ISSN 2281-5147

Opinio Juris in Comparatione

Studies in Comparative and National Law

Legality attentive data scientists for everyone Special Issue 2024

Founded by the European Union's H2020 ITN project LeADS-Legality Attentive Data Scientists
Grant Agreement ID:956562

TABLE OF CONTENTS

ISSN 2281-5147

INTRODUCTION TO "LEGALITY ATTENTIVE DATA SCIENTISTS FOR RVERYONE"

by Giovanni Comandé*

This special issue of Opinio Juris in Comparatione includes 14 contributions by the Early Stage Researchers (ESRs) of the "Legality Attentive Data Scientists" (LeADS) Project, funded by the European Commission (GA number 956562). These contributions were selected through a rigorous peer-review process conducted by a pool of international and interdisciplinary reviewers.

The review and selection process spanned nearly nine months, during which the researchers were guided in creating contributions that were both scientifically rigorous—based on three years of research within the project—and intentionally accessible to non-experts (at least to those not specialized in the researchers' respective core disciplines). The cultural goal aligns with the objectives of LeADS, which aim, among other goals, to create "a new interdisciplinary professional figure that we call Legality Attentive Data Scientist or LeADS. LeADS will be an expert in data science and law expected to work within and across the two disciplines, a leader in bridging scientific skills with the ethic-legal constraints of their operating environment."

Credit must also be given to all the ESRs, even those who did not pass the lengthy and rigorous review process, for their intense cultural effort in crafting a language that is scientifically rigorous yet imbued with a popularizing spirit. This approach often required a certain heterogeneity in the notes, or even their removal and simplification.

* Full Professor of Comparative Private Law at Sant'Anna School of Advanced Studies and Coordinator of LeADS Project (www.legalityattentivedatascientis.eu). All contributions in this Special Issue were financially supported by the European Union's Horizon 2020 - Innovative Training Networks programme under the grant agreements ID: 956562.

ISSN 2281-5147

Each researcher interpreted the task according to their individual talents and the constraints imposed by their core discipline, resulting in styles that were often very different but all deserving of careful reading and re-reading to appreciate their diverse nuances.

We are confident that all the contributions will inspire the readers. Possibly those articles that explore the sectorial boundaries and overlaps of legislations, may help legislators and policymakers to propose regulatory innovation, support legislative reforms and devise appropriate policies bridging these players with in depth analysis and a less esoteric language. Eventually, theoretical and critical contents may help judges, independent authorities and legal experts to give a turn in their understanding of the emerging digital regulatory framework. Similarly, the technical and empirical results may increase the preparedness of developers, engineers, business owners and governmental organizations to more efficiently design and develop technologies and implement norms in real case scenarios.

Overall, we hope that citizens at large may be directly impacted by the overall results presented in these pages, for example through technological advancements for smoother user-centered privacy-friendly management of personal data or through access to fairer automated decision-making in key sectors such as justice and employment.

This issue is organized in sections around four core arguments.

The first one, devoted to "*Privacy, consumers and competition*", contains 3 rich contributions.

Onntje Hinrichs wrote on "*Why Your Data is not Your Property (and Why You Still End Up Paying With It)?*" exploring three interrelated topics that reveal tensions in the European approach towards the regulation of the data economy: (i) data as property (ii) data and fundamental rights, and (iii) data as currency. Qifan Yang put her twofold skills of statistician and jurist at work to understand the complex relationship between the GDPR and market competition in her article "*Your Data Rights: How does the GDPR Affect the Social Media Market?* "Last but not least Tommaso Crepax, with his very intriguing style drives everyone into the realm of data portability in the contribution entitled "*Bring Your Own Data. The struggle of re-using data in a world of heterogeneous systems*".

ISSN 2281-5147

The second section, devoted to "*Privacy and security in practice*" is definitively dominated by a team of researchers with heightened technological skills. Cristian Lepore's "*Self-Sovereign Identity: The Revolution of Digital Identity*" drives us through the complex world Self-Sovereign Identity (SSI). Meanwhile, Louis Sahi, through his summary of interviews with experts and background analysis in "*Evaluation and Harmonization of Data Quality Criteria: Insights from Expert Interviews for Legal Application*" escorts the reader in understanding the technical and legal role of data quality criteria and the need for collaborative data processing (CDP) in decentralised environments.

Armend Duzha ("*Extracting Data Value through Data Governance*") and Christos Magkos ("*Personal Health Information Management Systems (PHIMS) for user empowerment: A Comprehensive overview*") continue this section. Mr. Duzha explores a new approach for data governance developed to extract data value respecting the ever-delicate balance between transparency and privacy, relating it to novel technologies such as Artificial Intelligence, Federated Learning and Blockchain, and illustrating how these can be integrated in a data governance program. Mr. Magkos devotes his attention to personal health information management systems (PHIMS) and on how integrating raw data could provide a method for the storage, management, and regulation of personal health data access. The key message is how PHIMS can empower users to take control of their own healthcare.

Privacy risks entailed in the advent of AI is at the core of the last contribution of this Section devoted by Soumia al Mestari to "*What AI is stealing! Data privacy risks in AI*". She discusses that this risk of AI's leaking personal data is not only hypothetical and suggests how to mitigate it.

