



Themes in data strategy: thematic analysis of ‘A European Strategy for Data’ (EC)

Graca Carvalho^{1,2} · Emre Kazim¹

Received: 1 May 2021 / Accepted: 20 September 2021 / Published online: 27 October 2021
© The Author(s) 2021

Abstract

In March 2021, the European Commission announced Europe’s Digital Decade (Europe’s Digital Decade: Commission sets the course towards a digitally empowered Europe by 2030. European Commission Press Release. Access on https://ec.europa.eu/commission/presscorner/detail/en/ip_21_983). Here the Commission sets the course towards a digitally empowered Europe by 2030. In February 2020, the European Commission published ‘A European Strategy for Data’ (European data strategy: Making the EU a role model for a society empowered by data. European Commission, February 2020. Access on: https://ec.europa.eu/info/strategy/priorities-2019–2024/europe-fit-digital-age/european-data-strategy_en (2020)) as part of a wider drive concerning digital transformation and policy. In this article, we analyse the publication as it touches on broader themes ranging from digital literacy, to cloud infrastructure and artificial intelligence. Within this context, in this article, we use the EC publication as a point of departure to explore themes central to national and international digital transformation and policy writ large. As such, this article is to be read as a thematic analysis rather than a close reading of the EC’s publication. The article is divided into three parts: an executive findings and recommendations section (where our main findings are articulated); Themes and Key Takeaways (where we thematically flesh out the document); and, EU Data Strategy document summary (where we provide an overview summary the document itself).

Keywords Data strategy · Data governance · Infrastructure · Emerging technologies · Ethics · Artificial intelligence · IoT · Blockchain · Digital literacy · Fairness · Digital transformation

1 Introduction

In March 2021, the European Commission announced Europe’s Digital Decade [1]. Here the Commission sets the course towards a digitally empowered Europe by 2030. This vision builds upon the Commission’s digital strategy as published in February 2020. In February 2020, the European Commission published ‘A European Strategy for Data’ (referred to in this article as ‘EU data strategy’) as part of a wider drive concerning digital transformation and policy. We read digital policy in broad terms to encompass the

entire range of policy concerns—from infrastructure planning to competition law, data protection, digital literacy, etc. Indeed, we also understand such policy in terms of geopolitical strategy and economic growth. In this article, we will analyse the February 2020 publication as it is a more detailed document which touches on broader themes ranging from digital literacy to cloud infrastructure and artificial intelligence. Within this context, in this article we use the EC publication as a point of departure to explore themes central to national and international digital transformation and policy writ large. As such, this article is to be read as a thematic analysis rather than a close reading of the EC’s publication.

The article is divided into three parts:

- Part 1: Executive findings and recommendations: in this section, from our thematic reading of the EU data strategy, our main findings are articulated. Here we provide our highest level commentary.

✉ Graca Carvalho
graca.carvalho@ucl.ac.uk

Emre Kazim
e.kazim@ucl.ac.uk

¹ Department of Computer Science, University College London, London, UK

² Centro Algoritmi, Universidade do Minho, Braga, Portugal

- Part 2: Themes and key takeaways: in this section, we thematically flesh out the document and dig deeper into themes we find most pertinent.
- Part 3: In this section, we provide an overview summary of the EU data strategy document.

The intended audience for parts 1 and 2 are readers from policy, governance, industrial strategy, digital ethics, and law. Those interested in a succinct overview or to familiarise themselves with the document will find Sect. 3 most useful.

2 Part 1: Executive findings and recommendations

From our thematic reading of the EU data strategy, our main findings are:

1. Greater vision needed: whilst commending the EU for this vision document we find that there is a lack of vision and boldness. An example of blue-sky thinking is discussion of ‘Destination Earth’ (digital twin of the Earth), which is bold, positive and a clear target for the EU to be a global leader in. In general, we believe that at the level of nations, data strategy (digital strategy) requires as much thinking about the practicalities (standards, infrastructure, etc.) as it does about the kind of country and/or union that is desired (e.g. what about digital ID, what about fully integrated systems, etc.). Estonia’s vision drove its remarkable transformation to a digital hub of the world (e-Estonia1) and we argue that such confidence is lacking in this document.
2. Prioritise data standardisation: If data are to be considered an economic asset, it must be carefully and completely described: data formats, data quality metrics, data usability conditions, data sources qualification, and other data properties must be collected in the metadata. Equally important is understanding data value, data use terms and conditions and to create a common taxonomy. Quantitative metrics to assess potential of data are needed. Risk assessment models are also needed [2].
3. Data strategy in parallel/in situ with AI, blockchain and IoT strategy: as we have raised on several occasions, we read ‘data’ strategy as part of broader concept of digital strategy, which includes digital transformation and the utilisation of all novel computer science-based technologies (AI, blockchain, IoT, etc.). It appears the current thinking is sequential, where there is digital transformation, followed by a data strategy, after which perhaps AI strategy will proceed. The disadvantage of this is that the full potential benefits of the benefits of those ‘subsequent technologies’ will be severely curtailed because of not fully incorporating them into the thinking of the data

strategy. For example, the debate about standardization should be fully informed and premised on the idea that the standards will facilitate IoT, AI technologies, etc.

