SOME REFLECTIONS ON 'DATAFICATION':
DATA GOVERNANCE AND LEGAL
CHALLENGES

Lorena Elena STĂNESCU
Raluca ONUFREICIUC

Covered in:
CEEOL, Ideas RePeC, EconPapers, Socionet, HeinOnline

Published by:
Lumen Publishing House
on behalf of:
Stefan cel Mare University from Suceava,
Faculty of Law and Administrative Sciences,
Department of Law and Administrative Sciences
Abstract

Data is considered to be the new fuel for a business growth. By the Data Strategy, the EU made a big step in designing a vision of how data governance can be integrated in organizations’ management and activity, in governments’ systems and, ultimately, in people’s (digital) life. Translating data into competitive business outcomes through digital innovation while ensuring the protection and security of individuals’ (non-) personal information raises ethical and legal questions that we need to prepare to answer.

Keywords

Data governance; legal challenges; high-value data; business asset; data innovation EU Data Strategy; data protection.

Introduction

Data: business risk or business asset?

The growth and evolution of data governance as a practice nowadays requires an improved data availability, transparency, a higher quality data and a greater accountability. As organizations advance further on into the data-driven age, the speed at which they generate, store and consume, the volume of data increases and becomes uncontrollable. However, more data means more responsibilities, which come with more challenges. For example, many organizations face difficulties in designing data governance policies for implementing in their organizations and lack the business understanding required being unable to communicate them as an indispensable need.
Data governance can bridge the gap between the massive amount of data a business has to deal with and let the business carry out its activity. In addition, the big organizations “are beginning to realize that data-related risk can show up in many areas of their business and information governance issues touch every function and business unit.”[1]. Nonetheless, the perspective on the data is changing from being catalogued as risk to being capitalized as asset. For this matter, the question it raises is how data is turned into value. Value is not only about financial profit, but it may refer also to social and healthcare benefits or customer experience. Translating data into competitive business outcomes through digital innovation while ensuring the protection and security of individuals’ (non-) personal information raises ethical and legal questions that we need to prepare to answer.

**European Legal Framework**

The European Union legislator has developed an early strategy to make its data more open and secure [2]. Between the 1970s-1980s, the very first wave of structural reforms included privatizations, and liberalizations such as market-oriented reforms. The second wave regarded regulatory reform in the decade of 1980s-1990s involving ex ante filters. The first document adopted by the European Commission in the field of reusing public sector information was adopted in 1989 with the intention to be "a positive initiative from governments, to encourage the use and exploitation of public sector data and information"[3]. A further step and of utmost importance was taken in 1995 together with the adoption of European Union Data Protection Directive. In the following year, in order to protect non-personal data too, the European Commission adopted the Database Directive which contains clauses about copyrights and sui-generis rights with respect to databases of non-personal data. In 1998, on a background of lack of transparency and legislative coordination between the Member States in what regards technical and legal procedures for a better use of non-confidential information for the benefit of business and citizens, the European Commission adopted the Green Paper on Public Sector Information. In the same year, at the Fourth Environment for Europe Ministerial Conference was adopted the Aarhus Convention which entered into force on 30th of October 2001. Encouraging public participation in environmental decision-making and a better access to information, the main objective of the Aarhus Convention was the “protection of the right of every person of present and future generations to live in an environment adequate to his or her health and well-being [4]”.
Establishing a regulatory governance framework, one of the main achievements of the third wave was analyzing policy making as a process by defining the policy cycle concept. In 2002, the first ePrivacy Directive regulated how providers of electronic communication services should manage their subscribers’ data. It was amended in 2009 and presently on the edge of being replaced by a new proposed regulation targeting companies operating in the digital economy and what specific requirements on processing of personal data they will have to meet. Focusing more on reusing information for economical purposes, the PSI Directive [5] was revised in 2013 and 2017 and finally replaced by the Open Data Directive which entered into force on 16th of July 2019. A unique case of legislative “regional” approach [6] concerning the infrastructure for spatial information, the INSPIRE Directive delimited common implementing rules in areas such as metadata, data and service sharing, monitoring and reporting.

