become an extremely important skill. How do you engage with ecosystem partners of the scoping of different data interpretations? How do you manage differing interpretations? On which basis do you select which actions to take as interventions, and based on which (collectively accepted) data? This is currently an under-researched topic.

4) Data-Technical skills

Something that is mentioned more often in literature are the technical skills related to data, such as data engineering, data analysis and data tooling. As more data becomes locally available within organisations, employees will be able to create their own data products to support their individual tasks and processes. This requires more general knowledge and skills in being able to capture data, manage data responsibly and accurately, define analytics around this data based on sound hypotheses and implement these with sound data analysis methodologies, and using the relevant tools competently. Finally, this also requires skills to accurately interpret outcomes of data analysis and translate them into business insights for intervention.

Wrapping data up: some takeaways

Our world is increasingly becoming more data-driven.

This crosses all aspects of our work, education and labour systems, or societal groups. Data-based tools can support organisations by automating some data analytics, thereby improving decision-making and adding value. But they hold challenges.

As users, we should be aware of the scope of such tools, if we want to know how to work effectively (and efficiently) with them. Tool-makers and developers also need to meet users halfway: by improving access to information about certain tools – and in particular, giving users the means to check the reliability and validity of data, and the analytics

methods used. Improving the way we manage, collect and store data is central to the uptake of these tools, and to breaking barriers associated with their use.

And the future?

Data-driven for sure.

References

- Data Spaces Support Centre (DSSC), 2023. *Data Spaces Building Blocks – Functionality*. Available at: <https://dssc.eu/space/SK/31424817/6+Functionality> [Accessed 18 August 2023].
- European Commission, 2023. *What personal data is considered sensitive?* Available at: https://commission.europa.eu/law/law-topic/data-protection/reform/rules-business-and-organisations/legal-grounds-processing-data/sensitive-data/what-personal-data-considered-sensitive_en [Accessed 18 August 2023].
- European Commission, 2023b. *The Digital Markets Act: ensuring fair and open digital markets*. Available at: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/digital-markets-act-ensuring-fair-and-open-digital-markets_en [Accessed 18 August 2023].
- European Commission, 2023c. *The Digital Services Act package*. Available at: <https://digital-strategy.ec.europa.eu/en/policies/digital-services-act-package> [Accessed 18 August 2023].
- European Commission, 2023d. *Data Act: Commission proposes measures for a fair and innovative data economy*. Available at: https://ec.europa.eu/commission/presscorner/detail/en/ip_22_1113 [Accessed 18 August 2023].
- European Commission, 2023e. *European Data Strategy*. Available at: https://commission.europa.eu/strategy-and-policy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en [Accessed 18 August 2023].
- European Commission, 2023f. *European Data Governance Act*. Available at: <https://digital-strategy.ec.europa.eu/en/policies/data-governance-act> [Accessed 18 August 2023].
- Gartner, 2022. “Press Release: Gartner Forecasts Worldwide Public Cloud End-User Spending to Reach Nearly \$600 Billion in 2023”. Available at: <https://www.gartner.com/en/newsroom/press-releases/2022-10-31-gartner-forecasts-worldwide-public-cloud-end-user-spending-to-reach-nearly-600-billion-in-2023> [Accessed on 18 August 2023].
- Gartner, 2023a. *Infuse Enterprise Business Strategy With Data, Analytics and AI*. Available at: <https://www.gartner.com/en/data-analytics/insights/data-analytics-business-strategy> [Accessed on 18 August 2023].
- Gartner, 2023b. *A Data and Analytics Leader's Guide to Data Literacy*. Available at: <https://www.gartner.com/smarterwithgartner/a-data-and-analytics-leaders-guide-to-data-literacy#:~:text=What%20is%20data%20literacy%3F,case%2C%20application%20and%20resulting%20value> [Accessed on 18 August 2023].
- IBM, 2023. *What is data architecture?* Available at: <https://www.ibm.com/topics/data-architecture> [Accessed on 18 August 2023].
- Klee, S., Janson, A., and Leimeister, J. M., 2021. “How data analytics competencies can foster business value – A systematic review and way forward”. *Information Systems Management*, 38(3), pp. 200-217. Available at: <https://www.tandfonline.com/doi/pdf/10.1080/10580530.2021.1894515> [Accessed on 18 August 2023].
- Mazzei, M.J. and Noble, D., 2017. “Big data dreams: A framework for corporate strategy” in *Business Horizons*, 60(3), pp. 405-414. Available at: <https://www.sciencedirect.com/science/article/abs/pii/S0007681317300101> [Accessed on 18 August 2023].
- Lucivero, F., 2020. “Big data, big waste? A reflection on the environmental sustainability of big data initiatives”. *Science and engineering ethics*, 26(2), pp.1009-1030. Available at: <https://link.springer.com/article/10.1007/s11948-019-00171-7> [Accessed on 18 August 2023].

van Leeuwen, A., and Rummel, N., 2019. “Orchestration tools to support the teacher during student collaboration: a review”. *Unterrichtswissenschaft*, 47(2), pp. 143-158. Available at: <https://link.springer.com/article/10.1007/s42010-019-00052-9> [Accessed on 18 August 2023].