4. Ensure that data spaces are comprehensive: Training datasets and testing datasets that reflect EU principles and values like privacy and consumer protection. To achieve that the availability of training and testing datasets that prevent bias and discrimination is essential.
5. Invest in technological solutions that allow easy, transparent, and fair sharing of value generated by data amongst stakeholders: smart contracts, blockchain, Distributed technologies, all contributing to an automated, fully digital, distributed solution is needed to support Digital Value Chains. This is an area still not addressed in a satisfactory way.
6. Promote the creation of ecosystems of excellence and trust: The approach to create strong public–private partnerships that can co-create architecture reference models, define terminology, and develop business models should be the preferred path. The sustainability of assets created by funded projects is as important as the creation of the assets themselves. This is the only way to guarantee that investments have future impact. The example of the creation of the Alliance for Internet of Things Innovation [1] in 2015 [aioti.eu] shows how an ecosystem can exploit the results of previous projects and programmes and act as sustainability agents for future use and continued development connecting researchers, developers, service providers and final users.

3 Part 2: Themes and key takeaways

From our reading of the EU data strategy, we note several important themes and draw key findings from these themes. We summarize these themes and key findings in Table 1 and expand upon each theme in the subsequent subsections.

3.1 Disambiguating data

The document states, ‘Data is the lifeblood of economic development’ and the strategy is built around the concept of data as an economic asset. However, the landscape of data is far from simple and there are several dimensions to be considered (listed below). *A central recommendation is to move from the broader ‘data’ to disambiguate the various uses of the term.* This recommendation recognises that conceptual confusion and impreciseness leads to friction in making the most of the various taxonomy of data applications and potential uses.

- Public data versus private data: Data can be created in the context of Public and Governmental projects and activi-

Table 1 Summary of key takeaways from thematic analysis of EU data strategy

Theme	Key takeaway
Disambiguating data	A central recommendation is to move from the broader ‘data’ to disambiguate the various uses of the term.]
Infrastructure	It is critical that infrastructure is understood in holistic fashion to include the physical and cyber realms, and that plans are drawn with each part of the infrastructure and their interaction taken in tandem. [See part 1 Recommendations 3, 4, 6]
Lopsided development	Similar to a national industrial strategy the data strategy will require significant investment over long periods of time. [See part 1 Recommendations 2, 4, 6]
Integrated Technologies	Data should encompass ‘digital’ technologies of Blockchain, the Internet of Things, and Artificial Intelligence: these technologies must be at the forefront of how the data are collected, stored, and accessed. [See part 1 Recommendations 3, 4]
Governance	We argue that questions of governance are perhaps the central questions of ensuring data are managed appropriately and ultimately for the public good. [See part 1 Recommendations 2, 3, 4, 5, 6]
Data-driven policy	Data strategy should be discussed in terms of data as a driver for testing and creating policy itself i.e., to make big data analytics a central component of the data strategy. This is also relevant when thinking about issues of consent and fairness. [See part 1 Recommendations 2, 4, 5]
Upskilling	Actioning data strategy requires upskilling [See part 1 Recommendations 4, 6]
Empirical examples	We believe that Estonia is a paradigmatic example of coherent and successful data strategy and governance, and therefore encourage a thorough and detailed study to find what is scalable from the Estonian example. [See part 1 Recommendations 2]
Vision	Setting the agenda is critical for the data economy—it is crucial to have a vision with respect to data. [See part 1 Recommendations 1, 3, 6]

ties or by private corporations, through their services and platforms. Currently, a small number of Big Tech firms hold a large part of the world’s data [3],

- Open data versus closed data: Data access can span from Open Data, typically associated with data capture through publicly funded means to closed Data where private organisations limit the access to datasets through licensing and contractual agreements;
- Single owner versus co-created data: Datasets can be combined to create other datasets and many times these can be originated from different stakeholders. Shared ownership of data implies clear data provenance mechanism and transparency in sharing potential value created;
- Data Sources: Data can be originated in a multitude of contexts: Data from things, data from people, data from the physical world.
- Data attributes: Datasets can differ considerably in terms of volume (the quantity of data generated or consumed), variety (in terms of formats and multidimensionality of data fields), velocity (in terms of frequency of data generation and/or frequency of data delivery), veracity (inherent unpredictability and quality of data might require considerable analysis prior to use), and value (the extent to which data has the potential to generate economically worthy insights) (Fosso Wamba et al. [4]. Work on standardizing data attributes that can be part of metadata.

A case study in the need to disambiguate the term ‘data’ is in the discussion of the “Data-agile economy”. In the

period 2021–2027, the Commission will invest in a High Impact Project on European data spaces, this is an initiative directed at the creation of vertical data spaces across Europe. We welcome this, however, the participation of private sector organisations and the creation of operational models that support public and private stakeholders’ participation is fundamental [5]. Also important is that the pathway to production is designed from the outset of these projects so that continuity of availability is not impacted. Indeed, here interoperability is critical. The European Interoperability Framework is pointed as the initiative that will address the challenge around data interoperability and data quality. The ISA2 initiative DCAT-AP [6] has been implemented by 12 countries in Europe for public data exchange in Government-to-Government context however it is not clear how the business community is participating in such initiatives. The approach to create strong Public–Private–Partnerships that can co-create architecture reference models, define terminology, and develop business models should be the preferred path. The strategy mentions that organisations contributing data would get a return in the form of increased access to data of other contributors, analytical results from the data pool, services such as predictive maintenance services, or licence fees. However, no detail is provided in terms of the possible technological platforms that will enable such return to be shared amongst the stakeholders that have contributed with different data assets. Novel approaches to owner-centric sharing of data and data-driven insights are emerging [7] and should be explored. As such, even when financial

incentive, it is clear, without clearly defining terms, it will almost guarantee that such interoperability will fail.