Since the 2010s the legislator’s role was re-designing the government by providing an adaptive legal framework built on experimental-design-based rules and embedded compliance. During this period, we assist at certain important moments reflecting the concern of the European legislator towards data: the adoption of Digital Single Market Strategy (2015), the enactment of General Data Protection Regulation with respect to personal data of the European Parliament and of the Council of 27 April 2016, in force starting with May 25th, 2018, the recast of the Directive (Eu) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information and the publishing on 19 February 2020 of the European Commission’s data strategy. Discoveries of new technologies motivate a new algorithmic and data-driven legislation [7]:“interest in regulating AI systems increased, with a focus on data protection, algorithmic accountability, and biometric/facial-recognition safeguards. Building on the emergence of globally oriented data protection approaches such as the European Union’s General Data Protection Regulation (GDPR), policymakers are moving quickly, driven both by the current sense of urgency to regulate the mass deployment of AI technologies lacking discernible safeguards and by the failure of ethical frameworks to adequately answer the call for accountability and justice.”[8].
Premises

People have lost control of their data

Data comes from everywhere. The individuals create data. The smart devices produce data. The business organizations generate data. There are various and almost unlimited number of streams as news/online media, digital traces, social media, medicine labs and many more. Everything is connected. Anything generates data and is becoming automated. Everything becomes digitally conversed and needs to be protected. So, how can we - as a society - accomplish this?

At the beginning of 2020, the digital universe was estimated to consist of 44 zettabytes of data meaning that the digital universe was 40 times more than the number of stars in the observable universe [9]. Like the light, the data is traveling and, though the speed values are comparable, “researchers are also working on transmitting data via light through the air, say using lightbulbs for WiFi, or transmitting laser beams from building to building. Again, light through the air does move at close to light speed, but nothing we have now is surpassing the speed limit” [10] with the aim to “achieve actual faster-than-light transfer”.

So, as data crosses borders, it is difficult to be followed and monitored under aspects of integrity, accuracy and purpose of use. Also, the fragmentation of the present digital ecosystem makes data unmanageable. Nonetheless, due to limited knowledge of the users about cybersecurity and protection of their digital rights, the personal data is easy to be accessed by hackers.

We can see how digital data is frequently repurposed for marketing, commercial [11], behavioural studies or even surveillance purposes [12]. For instance, there have been situations when “platforms have granted additional powers to particular trusted groups to allow them to more easily flag a larger volume of posts or access expedited moderation processes. Many platforms are responsive to requests from law enforcement agencies, but do not explain that content was identified by law enforcement when moderating content under their terms of service. This leads to confusion about decisions and fears about potential overreach from public agencies in ways that would raise due process or constitutional issues if the action were taken directly by the public agency.”[13].

Another issue is that most data today is controlled by a handful of monopolies [14]: “Google, Facebook, Microsoft, and Amazon are custodians of at least 1,200 petabytes of people’s information. Some of them could acquire more global data created daily as they branch out. For instance, Facebook wants to establish a new financial world order with the
launch of Libra, a stable coin-based payment network. In the event that the social media titan’s plan comes to fruition, it could rival the power of central banks. Facebook could then exercise some monetary policies as it sees fit in order to manipulate and stabilize the value of its own cryptocurrency”. The main concerns are a future possible concentration of powers in the hands of a few entities can lead to economic inefficiencies and inequalities [15]. Acknowledging this situation, the European Data Protection Supervisor (EDPS) advised that “processing of data for the public good should not create or reinforce situations of data oligopoly (dependency of the public sector, SMEs, etc. on few powerful IT companies, so called Big Tech). This is also relevant from a data protection perspective since monopolies and oligopolies create situations of users’ lock-in and ultimately restrict the possibility for individuals to exercise effectively their rights” [16].

Nevertheless, we have to mention the technology incidents [17] of “2018 have shown that the gap between those who develop and profit from AI—and those most likely to suffer the consequences of its negative effects—is growing larger, not smaller.” Lack of government regulation together with a dynamic AI sector are gaps that raise serious concern about discrimination, due process and overall injury responsibility [18].

**So how could people regain control on their data?**

For the purpose to discuss some possible answers to this question, in the following rows we shall present the concept of data governance and how is it reflected in the EU Data Strategy.

**What is data governance?**

According to The Data Governance Institute, by “data governance” it can be understood “the exercise of decision-making and authority for data-related matters” [19]; and an extensive meaning: “data governance is a system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods. Also, data governance is seen as “the practice of organizing and implementing policies, procedures and standards for the effective use of an organization's structured/unstructured information assets” [20].
Furthermore, data governance is considered to be "a system of decision rights and accountabilities for information-related processes, executed according to agreed-upon models which describe who can take what actions with what information, and when, under what circumstances, using what methods."

Also, data governance is seen as a process of managing the availability, usability, integrity and security of the data in enterprise systems, based on internal data standards and policies that also control data usage" [21].

Nonetheless, by data governance it can be understood the “execution and enforcement of authority over the management of data assets and the performance of data functions” [22].

The data governance strategies, policies and programs are closely linked with IT governance practices which help organizations achieve great data performance. Even though the two concepts are very similar, they have independent goals, but aim only one target: generate greater business value for the organization.