The third section is devoted to "*Sharing (personal) data*", a title that would have suited a number of the contributions in the previous sections. Yet her ethe focus is more on the sharing *in practice*. Barbara da Rosa's "*Can business-to-government data sharing serve the public good?*" explores a number of regulations enacted by the European Union and their overlaps and analyzes if they indeed assist business-to-government data sharing. Xengie Doan ("*Collective consent, risks and benefits of DNA data sharing*") uses genetic data sharing as a use case to better understand what tools and methods can enhance a user-friendly, transparent, and legal-ethically aware collective consent. Still in the domain of health data is the contribution of Fatma Dogan ("*To Use or Not to Use? Re-using Health Data in AI Development*") focusing on the re-use of health data in

ISSN 2281-5147

the context of AI development, concentrating on regulatory frameworks governing this practice under the European Health Data Space. Her aim is to assess whether health data can be re-used for AI-driven healthcare advancements without undermining individuals' data protection rights.

In the last contribution of the section Maciej Zuziak ("*How to collaboratively use statistical models in a secure way*") empowers the curious reader with a set of links and pointers that would allow them to go deeper into a well of data governance and large AI infrastructure but only after having introduced the reader to the nuanced world of decentralised learning systems and statistical learning explaining the basic technocratic lingo in an engaging way.

The last Section ("*Preparing for AI* ") is opened by Mitisha Gaur's "*Policing the AI Judge: a Balancing Act*" where she analyzes AI backed automated decision-making systems used by public authorities. She advocates for a strict governance framework based on risk management and algorithmic accountability practices focused on safeguarding fundamental rights and upholding the rule of law by adhering to the principles of natural justice.

Robert Poe's challenging "*The perils of Value Alignment*" is a program already by the title. The article vigorously argues that global AI governance risks institutionalizing violations of fundamental rights. It argues that the current ethical foundation of AI governance can lead to conflicts with the rule of law. It calls for a re-evaluation of AI governance strategies, urging a realistic approach that respects citizens, legal precedent, and the nuanced realities of social engineering.

ISSN 2281-5147

ISSN 2281-5147

WHY YOUR DATA IS NOT YOUR PROPERTY (AND WHY YOU STILL END UP PAYING WITH IT)

Onntje Hinrichs*

Abstract

This essay explores three interrelated topics that reveal tensions in the European approach towards the regulation of the data economy: (i) data as property (ii) data and fundamental rights, and (iii) data as payment. By retracing how scholars and policy makers have attempted to find an appropriate regulatory framework for the data economy, this essay shows that up to this day, contradictions in the EU's approach to the data economy persist and become evident in our everyday lives online. Despite not owning our data, we pay for digital content and services with it. This essay clarifies this paradox and its role in ongoing legal battles between the large corporations, civil society and the EU.

Table of Contents

* Onntje Hinrichs is a PhD Researcher at the Research Group on Law, Science, Technology and Society (LSTS), Vrije Universiteit Brussel (VUB) and a Marie-Sklodowska Curie Action Fellow in the Legality Attentive Data Scientists project. His research explores emerging forms of data governance in EU law and how consumer law shapes the regulation of data. E-mail address: onntje.marten.hinrichs@vub.be
This work is supported by the European Union's funded project Legality Attentive Data Scientists (LeADS) under Grant Agreement no. 956562

ISSN 2281-5147

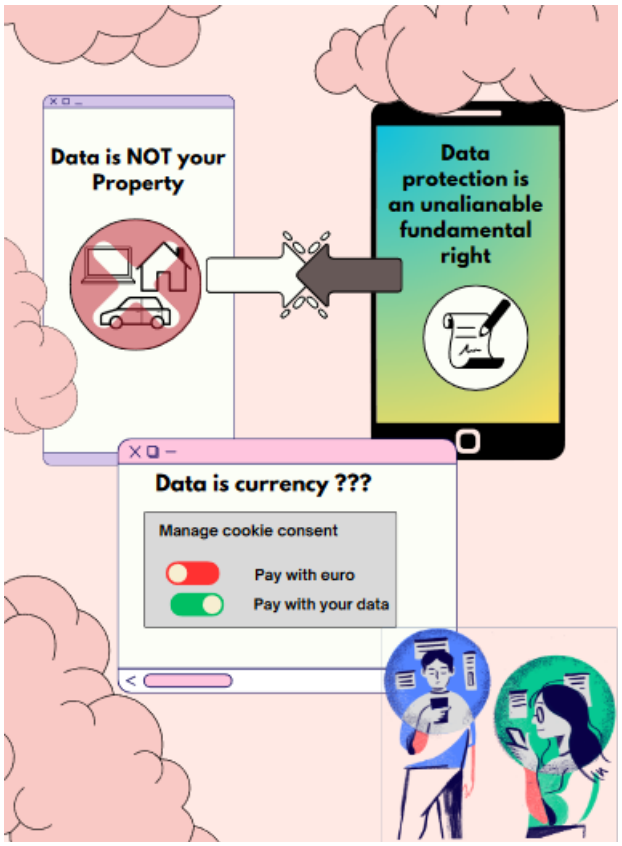
Keywords

Data Protection – Data Economy – Data Property – Consumer Protection – Pay-Or-Consent

ISSN 2281-5147

1. Introduction

If data is considered as the new oil, shouldn't we as consumers somehow benefit monetarily from allowing others to harvest 'our' personal information? Put more polemically, if our shopping habits online are the new oil fields of the 21st century, does that mean we will all become rich?



We do not 'own' our data. It is protected as a fundamental right instead – but we experience online that in practice data constitutes a form of payment. This essay explains this paradox and the challenges it creates.