The strategy recognised the importance of identifying incentives to stimulate data sharing. The World Economic Forum addresses this topic in different contexts [8] and we recommend that incentives are defined taking into account the different stakeholders and actors that participate in the data ecosystem: individual citizens, who could be incentivised through better prices and/or offers or improved services; the public sector, sensitive to incentives related to improvements in service offering and increased efficiency in expenditures; and lastly, the private sector, that could be incentivised through the benefits of understanding better the customer base and the ability to predict trends [9].

By way of concluding this theme, as point of principle it is important to reiterate that the funding of architecture creation, governance mechanisms and business models' definition is welcome. However, funding of data-sharing tools and infrastructures that need to be operational and economically viable should be governed by a different set of mechanisms tied to migration to operation and commercial pathways.

3.2 Infrastructure: as physical, as cyber, as interrelated

Data in the context of issues related to infrastructure are discussed. Crucially infrastructure as it relates to data can be fleshed out in several ways:

- **Physical infrastructure:** here, where data are physically stored (servers, data centres, personal data stores) is at issue. In this case, there are factors such as where the physical infrastructure will be located (in the context of the EU this could be via central 'data vaults' located in a few locations in Europe, or each country having its own data vault organised in a federated manner), the security of the locations, the integration of the locations with other communication/networking infrastructure (both national and international), and mechanism for fail safes i.e. fall back plans if and the infrastructure is compromised. A corollary to this is the environmental dimension. Here not only the energy consumption—and how any data centre would be integrated with national/international energy grids—but also issues to do with cooling and protection against natural disaster must be considered. A loose analogy with nuclear sites can be made, where a breach in the infrastructure has the potential to lead to widespread social and economic consequences.
- **Cyber infrastructure:** here the manner in which data are stored and granted access too is at hand. Concerns relate to what is the most feasible technological solution to securing data (below we touch on cloud and blockchain solutions). Within the context of the EU, we note that

Gaia-X [10] 'a federated data-infrastructure for Europe', whose aim was to address 'Data spaces should foster an ecosystem (of companies, civil society and individuals) creating new products and services based on more accessible data' was supposed to have prototypes for early 2021 yet nothing has come of this. More generally, we note that in the EU data strategy great emphasis is placed on infrastructure in terms of cybersecurity, and that physical security of infrastructure is not discussed with the same amount of treatment.

As a general point we believe that prioritizing questions of physical infrastructure at earlier stages can protect against a scenario where infrastructure is effectively monopolised and 'abroad'—which is likely to be the case in the coming decade.

The evolution towards more decentralized infrastructure models is recognised as a paradigm shift. The report mentions "cloud" very broadly and we think that some finer analysis is needed to understand the phenomenon and the opportunities and challenges that might arise in this domain. The report mentions that emerging Edge Computing capacities—a set of enabling technologies that move data storage, computing and networking closer to the point of data generation or consumption and away from a centralized computing location—should be considered from the start but does not develop the topic [see aioti.eu]. Indeed, edge computing will open business opportunities for local providers that can offer local resources and services, as long as interoperability with central cloud providers is offered. Another aspect to consider is that the infrastructure design for edge computing will be tied to the business application and therefore it is fundamental that the cloud services marketplace addresses the bespoke services that will be required to support adoption and implementation.

Lastly, given the accelerated pace of the cloud offering market, the proposed timeline, Q4 2022, seems rather late to secure leadership in the cloud infrastructure market or even for a common European standards and requirements for the public procurement of data processing services, given that procurement of cloud services will continue to accelerate. As of February 2021, Canalys [11] reports that the worldwide cloud market grew 32% that quarter to \$39.9 billion. For the full year of 2020, cloud infrastructure spending grew 33% to \$142 billion. AWS has 31% of the market, followed by Azure at 20%, Google at 7%, Alibaba Cloud close behind. The top four cloud providers account for 65% of the total cloud spend in Q4 2020. The strategy document does not address how the investment in federated cloud infrastructures will change the current landscape in terms of cloud services commercial offering.

In closing, we note that infrastructure has significant ethical implications, regarding not only safety and security, but

also issues of national and individual autonomy, privacy, and fairness.

3.3 Lopsided development

Data strategies and digital strategies more generally, are aspirational in nature. Read as vision documents they can drive programmes of development and indeed legislation. However, when taken as a vision-statement, implicitly a commitment is being made by the purveyor of the vision statement (this is particularly true when the strategy is articulated from a national/international governance structure such as the UK government or the European Union). Similar to a national industrial strategy—indeed digital/data strategy can be read as an/the industrial strategy—the data strategy will require significant investment over long periods of time. As a multinational union, the EU is well placed to direct and lead in the space of data; however, unlike the nation state, the EU will have to contend with national autonomy of member states. This raises the broader issue of lopsided development, where in the case of nation states regions are often unequally developed, which often leads to targeted investment from central authority. The reason why this is relevant in data strategy is because it will require standardization and regulation (imposing legislative agendas across regions and nations, thereby challenging autonomy) and integrating physical infrastructure across nations (again, challenging autonomy). When viewed as a resource, ‘data autonomy’ of nations and communities becomes geopolitical and constitutional concerns. In other words, data strategy will change the way nations are nations, and there are real dangers that less developed nations will lose autonomy.