Data Governance programs are designed to satisfy specific exigencies of an organization, like meeting compliance, corporate or data integration requirements. Referred to as “focus areas” [23] the six most common sectors where data governance should be implemented are focus on management alignment, focus on data warehouses and business intelligence (bi), focus on architecture/integration, focus on privacy/compliance/security, focus on data quality, and focus on policy, standards, strategy. In this paper, our attention is directed to legal compliance, which includes the last three above mentioned areas of focus.

**Data governance model. How does it work?**

Digital transformation cannot be stopped. Generating data cannot be stopped. There are many legal challenges. There are many ethical dilemmas. But what can we do? The best efforts to ensure the respect of the citizens' human rights and legitimate interests in all circumstances.

Historically, the cost criteria had always been a key point for data governance programs. Focusing more on people and process and with poor funding, data governance programs had frequently started from scratch. As a consequence, tools such as Microsoft Excel and SharePoint were often an obvious choice to document and share data governance artifacts. Though certain business intelligence companies may promote the data governance model as restoring order to chaos, in the spirit of honesty, we
should recognize that the data chaos is unknown. However, the experience of COVID-19 pandemic situation motivated the European Commission to accelerate the plans for the “digital transformation” and provide a data governance model encompassed within the EU Data Strategy [24]. In relation with the subject matter of the present paper, we highlight three directions that will define the use of data in the future.

Aiming to create a single market for data, as a first recommendation to the companies to pool and share their personal and non-personal data, the European Commission proposes a “common European data spaces in strategic sectors and domains of public interest” that leads “to the availability of large pools of data in these sectors and domains, combined with the technical tools and infrastructures necessary to use and exchange data, as well as appropriate governance mechanisms”.

For this purpose, the European Commission is looking forward to “foster business-to-government data sharing for the public interest” and encourage “business-to-business data sharing, in particular addressing issues related to usage rights for co-generated data (such as IoT data) typically laid down in private contracts”, while clarifying the legal framework for achieving these goals.

Concerning projecting a wider set of strategic EU investments in new technologies, the European Commission plans for the period 2021-2027 an investment “in a high impact project on European data spaces and federated cloud infrastructures”, concerning “funding for edge computing, high-performance computing, cyber-security, low-power processors and 6G networks. These investments are essential for the EU’s data infrastructure of the future, to equip Europe with the right infrastructures, computing power, encryption capacity and cybersecurity tools to process data”.

In what regards restoring individuals’ control on their data, the European Commission is empowering individuals with respect to their data by supporting them to “be in control of their data through tools and means to decide at a granular level about what is done with their data”.

However, we should not forget that the implementation of the EU Data Strategy legal-ethical has to comply with the framework regarding human rights such as: privacy, data protection, dignity, autonomy, freedom of expression, non-discrimination, equality, participation and social/economic goals: welfare, competition, efficiency, science, art and culture, cooperation, civic dialogue, democracy [25].
Data protection: Legislation enforcement needed to protect citizens data

As the technology and ethics professor Helen Nissenbaum stated: “information technology is considered a major threat to privacy because it enables pervasive surveillance, massive databases, and lightning-speed distribution of information across the globe.”[26]. Even though we have a lot of knowledge and creativity, there is not yet a clear representation of the effects of the big data on our lives. How should we be in control with our data if we do not know how to avoid or mitigate the risks?

A reasonable first step is to perform impact assessments as EDPS recommends: “The use of data for the public interest may involve large-scale processing, which combines data from a variety of sources, potentially involving special categories of data or vulnerable groups data”. In accordance with article 35 of the GDPR data controllers are obliged to conduct data protection impact assessment in a transparent and trustworthy climate [27]. Inspired from the data protection impact assessment, but with specific application to AI technology, it is developed a new concept of algorithmic impact assessments (AIAs): “Built on the success of data-protection, environmental, human-rights, and privacy-impact assessments, AIAs require AI vendors and their customers to understand and assess the social implications of their technologies before they are used to impact people’s lives.” [28].

Secondly, we have encountered many arguments that support the right to data portability. Stressing out the individuals’ rights to control and protect their data, in a recent Opinion, EDPS “welcomes the objective of the Strategy to empower the individuals to be in control of their data”. Also, another strategy’s target underlines “the portability right of individuals under Article 20 of the GDPR”. Analyzing the right to data portability in correlation with the concept of “real data sovereignty” we think that the EU Data Strategy’s implementation must give more importance to the consequences and responsibilities provided by wide open access to data. In this regard, we relate to the argument presented in a must-read article about the current understanding of the term of transparency, by which the authors refer to the transparency system not only in what regards the information content, but also in the accountability element [29]. Using the machine execution language to discuss the relation between transparency and accountability, we
see how new concepts and terminologies are born such as algorithmic transparency.