The last question is evidently polemic as it simplifies and caricatures the metaphors of 'data as the new oil' or 'data as the new gold to be mined' which have increasingly been employed by businesses and policy makers to highlight the importance of data for the information economy. Nevertheless, as is often the case with caricatures, they contain elements of truth and reflect important questions we as society have to deal with. Moreover, these topics are grounded in debates that have marked academic and policy discourse over the past decades on how to regulate the data economy. This essay is structured around three interrelated topics that are still central to the regulation of the data economy: (i) **data as property**, (ii) **data and fundamental rights**, and finally (iii) **data as payment**.

The first section explains why 'our' data cannot be considered as our property – contrary to resources such as oil or precious metals such as gold. It outlines, how scholars and policy makers have debated whether introducing property rights in data could be a viable way for individuals to control and benefit economically from their data. Furthermore, it might facilitate the emergence of competitive data markets. The second section on data and fundamental rights, elaborates upon how 'our' data is currently protected within the legal data protection framework of the EU. It explains

ISSN 2281-5147

why the European approach has been described as incompatible with the idea of 'data property' (Mayer-Schönberger, 2010).

Whereas 'property' rights might typically enable us to sell, destroy or rent objects we own, fundamental rights are for evident reasons very different. We should not be able to transform our fundamental right into a commodity that can be bought and sold on a market. It should be impossible to put a price tag on human dignity which is thus described as an unalienable fundamental right. However, if data property stands in direct opposition to the fundamental right to data protection, there should be no necessity to deal with the last topic of this essay: data as payment. Counter-intuitively, however, this remains a highly contentious issue and the last section of this essay explains why. It exemplifies how seemingly trivial things such as cookie banners reflect complex legal questions to which thus far no conclusive answer exists. This essay clarifies this paradox and its role in ongoing legal battles between large corporations, civil society, and European Institutions.

2. Data as Property

In our everyday language, it has become perfectly normal to refer to 'our' data when we speak about the traces which we leave online, and which are used by companies for commercial purposes. For instance, whenever we visit online shops, our data is used to create profiles about our consumption habits. The objective: to show us other products and services which, based on the collected data, we might be interested in. When we speak about 'our' data, it expresses how we feel about information that relates to us. Since the traces which we leave when we browse the internet oftentimes reveal private and intimate information about us, such as our sexual orientation or political views when we 'like' certain content on social media, it makes sense to intuitively claim ownership over such data and to prevent others from using them. 'Our' data should belong to us. Such language thus reflects how the notion of ownership can be a psychological concept that expresses a sense of belonging. From a legal perspective, however, 'ownership' and 'property rights' denote something entirely different.

From a legal perspective, property rights do not come automatically into existence because of a felt claim over something. Instead, property rights have to be artificially

ISSN 2281-5147

created by the law. It is the law which exclusively defines what can and cannot be owned. It is the law which defines to what extent and under which conditions others can interfere with our property. Whereas a contract is only binding on the contracting parties, property rights can be enforced against everybody else. When you own a house, you can exclusively and selectively determine, who should be able to enter. You may decide to rent, to sell, or even destroy it. Your ownership thus lasts until you decide to end it. Deciding why certain things should (not) be treated as objects of property law, is thus a highly sensitive question as the rights conferred are far-reaching and are binding to others. Think again about the metaphor of owning a house. You can prevent people from entering your house. If they still enter against your will, you can call the police to enforce your rights.

Should data be considered as an object of property law? Should the law create new property rights that can be invoked between individuals and businesses with regard to data? Again counter-intuitively, this is not a new question but has been a hotly debated issue for several decades. Already in 1968, decades before the collection of data both online and offline has become ubiquitous, authors argued that the right we as individuals have over our personal information should be understood as a property right (Westin, 1968). Conceiving our personal information as property would give us control over our personal information. Just like property rights over tangible objects give us the possibility to exclude others from using them, property rights over (intangible) information should empower us to exclude others from using it without our permission.

This argument gained prominence with the growing importance of the internet during the 90s. Whilst our personal information was already being collected in the physical world, it still required comparably more effort. With the emerging architecture of digital cyberspaces, on the other hand, the collection of our personal information was becoming the new default. Data property was thus believed by many as a solution to empower us online with regard to our data (Lessig, 1999). Without our approval, companies would not be allowed to use any of our personal information. If, however, we wanted to sell our data to the highest bidder, a property rights regime would empower us to do so. Data property would thus not only be a useful instrument to protect citizens' data, but at the same time it would benefit the economy since it could

ISSN 2281-5147

facilitate the emergence of data markets and thus the availability of data for companies.

Today, however, we still have not created property rights for data. We might talk about 'our' data in everyday language, but this does not mean that 'your' data is 'your' property from a legal perspective. It expresses a felt entitlement over 'our' information, but it does not translate into corresponding legal property rights. Therefore, you do not own 'your' data in a similar way you own your car, laptop, or house. The reasons why we have not taken this path are many. Some relate to the traits of data (as a non-physical object, how do you transfer property rights from one person to another?), others to determining its value (what is the precise value of our data – can you ascribe monetary value to it in a similar way we do with physical objects?), and others to general societal objectives and how they can be best achieved through law (shouldn't information be 'free as the air we breathe' to foster culture and artistic expression or innovation?). Most importantly in Europe, however, the creation of property rights in data was believed to be diametrically opposed to the foundation of European data protection law: Data protection in Europe constitutes a fundamental right which is enshrined in article 8 of the Charter of Fundamental Rights of the European Union.