3.4 Technologies: blockchain, IoT and AI

The EU data strategy hinges on the ‘data’ as a driver for economic growth; however, the report recognises that the functioning of the European data space will depend on the capacity of the EU to invest in next-generation technologies and infrastructures. Also important is to understand how data are transformed into Information, Knowledge and finally into Actionable Intelligence [12]. There is a broader question regarding ‘digital strategy’, which can often be confused with data strategy. We use the term ‘digital’ loosely to encompass technologies that have primarily emerged out of research and developments in computer science, these include the digital technologies of Blockchain, the Internet of Things, and Artificial Intelligence. In this umbrella notion of ‘digital’ data can be read as a powerful enabler for these other technologies, however, we argue that for data to be harnessed for these technologies, these technologies must be at the forefront of how the data are collected, stored, and accessed.

With respect to the EU data strategy.

- Blockchain: is only briefly discussed as a passing note (in the form of a highlighted text box) rather than a deeply considered suggestion. This technology is of particular interest because it may be a solution to issues of cyber security and physical security. More broadly the use of Distributed Ledgers can support the traceability of ownership of data assets and therefore be used as a value sharing infrastructure.
- IoT: is mentioned in terms of its use in Industrial settings and Smart City contexts. Here some standout points are:
 - Calls for standardisation: for the data to be useful it must be standardised. Such standardization can be at the point of collection, storage, or access (or all these points). We believe that this is perhaps the central barrier in realising the benefits of IoT (Taylor et al. [13] and for this reason, we argue that the treatment in the EU data strategy should warrant a separate treatment in itself.
 - Public sector: The role of the public sector as a catalyst for change in usage rights of co-generated data that derives for IoT data collected from public assets should be explored and the leading by example approach continue to be promoted.
 - Digital twin cities: IoT data are mentioned in terms of the data that the built environment can provision for the enabling of smart cities. A particular stand-out from the EU data strategy is ‘Destination Earth’ (digital twin of the Earth) initiatives, which will enable research and simulation on a global scale (c.f. climate change modelling, national and continental modelling).
- Artificial intelligence: we note that throughout the report data as enabling AI is commented upon but always briefly. We believe that this is a critical gap, and that innovation in AI, where know-how and high-quality (labelled) data has the potential to position a nation/economic block, at the forefront of the global economic community [14, 15]. The availability of training datasets, with specific control mechanisms to prevent bias and discrimination, is as important as the availability of datasets to allow the harvesting of insights and analytics and ultimately to create innovation. Equally important are technologies that can be more economically viable and that allow re-use of trained models, ex. use of Transfer Learning and Meta Learning paradigms.

Highly related to the above is the EU’s ‘Report on the safety and liability implications of Artificial Intelligence, the Internet of Things and robotics’ (February 2020). In this

report, the terms safety and liability are synonymous with safety for customers and liability for product developers and service providers. Indeed, the document is read as consumer centric, with risk defined in terms of mental and physical harm to users—principally individuals (e.g. digital addiction is discussed). We note this with particular interest as the data and AI ethical consequences for failing to appropriately address technology integration can have consequences in terms of lack of development and loss of opportunity for nations and welfare.

3.5 Governance, legislation and ethics

‘Governance’ is raised in the EU data strategy—which we read in terms of ‘data governance’—however it is raised in only a cursory manner. We argue that questions of governance are perhaps the central questions of ensuring data are managed appropriately and ultimately for the public good. This is interesting given that work on data governance has been on-going for some time now [16]—admittedly in terms of data governance at a company/institutional level. Indeed, when talking about the single market, the EU data strategy spoke about ‘legal parity’, which is a notion desperately in need of fleshing it. For example, what about parity in terms of contributing data, and infrastructure, and security?

In particular is the issue of ethics—contributing data with parity is needed for issues of fairness in the context of data driven policy and training sets for machine learning technologies. In this context there is a need to develop audits and reporting mechanisms (as is being done in the case of AI and data use) [17] that accounts for more than data privacy (as exemplified by GDPR) [18, 19].

EU data strategy is strongest when it provides a sustained treatment regarding the enablers of competitive, secure, and fair European cloud services (Sect. 5.2). Here, a ‘cloud rulebook’ covering a compendium of existing cloud codes of conduct and certification on security, energy efficiency, quality of service, data protection and data portability is listed. While this cloud rulebook has a strong role in increasing the transparency of cloud services offered in the EU, it is not clear if that will be sufficient to increase the uptake of cloud services by SMEs. Another important aspect to consider is the need to distinguish different cloud offerings so that the barrier to entry to new edge cloud service providers is not increased to a disproportionate level.