There are also discussions about regulating the “right to explanation”. This right is seen to be used because of the increasing cases of using automated decisions: “algorithmic decision systems are often touted for their putative benefits: mitigating human bias and error, and offering the promise of cost efficiency, accuracy, and reliability. Yet within health care, criminal justice, education, employment, and other areas, the implementation of these technologies has resulted in numerous issues” [30] as presented in Litigating Algorithms 2019 US Report: New Challenges to Government Use of Algorithmic Decision Systems.

Some other suggestions for protecting the users date include limited lifetime and uses of data, keeping privacy relevant, updated laws to the new technologies developments untraceable, unlikable identities: Attributed Based Credentials, smart rules: creative commons licenses[31], BCN Pilot: data-driven participatory democracy (DECIDIM)[32] or IoT pilot: personalized data commons dashboards [33].

**Governments Data Innovation**

According to EDPS, access to and processing of data held by public authorities should be implemented by ensuring: “transparency and societal participation on the purpose of the reuse vis-à-vis the citizens/data subjects, as well as transparency and clear purpose definition between the licensor (the public authority) and the licensees”. In all cases, before any reuse is more than recommendable a data protection impact assessment for data processing according to GDPR to identify the risks and the appropriate data protection safeguards [34]. Also, EDPS underlined that especially in the tech sector which is able to control a considerable amount of valuable data there is a corporate secrecy issue. In the context of the pandemic crisis, for healthcare purposes, the EU Commission affirmed that a new regulated access is needed. Therefore, Government open data innovation may be an answer at the problem of limiting the market power and help citizens gain power in front of the natural monopolies. For this reason should be taken in consideration alternatives like decentralized networks, participatory design and responsible research and innovation should be taken into consideration [35].
Some primary examples showing the effects of the direct use of Artificial intelligence and big data by private and public organizations are the research activities conducted at the European University Institute. Firstly, the ARTSY[36] project which extensively documented and evaluated “the market practice of the corporate use of AI, map the scholarly debates about (consumer) law and artificial intelligence, and present a list of twenty five research questions which, in our opinion, require attention of regulators and academia”. Secondly, in the CLAUDETTE project – “automated CLAUseDETectEr”[37] the main objective was to test, by using machine learning approaches, the limitations of automate reading and legal assessment of online consumer contracts and privacy policies and to what extent it is compliant with EU’s unfair contractual terms law and personal data protection law.

In what regards a future possible regulation, there are some difficulties to overcome. One of them is represented by fragmented legal infrastructure that governs the (digital life on) internet. Also, the constant development of the digital technologies may cause violations of citizens’ rights by using the online data provided by the users themselves. To overcome these challenges, a new concept of laws written in programing language appears: RegTech [38]. Furthermore, there were developed many computer programs that exercise the functions of lawyers or compliance officers. As we have mentioned before, within the CLAUDETTE research there was developed an automated detector of potentially unfair clauses in online terms of service “where machine learning is employed to automatically detect such potentially unfair clauses.”[39]. Moreover, using CLAUDETTE could be performed automated evaluations of privacy policies using artificial intelligence [40].

Given the properties of law and computer systems, the question remains to what extent is it possible to have machines that can do what legislators or lawyers do [41].

Conclusion

Considering the factual situation, the rapid advance of new technologies and the legislative framework in progress, for business innovation, data governance can be a possible solution to managing risks such as security, privacy or liability. The question that still remain unanswered is if data governance can be a possible solution for the reconciliation of different
regulatory frameworks for the future of digital information. Data governance may be a way to ensure that a business is compliant with the regulatory framework, but it depends of people if data is a business risk or a business asset. Its important role in the future should not be underestimated by organizations, nor citizens.

Acknowledgements: This paper was co-financed from the Human Capital Operational Program 2014-2020, project number POCU / 380/6/13/125245 no. 36482 / 23.05.2019 "Excellence in interdisciplinary PhD and post-PhD research, career alternatives through entrepreneurial initiative (EXCIA)", coordinator The Bucharest University of Economic Studies”.

REFERENCES


SOME REFLECTIONS ON 'DATAFICATION': DATA...


[14]. Cullen C. Over 43% of the internet is consumed by Netflix, Google, Amazon, Facebook, Microsoft, and Apple: Global Internet Phenomena Spotlight; 30 August 2019. (Internet). Available from: https://www.sandvine.com/blog/netflix-vs-google-vs-amazon-vs-


[23]. Thomas G. Focus Areas for Data Governance. The Data Governance Institute. (Internet). Available from:
SOME REFLECTIONS ON 'DATAFICATION': DATA…