3. Data Protection as a Fundamental Right

Why is the conception of data protection as a fundamental right difficult to be reconciled with the creation of property rights in data? Rights that are characterized as being 'fundamental' denote core values of our European society. Whereas some of such rights, whether it is the fundamental right to data protection, freedom of expression, or freedom of assembly, can be limited under certain circumstances – their core is absolute. This implies furthermore, that they are not considered as simple commodities which can be bought and sold on the market. Considering data as property would, however, make it akin to any other object which companies can purchase from citizens or which citizens can sell to a price they deem appropriate. Within the European Union, it is therefore difficult, if not impossible, to conceive data as property since it might turn a fundamental right into a commodity that can be

ISSN 2281-5147

bought and sold on the market – it would put a price tag on the right to data protection.

If not through property rights that grant us ownership over our data, how does European data protection law intent to protect or empower us as citizens with regard to our personal information? Most famously, through the General Data Protection Regulation, short GDPR. Often criticized for the bureaucratic burdens which it would impose on any type of business, regardless of its size and field of activity, it is the GDPR that imposes constraints on what public bodies and businesses can and cannot do with our data. It is the GDPR which has been used to inflict heavy fines on some of the largest corporations for their violations of EU data protection law: 1.2 billion euro for Facebook, 746 million euro for Amazon, 345 million euro for TikTok[1] – the numbers constitute a powerful reflection of the 'value' which the law ascribes to the protection of our data.

The regime for fines which the GDPR created for breaches of European Data Protection law, should of course only constitute one of the last means to ensure that our data is protected. It would be preferrable if companies and public bodies only use our data in a way that is compliant with the GDPR. Over 88 pages, the GDPR outlines rules which any entity that processes personal data has to comply with. It creates a set of rights which should empower us as citizens over our data – such as the right to access or delete data which companies have collected about us. It creates a variety of obligations to empower citizens through information. Privacy notices are an example of this. Companies must be transparent about what they do with our data. The law wants to put us in the position to better understand what happens to our data and to act accordingly (De Hert & Gutwirth, 2009). Companies and public bodies, on the other hand, must show at all times that they comply with the GDPR. If they fail to do so, they can be held accountable, for instance, through the imposition of fines.

This regulatory regime thus forces entities that use our data to always have in mind the obligations of the GDPR whenever they use our data. For instance, the GDPR classifies certain types of information as 'sensitive data', such as data concerning our health, sexual orientation, or political opinion. Whenever such data is being used, the GDPR imposes much stricter use conditions – reflecting again the fundamental rights

[1] For an overview of fines under the GDPR see for example https://www.enforcementtracker.com/

ISSN 2281-5147

rationale of the GDPR. Finally, in all such cases, the processing of our data must be 'lawful'. What does 'lawful processing' of our data mean within the context of the GDPR? The GDPR creates six major justifications ('legal basis') which companies or public bodies must rely upon when they want to use our data. If they fail to do so, the use of our data would be illegal. For instance, in some cases our employers may have to report our data to social security or tax authorities. Since the law imposes this obligation on companies, the GDPR authorises the use of our data in such circumstances. One of the most (in)famous justifications companies can rely upon when they want to use our data is 'consent'. What does 'consent' under European data protection law mean?

The GDPR needs roughly 900 words, spread over different articles and recitals of the GDPR, to explain what 'legal' consent means. The European Data Protection Board, a body which consists of representatives from all national data protection authorities and which is tasked with giving guidance on the application and interpretation of European data protection law, has issued two guidelines on what consent is under the GDPR (both put together total roughly 60 pages or 35.000 words). The answer to 'what constitutes legal consent' under the GDPR therefore does not seem to be an easy one.

Put in GDPR terms, consent should be a freely given, specific, informed and unambiguous indication of our wishes. Can we 'freely consent' when our employer asks us if it can process our data? Probably not. The dependency inherent to an employer/employee relationship and the possible repercussions if we say 'no' to our employer might stop us from 'freely' consenting. Furthermore, we should always be given a genuine free choice – for instance, consent should not be tied to the provision of a service or the access to digital content. If you only get something you want if you must consent to the processing of your data, it is not really 'freely given'. Instead, it seems more akin to a sort of payment which you provide in return for something you want.

4. Data as Payment

The first two sections of this essay outlined how data is protected as a fundamental right in the EU and how it cannot be considered as our property. As outlined in the

ISSN 2281-5147

first part of this essay, the idea of creating 'property rights' in data was rejected since we should not be able to 'own' or 'sell' our data on the marketplace. The second part explained this particularity of the European approach by elaborating upon the fundamental rights rationale that underpins European data protection law. It intends to protect our data through a variety of rules created by the GDPR. If data is not property but protected as an unalienable fundamental right instead, can data constitute some form of digital payment?

Following this line of reasoning, the obvious answer should be clearly: no, it cannot.[2] This is further exemplified by the characterisation of 'consent' as 'freely given'. The GDPR understands consent to the use of data as something that is not conditional for subsequent access to digital content.

However, the reality for most of us is different when we try to access digital content or digital services. Whenever you visit the website of a news publisher to gain access to articles for 'free', you have probably encountered so called 'cookie-walls'. 'Cookie-walls' (see image in beginning of the article) provide visitors of a website with the following binary choice: if they want to access the digital content of the website, visitors either have to agree that the publisher can use their data (for instance for marketing purposes) or they have to pay for a subscription. However, if digital content can only be accessed in exchange either for data or for money it is easy to perceive our personal information as some form of alternative payment. Such an approach thus seems contrary to the perception of the right to data protection as an unalienable fundamental right. Are such practices contrary to European data protection law?