3.6 Data-driven policy

In the EU data strategy, there is little discussion in terms of data as a driver for testing and creating EU policy itself, i.e., to make big data analytics a central component of the data strategy. In order to do this, the EU would have to demarcate the domains/jurisdictions within which the policy

decision-makers would be acting. This is relevant when thinking about issues of consent and fairness [18, 20]; consent is a concern because if data are driving policy then in part when it is collected it is done so for policy and as such, people should be aware of this; fairness is a concern because the data that is driving policy should reflect the people it is ultimately going to effect.

3.7 Upskilling

In addition to data, a highly skilled workforce will ensure competitiveness. This is mentioned but more is focused on digital literacy, which we read in terms of citizens being able to navigate the basics of the digital works (including things as basic as searching the web and accessing information). We argue that a two-pronged approach is needed.

- General digital literacy: Here the concern is that a minimal level of competency and awareness is ensured through educative programmes for all people. This may become as basic as numeracy and literacy.
- World leading research: here the concern is that any data strategy will require a constituency within the community of highly competent, world leading experts, in data structure, security, governance etc. We note that this is an issue of skills and innovation rather than one of pulling the relevant legislative leavers.

Consequently, there is an ethical drive and imperative to get this right. Namely, citizens have a right to certain public goods and resources, as well as the ability to engage in civil society and the democratic processes—it is likely that digital competency will ground much of the future of civic participation.

3.8 Empirical example—Estonia

Estonia is not mentioned in the EU data strategy. We believe that Estonia is a paradigmatic example of coherent and successful data strategy and governance [21], e-Estonia [22–24]. As such, we encourage a thorough and detailed study. Indeed, there is ample research and communities of experts to draw from. The case of Estonia is of course specific to its population size, economy, existing infrastructure, etc. (e-Estonia2), so the key lessons to be sought must concern what is scalable from the Estonian example.

More generally, there is a dearth with respect to empirical studies more generally. The Estonia example is instructive because it has real-world lessons and expertise. It also poses an alternative route to digital transformation to the one implied by the EU data strategy, namely one that starts with the nation and expands to multinational union, i.e., a bottom-up approach. Although in the EU data strategy asserts such

as this are made ‘the Commission deliberately abstains from overly detailed, heavy-handed ex ante regulation, and will prefer an agile approach to governance that favours experimentation (such as regulatory sandboxes), iteration, and differentiation’, we challenge the level of abstraction and argue strongly for empirical case studies and experiments. The policy should be to ‘try and test, and then expand’.

3.9 Vision—data as resource, data as governance

Setting the agenda is critical for the data economy. Indeed, notwithstanding that much of the criticism we offer in our above comments is that major issues are skirted over, it is nonetheless crucial to have a vision with respect to data. In this respect the EU data strategy is commendable, and places it ahead of other major jurisdictions (including the UK and the USA). However, there are some issues with respect to how the EU data strategy seems to envision the EU on the global and international stage. For one, it appears to be suggested/implied that the EU can be ‘the’ data space—similar to the manner in which GDPR became de facto ‘the’ privacy provision, there is a strong hint that the EU sees itself as well positioned to achieve the same in the data space.

We believe that there is a danger in the EU’s confidence that data in terms of the EU being the ‘safe’ or ‘trusted’ data space, thereby potentially facilitating global data sharing via the EU as the trusted party (following something akin to the experience of GDPR, where the EU’s data privacy provisions became the global standard and thereby empowered the EU to be the global leader in this space) is ungrounded. We believe that there is an asymmetry here: in terms of GDPR there was a framework of protection (so it was easy for others to adopt) but data are a product and resource itself (so much less likely to be adopted by others) as this would be akin to the sharing of national resources.

4 Part 3: Summary of EU data strategy

In this section, we summarise the EU Data Strategy document [26]. Those interested in a succinct overview of the document will find it here.

1. Introduction

The publication begins by noting the ‘enormous’ benefits that data-driven innovation will bring for citizens and immediately turns to calling for ‘the interests of the individual first, in accordance with European values, fundamental rights and rules. Following this, data as a source of economic growth is spoken of in terms of a ‘data-agile economy’ and the vision that “the EU can become a leading role model”.

2. What is at stake?

The significance of an appropriate data strategy is then discussed. Highlighted are the following:

- Growing data volumes and technological change: there is de facto a massive increase in collected data and innovations in how the data are generated, collected, processes, stored and secured.
- The importance of data for the economy and society: here public services are mentioned, as well as economic development (‘Data is the lifeblood of economic development’) and achieving the European Green Deal.
- The EU and the data economy of the future: here the geo-economic factors are discussed in terms of keeping the EU competitive, challenging the seeming data monopoly of ‘Big Tech’ firms, and high lighting that in addition to data a ‘highly skilled workforce’ will guarantee competitiveness.

In terms of work that has already been done and started that can be built upon, the section closes by surveying actions that the EC has taken since 2014. These are: GDPR, regulation on the free flow of non-personal data (FFD), the Cybersecurity Act (CSA), Open Data Directive, Sector-specific legislation on data access such as automotive, payment service providers, smart metering information, electricity network data, or intelligent transport systems, and, The Digital Content Directive [See: Regulation (EU) 2018/1807; Regulation (EU) 2019/881; Directive (EU) 2019/1024; Regulation 715/2007 as amended by Regulation 595/2009; Payment Service Directive 2015/2366; Directive 2019/944 for electricity, Directive 2009/73/EC for gas meters; Commission Regulation (EU) 2017/1485, Commission Regulation (EU) 2015/703; Directive 2010/40/EU; Directive (EU) 2019/770.)

3. The vision

A bold vision is stated: The aim is to create a single European data space—a genuine single market for data, open to data from across the world—where personal as well as non-personal data, including sensitive business data, are secure and businesses also have easy access to an almost infinite amount of high-quality industrial data, boosting growth and creating value, while minimising the human carbon and environmental footprint.

There is much to this vision; however, in particular we note the phrase ‘open to data from across the world’ as an attempt to establish the EU as the global data space. This is augmented with the cursory suggestions regarding what benefits those who are within this space will gain, namely ‘organisations contributing data would get a return in the form of increased access to data of other contributors, analytical results from the data pool, services such as predictive maintenance services, or licence fees’. Indeed, it is noted

that the vision is ‘needs to be complemented with a broader industrial strategy for the data-agile economy’.

The problems

The publication highlights several problems that hold the EU back from realising its potential in the data economy:

- Fragmentation between member states: this is mentioned in terms of legal framework.
- Availability of data: The value of data lies in its use and re-use. This includes use of data in terms of:
 - Use of public sector information by business (government-to-business—G2B—data sharing).
 - Sharing and use of privately-held data by other companies (business-to-business—B2B—data-sharing). The argument is made that this is due to a lack of economic incentives (including the fear of losing a competitive edge), lack of trust, imbalances in negotiating power, the fear of misappropriation of the data, and a lack of legal clarity (ex. for co-created data, in particular IoT data).
 - Use of privately-held data by government authorities (business-to-government—B2G—data sharing). Here recommendations of an Expert Group created by the Commission, include creation of national structures for B2G data sharing, development of appropriate incentives, and the suggestion to explore an EU regulatory framework to govern the public sector’s re-use for the public interest of privately held data.
- Imbalances in market power: here it is pointed out that there are market imbalances in relation to access to and use of data (acutely impacting SME), which is in addition to the concentration in provision of cloud services and data infrastructures.
- Data interoperability and quality: here standardisation is sought (c.f. ICT standardisation (<https://ec.europa.eu/digital-single-market/en/news/rolling-plan-ict-standardisation>) and a strengthened European Interoperability Framework (https://ec.europa.eu/isa2/eif_en; see COM (2017)134 final).
- Data governance: here we note with curiosity the fact that this subject is given only a few sentences. We believe this is a core theme.
- Data infrastructures and technologies: Here the cloud is discussed as the principal infrastructure and technology. However, several problems from both the supply and demand side of the cloud are mentioned:
 - On the supply side:
 - EU-based cloud providers have only a small share of the cloud market
 - Service providers operating in the EU may also be subject to legislation of third countries. One concern with this is that it generates uncertainty over legal compliance with applicable EU law.
 - Uncertainty about compliance of cloud service providers with important EU rules and standards, for example on data protection.
- On the demand side:
 - Low cloud uptake in Europe (1 company in 4, only 1 in 5 for SMEs). Significant divergences in cloud uptake exist between Member States.
 - Specifically, cloud uptake in the European public sector is low.
 - There is frequently insufficient visibility on the market of smaller, often European, providers of innovative cloud services.
 - European businesses often experience problems with multi-cloud interoperability, in particular data portability.

Empowering individuals to exercise their rights: Individuals suffer from the absence of technical tools and standards that make the exercise of rights simple and not overly burdensome (c.f. MyData movement, etc. <https://mydata.org/>; <https://www.decodeproject.eu/>; <https://solid.mit.edu/>, <https://radicalxchange.org/>).

Skills and data literacy: ‘Currently, big data and analytics are top of the list of critical skills shortages’.

Cybersecurity: Here the EU Cybersecurity Certification Framework and the EU Agency for Cybersecurity (ENISA—Regulation (EU) 2019/881—European Cybersecurity Act) are read as critical. Following this, importantly, a note is added regarding the potential use of *blockchain* in the management of data—we note the brevity with which this is discussed and suggest that a far lengthier treatment would have been appropriate.

The strategy

Here four pillars are discussed:

A cross-sectoral governance framework for data access and use: This seeks to facilitate data across borders, mechanisms of decision-making, interoperability and standardisation.

- Making more high-quality public sector data available for re-use, in particular in view of its potential for SMEs. It is hoped this will drive innovation (c.f. Open Data Directive)
- Explore the need for legislative action on issues that affect relations between actors in the data-agile economy to provide incentives for horizontal data sharing across sectors (cf. Data Act (2021)).

Enablers: Investments in data and strengthening Europe’s capabilities and infrastructures for hosting, processing, and using data, interoperability.