Again, one would expect a clear answer. If data must not be considered as something that can be bought and sold on the market (it cannot be considered as property) but is protected as an unalienable fundamental right, data should not be considered as

[2] See in that context also European Data Protection Supervisor (EDPS), 2017 Opinion 4/2017 on the Proposal for a Directive on certain aspects concerning contracts for the supply of digital content. When the 2017 proposal for a Digital Content Directive intended to introduce the idea of data as counter-performance (meaning legally recognizing consumer data as a form of payment) the EDPS warned against any provision that would introduce the idea that people can pay with their data. For the EDPS it was clear that fundamental rights 'cannot be reduced to simple consumer interests, and personal data cannot be considered as a mere commodity'. In drastic words the EDPS observed how 'There might well be a market for personal data, just like there is, tragically, a market for live human organs, but that does not mean that we can or should give that market the blessing of legislation.'). When the final text of the Digital Content and Digital Services Directive was adopted two years later in 2019, it no longer described data as 'counter-performance' and highlighted how personal data could not be considered as a commodity.

ISSN 2281-5147

payment and such practices should not be allowed under European data protection law. However, the answers given by national data protection authorities differ. Some countries in the EU, thus seem to accept such business models. The European Data Protection Board, a group where members of all national data protection authorities meet, stated in 2020 that cookie walls which do not give consumers any other option apart from 'consenting' to the use of their data are contrary to data protection law. The French data protection authority in 2022, gave a more nuanced opinion. Cookie walls can be lawful, if they give consumers a real choice. This choice can consist of (i) either consenting to the use of our data or (ii) paying a *reasonable price* for accessing the digital content. What is a *reasonable price*? This would depend on each case.

Similarly, the Austrian data protection authority in 2023 argued how such cookie walls can be in accordance with data protection law. Consumers should have a certain degree of autonomy to decide what happens with their data (and thus chose if they want to pay for a service or consent to the processing of their data). Similarly to the French data protection authority, the Austrians highlighted that each solution would require a careful balancing of interests – the interests of us as consumers, but also the interests of companies who offer their services without monetary payment but in exchange for data. The Austrian authority was cautious in its approach because simply accepting any 'pay-or-consent' model would risk that low-income consumers in particular would always have to pay with their data since they cannot afford the 'pay' variant.

The cases from the French and Austrian data protection authorities both concerned cases where cookie walls were used by news publishers. In both cases the news publishers argued, how the collection of our personal information constituted a necessary counter-performance for their work – for example, journalistic articles. Since our information is subsequently being commercially exploited and monetised through the deployment of personalised advertisement, it enables publishers to finance their work and to provide us their content for 'free' – without obliging us to pay but to merely consent to the processing of our data. From a business perspective it thus equally is understandable how the processing of our personal information is an economic necessity for companies that provide digital content without monetary counter performance.

ISSN 2281-5147

'This Service is Free and will Always Remain Free'

One of the largest social media platforms advertised its services for years with the slogan 'Our services are free and will always remain free'. Beyond being a slogan to marketize its product, it also reflected a mentality online that every digital content, every digital service we use and consume is (and should be) available for free. Up to this day, some of the most used digital services, such as E-Mail or social media services, are provided in exchange for data and not for monetary payments.

Does the law challenge such business models? Think about the marketing slogan from a consumer protection perspective. Should such advertising be declared illegal because it makes a claim that is factually wrong and therefore misleading consumers? Are such services actually free? Do or do we not pay with our personal data? What this essay has tried to show is how complicated such a seemingly trivial question is. From a data protection perspective, the marketing slogan might be considered truthful. It is data protection law that insists on the fact how data is not a commodity (that cannot be propertized) and how we cannot pay with our data. When Facebook was brought to court over the legality of this marketing slogan, it used precisely this line of defence. It cited the European Data Protection Supervisor that data is not a commodity and that we as consumers would therefore of course not pay with our data for Facebook's service – hence their service could be marketed as being 'free'. This constituted of course a smart, but absurd, line of argumentation since it was one of the most infamous infringers of data protection law that relied on a European data protection institution to justify and defend the legality of its commercial practices. As one of the most profitable companies worldwide, Facebook certainly does not intend to donate its services to consumers.

More recently, Facebook, again, had to adapt its commercial practices to make them compliant with European law. As outlined during the second section of this essay, it needed a legal justification to ensure that it could use the data of its users in a lawful manner.[3] Facebook decided to rely on consent. It offered the following binary choice to consumers: either agree that we can use your data for targeted advertisement or

[3] See CJEU case C-252/21 Meta v Bundeskartellamt, [2023] ECLI:EU:C:2023:537. In this case which opposed Meta and the German competition authority, the European court clarified that Facebook needed consumer's consent if it wanted to use their data for its advertising business model and could not rely on, for instance, the legal basis of the GDPR where data processing is necessary for the performance of a contract. See e. g. paragraph 150 of the judgment.

ISSN 2281-5147

pay us a monthly subscription fee (at the time of writing of this article 9,99€/month). Put differently, either pay with your data or with your money.[4]

5. Conclusion

Your data is thus not your property, but (at the moment at least) you will still end up paying with data for services and digital content. This essay showed this unintuitive conclusion through three sections. First, it explained why data is not considered as property in the EU. You are not the owner of your data in a similar way you can be the owner of a house. Instead, our data is protected through the unalienable fundamental right to data protection. It showed how both concepts are fundamentally opposed. Whereas legal property rights enable us to buy and sell objects on the market, the fundamental right to data protection wants to precisely prevent that our data is transformed into a mere commodity. The third section showed, however, that in practice we do end up paying with our data for services. This becomes explicit when we are presented with the binary choice of either consenting to the processing of our data or paying with our money for a service. Here, data is transformed into an alternative means of payment.