- In the period 2021–2027, the Commission will invest in a High Impact Project on European data spaces and federated cloud infrastructures.
- The project will fund infrastructures, data-sharing tools, architectures, and governance mechanisms for thriving data-sharing and Artificial Intelligence ecosystems.
- This project needs to be seen in the context of a wider set of strategic EU investments in new technologies that the Commission will present in March 2020 as part of its industrial strategy.
- High-impact project: developing common European data spaces and interconnecting cloud infrastructures.
- Establishment of EU-wide common, interoperable data spaces in strategic sectors. The spaces will include (i) the deployment of data-sharing tools and platforms; (ii) the creation of data governance frameworks; (iii) improving the availability, quality, and interoperability of data—both in domain-specific settings and across sectors.
- Enabling access to competitive, secure, and fair European cloud services
 - Seeking a coherent framework around the different applicable rules (including self-regulation) for cloud services, in the form of a ‘cloud rulebook’.
 - In a first instance, the cloud rulebook will offer a compendium of existing cloud codes of conduct and certification on security, energy efficiency, quality of service, data protection and data portability. In the area of energy efficiency, earlier action will be considered.
 - Development of common European standards and requirements for the public procurement of data processing services.
 - The set-up of a cloud services marketplace for EU users from the private and public sector will be facilitated by the Commission by Q4 2022.

Competences: empowering individuals, investing in skills and in SMEs.

- Empowering individuals with respect to their data.
- Investments in skills and general data literacy (c.f. Digital Education Action Plan).
- Dedicated capacity building for SMEs: e.g. start-ups often require legal and regulatory advice to fully capture the many opportunities ahead from data-based business models.

Common European data spaces in strategic sectors and domains of public interest.

Building on the European Open Science Cloud, nine common European data spaces are listed:

- **Industrial (manufacturing):** supporting the competitiveness and performance of the EU’s industry, allowing to capture the potential value of use of non-personal data in manufacturing (estimated at € 1,5 trillion by 2027).
- **Green Deal:** to use the major potential of data in support of the Green Deal priority actions on climate change, circular economy, zero-pollution, biodiversity, deforestation, and compliance assurance. The “GreenData4All” and ‘Destination Earth’ (digital twin of the Earth) initiatives will cover concrete actions.
- **Mobility:** to position Europe at the forefront of the development of an intelligent transport system.
- **Health:** essential for advances in preventing, detecting, and curing diseases as well as for informed, evidence-based decisions.
- **Financial:** to stimulate innovation, market transparency, sustainable finance, as well as access to finance for European businesses and a more integrated market.
- **Energy:** promote a stronger availability and cross-sector sharing of data, in a customer-centric, secure, and trustworthy manner.
- **Agriculture:** enhance sustainability performance and competitiveness of the agricultural sector.
- **Public Administration:** to improve transparency and accountability of public spending and spending quality, fighting corruption, both at EU and national level, and to address law enforcement needs and support the effective application of EU law and enable innovative ‘gov tech’, ‘reg tech’ and ‘legal tech’ applications.
- **Skills:** reduce skills mismatches between the education and training system on the one hand and the labour market needs on the other.

6. An open, but proactive international approach

We note significant strategic statements, namely:

- Building upon the strength of the Single Market’s regulatory environment, the EU has a strong interest in leading and supporting international cooperation regarding data, shaping global standards, and creating an environment in which economic and technological development can thrive, in full compliance with EU law.
- At the same time, European companies operating in some third countries are increasingly faced with unjustified barriers and digital restrictions.
- This would allow the EU to have an open but assertive international data approach based on its values and strategic interests [...] facilitating international data flows.

- The EU should take advantage of its effective data regulatory and policy framework to attract the storage and processing of data from other countries and regions, and to increase the high-value-added innovation that arises from these data spaces. Companies from around the world will be welcome to avail of the European data space, subject to compliance with applicable standards, including those developed relative to data sharing.

We believe that there is a danger in the EU's confidence that data policy/sharing is going to follow the experience of GDPR, where the EU's data privacy provisions became the global standard and thereby empowered the EU to be the global leader in this space. We believe that there is an asymmetry here: in terms of GDPR there was a framework of protection (so it was easy for others to adopt) but data are a product and resource itself (so much less likely to be adopted by others).

Appendix

The following is a list of the Common European data spaces in strategic sectors and domains of public interest found in the report.

- Common European industrial (manufacturing) data space
- Common European Green Deal data space
- Common European mobility data space
- Common European health data space
- Common European financial data space
- Common European energy data space
- Common European agricultural data space
- Common European data spaces for public administrations
- Common European skills data space
- European Open Science Cloud

Funding The authors would like to acknowledge Cisco Research Centre for their research Grant (2020-222054 (3696)) that partially funded this research.

Availability of data and material Not applicable.

Code availability Not applicable.