At the same time, this essay showed that there is no obvious solution to this. The law still struggles with the precise classification of data. On the one hand it stresses the fact that data cannot be a commodity. On the other hand, it is obvious that economic value can be extracted from our data. When companies collect our data to commercially exploit it for marketing purposes, it oftentimes provides them with the necessary financial gains that enable them to provide their content or services for 'free'. Whereas we as consumers benefit financially from using many online services without having to pay with our money for them, the risk from fully accepting such business models is equally clear. It would ultimately risk turning the fundamental right to data protection into a commodity we have to pay for.

[4] (According to a CEO of the "Pay or Okay" provider, when faced with the choice of either consenting or paying 1,99€ for the use of online services, 99.9% of the users chose to 'pay' with their data. See 'noyb files GDPR complaint against Meta over "Pay or Okay" [2023] accessible via https://noyb.eu/en/noyb-files-gdpr-complaint-against-meta-over-pay-or-okay.

ISSN 2281-5147

Is Facebook's reformed commercial practice of 'pay-or-consent' in line with European data protection law in particular and European law in general? The European Court of Justice in its 2023 *Meta* judgment highlighted that 'freely given consent' would imply that consumers are offered, if necessary for an *appropriate* fee, an equivalent alternative where the personal data is not being processed.[5] Are 9,99€ per month an *appropriate* fee and a fair alternative to consenting to the exploitation of our personal data? Thus far, no conclusive answer has been given. For Meta, its Pay-or-Consent model is compliant with the judgment by the European Court. For the European Data Protection Board, offering only a binary choice between either consenting or paying a fee will in most cases not be compliant with European data protection law as it would transform a fundamental right into a feature consumers have to pay for.[6] For the European Commission, Meta's 'Pay or Consent' model would breach the new 2023 Digital Markets Act as it would not offer a true choice to consumers – a truly equivalent alternative should allow consumers to choose an alternative version of the service which is free of (monetary) charge *and* relies on non-personalisation of advertisement.[7]

Which interpretation of the law is correct? If Facebook does not comply with the demands by the European Commission to adapt its commercial practices, the answer will yet have to be given either by the European Court of Justice or through new legislation that clarifies more precisely how we find the equilibrium between the protection of our personal information through the fundamental right to data protection and its economic exploitation by companies.

---

[5] See CJEU case C252/21 Meta v Bundeskartellamt, [2023] ECLI:EU:C:2023:537, para 150.

[6] See Opinion 08/2024 by the EDPB which in reaction to the judgment by the Court argued that offering only a paid alternative to services which process personal data for behavioural advertising should not become the new default way for companies. Whereas the EDPB does not oppose in principle the imposition of a fee to access an 'equivalent alternative', such a fee should not inhibit data subjects from making a 'genuine choice' – whether or not such a fee would be 'fair' in light of the GDPR should fall within enforcement duties of national data protection authorities.

[7] With the Digital Markets Act (DMA) the EU further regulates large digital platforms ('gatekeepers'). Article 5(2) of the DMA requires gatekeepers to obtain consent from consumers if they intend to use their data, for instance, for online advertising services. At the same time, gatekeepers must offer consumers a 'less personalised but equivalent alternative'. For the Commission, Meta's paid subscription model does therefore not constitute an 'equivalent alternative' to the 'free' model that uses personal data for targeted advertisement. See European Commission. (2024). Commission sends preliminary findings to Meta over its "Pay or Consent" model for breach of the Digital Markets Act. See also Euractiv. (2024]. European Commission accuses Meta of violating digital competition rules with 'pay or OK' model. Retrieved from https://www.euractiv.com/section/data-privacy/news/european-commission-accuses-meta-of-violating-digital-competition-rules-with-pay-or-ok-model/

ISSN 2281-5147

6. Selected Readings

De Hert, P., & Gutwirth, S. (2009). Data Protection in the Case Law of Strasbourg and Luxemburg: Constitutionalisation in Action. In S. Gutwirth, Y. Poullet, P. De Hert, C. de Terwangne, & S. Nouwt (Eds.), *Reinventing Data Protection?* (pp. 3–44). Springer.

De Hert, P., & Lazcoz, G. (2022). When GDPR-Principles Blind Each Other: Accountability, Not Transparency, at the Heart of Algorithmic Governance. *European Data Protection Law Review*, *8*(1), 31–40. https://doi.org/10.21552/edpl/2022/1/7

EDPS. (2017). *Opinion 4/2017 on the Proposal for a Directive on certain aspects concerning contracts for the supply of digital content*.

Euractiv. (2024, July 1). *European Commission accuses Meta of violating digital competition rules with 'pay or OK' model*. Www.Euractiv.Com. https://www.euractiv.com/section/data-privacy/news/european-commission-accuses-meta-of-violating-digital-competition-rules-with-pay-or-ok-model/

European Commission. (2024). *Commission sends preliminary findings to Meta over its "Pay or Consent" model for breach of the Digital Markets Act*.

González Fuster, G. (2014). *The Emergence of Personal Data Protection as a Fundamental Right of the EU*. Springer.

Lessig, L. (1999). The Architecture of Privacy: Remaking Privacy in Cyberspace. *Vanderbilt Journal of Entertainment & Technology Law*, *1*(1), 56–65.

Lynskey, O. (2015). *The Foundations of EU Data Protection Law*. Oxford University Press.

Mayer-Schönberger, V. (2010). Beyond Privacy, beyond Rights—Toward a 'Systems' Theory of Information Governance. *California Law Review*, *98*(6).

ISSN 2281-5147

ISSN 2281-5147

YOUR DATA RIGHTS: HOW DOES THE GDPR AFFECT THE SOCIAL MEDIA MARKET?

Qifan Yang*

Abstract

With the development of digitalisation, personal data has gradually become a valuable resource for social media companies to extract value and obtain market dominance. Personal data processing can raise serious concerns about privacy leaks and misuse. In response, the adoption of the General Data Protection Regulation (GDPR) enhances personal data protection and market competition, but also potentially influences economic interests, the rights of data subjects, as well as market dynamics. The chapter uses the social media market to understand the complex relationship between the GDPR and market competition.

Table of Contents

* Qifan Yang is an Early-Stage Researcher in the LIDER Laboratory at Scuola Superiore Sant'Anna di Pisa and a Ph.D. student in the National Doctorate in Artificial Intelligence. This work is supported by the European Union's funded project Legality Attentive Data Scientists (LeADS) under Grant Agreement no. 956562. The author would like to thank Professor Giovanni Comandé from Scuola Superiore Sant'Anna di Pisa for his valuable contribution in reviewing this article. The author also wishes to thank the anonymous reviewers for their valuable suggestions.

ISSN 2281-5147

Keywords

GDPR – Personal Data – Market Concentration – Competition Regulation – Social Media Market

1. Introduction: Direct and Indirect Network Effects

The technologies developed to allow the management of information and the exchange of communications have formed networks that connect people around the world and allow them to interact regardless of distance or time (Ibert et al., 2022). When using digital services - downloading software from an app shop, using Google to find the latest news, sharing a story on X or Facebook, watching a video on YouTube or TikTok, or buying an item on Amazon - one cannot help but notice the fact that digital services are dominated by a handful of well-known internet companies.

In real life, there tends to be greater flexibility of choice - one can buy coffee from a large chain such as Starbucks or homemade coffee from a local cafe operated by a neighbour. The freedom to choose between a global brand and a small local business is something we might take for granted offline. The diversity of providers allows consumers to never worry about the risk of no coffee owing to a boycott of Starbucks. But online, the providers of products or services seem to be more centralised, and

ISSN 2281-5147

users are drawn to the services of these large companies without much thought. But why is that?

Take a platform like YouTube, for example. At first glance, it might just seem like a space where people share and watch videos. In fact, YouTube operates in a complex balance involving three key groups of users: content creators, content viewers, and advertisers. For content creators, YouTube provides a stage to distribute their work to a global audience, giving content creators fame, fun, and revenue. Millions of content viewers use YouTube to watch a wide range of videos and become potential targets for advertisers to promote their products. Unlike one-sided markets that provide products and services to consumers like retailers (e.g. Walmart, Carrefour) and legacy media (e.g. newspapers, TV), Internet companies function more as an interactive platform for communication, sharing or trading among different parties (Jullien et al., 2021; Saura et al., 2021).

The more individuals join a platform, the more opportunities to interact, connect and share content, thus improving the user experience and creating a more valuable network. The value of a platform increases with the number of individuals joining, which is known as the "direct network effect"[1]. In essence, it is a virtuous circle, where users attract more users, which in turn enhances the service provided to users, like a snowball rolling down a hill.

As the number of users has increased, so has the change on the other side of the platform. Sellers, recognising potential consumers, are naturally attracted by platforms with a large pool of potential buyers, like YouTube, Instagram, and Amazon (Calvano & Polo, 2021). In addition, this large user base generates huge amounts of data that are of great interest for advertisers, who are eagerly seeking to match a product or service with a target audience (Sembada & Koay, 2021). The growing user base is more like a magnet pulling in outside businesses, advertisers, and service providers. This is the "indirect network effect"[2] - where the value of a platform is increased by businesses, advertisers, and other external entities being attracted by the large user base (Veisdal, 2020).

[1] The direct network effect means that the value or utility that users derive depends on the number of other users on the same side.
[2] The indirect network effect means that the utility of at least one group grows as the other groups grow.

ISSN 2281-5147

To sum it up, the direct network effect revolves around the interplay among users, whereas the indirect network effect involves the interplay between users and other stakeholders like sellers and advertisers, the number of which will increase as the user base expands.

When companies offer better products or services, users are willing to invest more time, data, and connections on social media platforms. As users spend more time building connections, sharing content and leaving digital footprints, the cost of moving to another platform increases - not just financially but psychologically as well (Buiten, 2021). When these economic and psychological costs of switching from one alternative to another become high, social media platforms can trap individuals in their own networks, which is known as the "lock-in effect"[3]. Consider that a user spends years building a social network on Facebook or Instagram. Upon leaving that platform, that user not only loses friends and followers, but also data and content. Rebuilding a new social network on another platform can be time-consuming and effortful, which creates a digital trap.

With the network effect and the lock-in effect, the expanding control of user personal data from online platform companies creates an insurmountable barrier to entry for market competition in this area (Newman, 2014), since personal data have gradually become a competitive asset in the online platform market. It has raised critical concerns about the protection of personal data and the potential abuse of market dominance In 2020, Germany's Federal Court of Justice highlighted that Facebook's massive user data collection exacerbated already distinct "lock-in effects" and their large user database enhances the possibilities to finance the social network by using the profits generated from advertising contracts.