Declarations

Conflict of interest The authors have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Europe's Digital Decade: Commission sets the course towards a digitally empowered Europe by 2030. European Commission Press Release. Access on https://ec.europa.eu/commission/press-corner/detail/en/ip_21_983
2. Kannan, K., Ananthanarayanan, R., Mehta, S.: What is my data worth? From data properties to data value. arXiv:1811.04665 [Cs]. <http://arxiv.org/abs/1811.04665> (2018)
3. Takala, V.: What are the alternatives to either big corporations or big states owning our data? UCL Institute for Innovation and Public Purpose. Blog. Access on <https://medium.com/iipp-blog/what-are-the-alternatives-to-big-corporations-or-big-states-owning-our-data-b76532e2fed1> (2019)
4. Fosso Wamba, S., Akter, S., Edwards, A., Chopin, G., Gnanzou, D.: How 'big data' can make big impact: findings from a systematic review and a longitudinal case study. *Int. J. Prod. Econ.* **165**, 234–246 (2015). <https://doi.org/10.1016/j.ijpe.2014.12.031>
5. Carvalho, G., Ritchie, A.W., Fradkin E.D.: Innovation ecosystems: Greater than the sum of their parts? In: *Living in the Internet of Things (IoT 2019)*. IET (2019)
6. ISA2: Access on: https://ec.europa.eu/isa2/solutions/dcat-application-profile-data-portals-europe_en (2018)
7. Cheung, K., Huth, M., Kirk, L., Lundbæk, L.-N., Marques, R., Petsche, J.: Owner-centric sharing of physical resources, data, and data-driven insights in digital ecosystems. In: *Proceedings of the 24th ACM Symposium on Access Control Models and Technologies (SACMAT '19)*. Association for Computing Machinery, New York, NY, USA, 73–81. <https://doi.org/10.1145/3322431.3326326> (2019)
8. Big Data, Big Impact: New Possibilities for International Development. World Economic Forum Report July 2012, edited July 2019. Access on: <https://www.weforum.org/reports/big-data-big-impact-new-possibilities-international-development>
9. Carvalho, G., Madaan, A.: Impact assessment for data-driven innovation in cognitive IoT architectures. In: *Living in the Internet of Things (IoT 2019)*, pp. 1–8. IET (2019)
10. Gaia-X: A Federated Data Infrastructure for Europe. Access on <https://www.data-infrastructure.eu/GAIAX/Navigation/EN/Home/home.html>
11. Global cloud services market surges by US\$10 billion in Q42020. Global Cloud Infrastructure report Q4 2020. February 2021. Canalys Innovation. Access on: <https://www.canalys.com/newsroom/global-cloud-market-q4-2020>
12. Jennex, M.: Big Data, the Internet of Things, and the revised knowledge pyramid. *ACM SIGMIS Database DATABASE Adv. Inf. Syst.* **48**, 69–79 (2017). <https://doi.org/10.1145/3158421.3158427>
13. Taylor, P., Allpress, S., Carr, M., Lupu, E., Norton, J., Smith, L., Carvalho, G.T., Westbury, P.S.: *Internet of Things Realising the*

- Potential of a Trusted Smart World. Royal Academy of Engineering report, London (2018)
14. White Paper on Artificial Intelligence: Public consultation towards a European approach for excellence and trust. Consultation results. European Commission. Access on: <https://ec.europa.eu/digital-single-market/en/news/white-paper-artificial-intelligence-public-consultation-towards-european-approach-excellence> (2020)
 15. Proposal for a Regulation on a European approach for Artificial Intelligence. European Commission. <https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-european-approach-artificial-intelligence> (2021). Accessed 24 Apr 2021
 16. Data governance and data policies at the European Commission, Secretariat- General, July 2020. European Commission. Access on: https://ec.europa.eu/info/sites/info/files/summary-data-governance-data-policies_en.pdf (2020)
 17. Koshiyama, A., Kazim, E., Treleaven, P., Rai, P., Szpruch, L., Pavey, G., Ahamat, G., Leutner, F., Goebel, R., Knight, A., Adams, J., Hitrova, C., Barnett, J., Nachev, P., Barber, D., Chamorro-Premuzic, T., Klemmer, K., Gregorovic, M., Khan, S., Lomas, E.: Towards algorithm auditing: a survey on managing legal, ethical and technological risks of AI, ML and associated algorithms. Available at SSRN: <https://ssrn.com/abstract=3778998> or <https://doi.org/10.2139/ssrn.3778998> (2021)
 18. Kazim, E., Barnett, J., Koshiyama, A.: Automation and fairness: assessing the automation of fairness in cases of reasonable pluralism and considering the blackbox of human judgment. Available at SSRN: <https://ssrn.com/abstract=3698404> or <https://doi.org/10.2139/ssrn.3698404> (2020)
 19. Kazim, E., Denny, D.M.T., Koshiyama, A.: AI auditing and impact assessment: according to the UK information commissioner's office. AI Ethics (2021). <https://doi.org/10.1007/s43681-021-00039-2>
 20. Wachter, S., Mittelstadt, B., Russell, C.: Why fairness cannot be automated: bridging the gap between EU non-discrimination law and AI. Available at SSRN: <https://ssrn.com/abstract=3547922> or <https://doi.org/10.2139/ssrn.3547922> (2020)
 21. Siikut, S.: How 20 years of digital transformation made Estonia a model for dealing with international crises. Emerging-europe.com. <https://emerging-europe.com/voices/estonia-how-20-years-of-digital-transformation-be-a-model-for-preparing-for-international-crises/> (2020). Accessed 1 June 2021
 22. IMF: E-stonia Takes Off. Ilves, TH. <https://www.imf.org/external/pubs/ft/fandd/2018/03/trenches.htm> (2018)
 23. <https://www.valitsus.ee/en/news/e-estonia-council-were-introduced-vision-digital-state-and-cybersecurity-2030>. Accessed 13 Sept 2021
 24. <https://e-estonia.com/>. Accessed 1 June 2021
 25. Alliance for Internet of Things Innovation (AIOTI). <https://aioti.eu/>. Accessed 1 June 2021
 26. European data strategy: Making the EU a role model for a society empowered by data. European Commission, February 2020. Access on: https://ec.europa.eu/info/strategy/priorities-2019-2024/europe-fit-digital-age/european-data-strategy_en (2020)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.