The General Data Protection Regulation (GDPR) represents a significant milestone in the ongoing effort to protect personal data. The GDPR aims to empower individuals by granting them greater control over their personal data while imposing stricter rules on how companies can process personal data. With the GDPR in place, businesses based on online platforms and personal data must rethink and restructure their strategies. The GDPR was expected to help reduce market concentration, but

[3] The lock-in effect means that customers are dependent on one product from a provider and cannot use the product from another provider unless they pay significant switching costs.

ISSN 2281-5147

what kind of impact did it actually have in terms of the company's market share[4]? This chapter introduces the framework and regulatory provisions of the GDPR and then selects the social media market as a use case. A statistical method for assessing the effects of interventions in comparative case studies (synthetic control method[5]) is employed to identify and quantify the causal effect of the GDPR reform on social media market share in the EU. The conclusion explores the impact of the GDPR along with a brief discussion of its reasons.

2. Personal Data Regulation Practices Under the GDPR Framework

Since the GDPR was adopted, each EU member state has taken significant steps to align its national laws with this robust framework. This alignment includes enhancing the capabilities of domestic data protection authorities (DPAs), which play a crucial role in investigating potential violations and implementing regulatory actions to protect personal data.

Since 2018 until June 2024, EU member states have dealt with 2,141 cases related to personal data protection violations. These cases have resulted in fines that collectively exceed €4,590 million - a clear indication that non-compliance can be quite costly. The main violations that most frequently trigger GDPR regulation and fines include non-compliance with general data processing principles[6], insufficient legal basis for data processing[7], insufficient technical and organisational measures to ensure information security[8], and others (CMS Legal Services EEIG, 2024).

Looking at the timeline from 2018 to June 2024, the number of reported cases has noticeably increased each year: just 9 cases in 2018, 143 cases in 2019, 342 cases in

[4] Generally, market share is defined as the percentage of a company's business out of the total revenue or sales in the market.
[5] The synthetic control method is a statistical method for assessing the effects of interventions in comparative case studies and it will be briefly described in 3.3.
[6] Non-compliance with general data processing principles includes failing to adhere to basic principles such as data minimisation, accuracy, and purpose limitation.
[7] Insufficient legal basis for data processing means that the company or the organisation lacks a sound reason for the data processing, such as consent or legitimate interest.
[8] Insufficient technical and organisational measures to ensure information security means the company or organisation does not provide appropriate and adequate technical and organisational measures to protect information security.

ISSN 2281-5147

2020, 462 cases in 2021, 536 cases in 2022, 510 cases in 2023 and 154 cases by June 2024. This upward trend shows that more people are becoming aware of their rights under the GDPR and DPAs are getting better at enforcing them (CMS Legal Services EEIG, 2024).

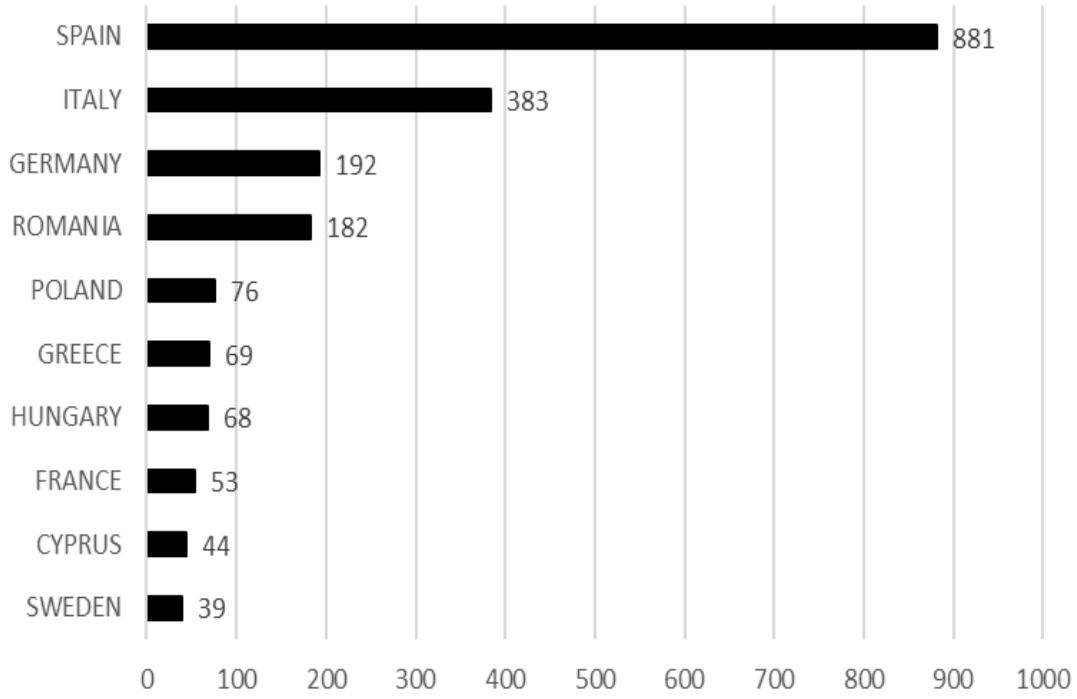


Figure 1. Top 10 DPAs by total number of fines

(Data Source: GDPR Enforcement Tracker as of June 2024)

ISSN 2281-5147

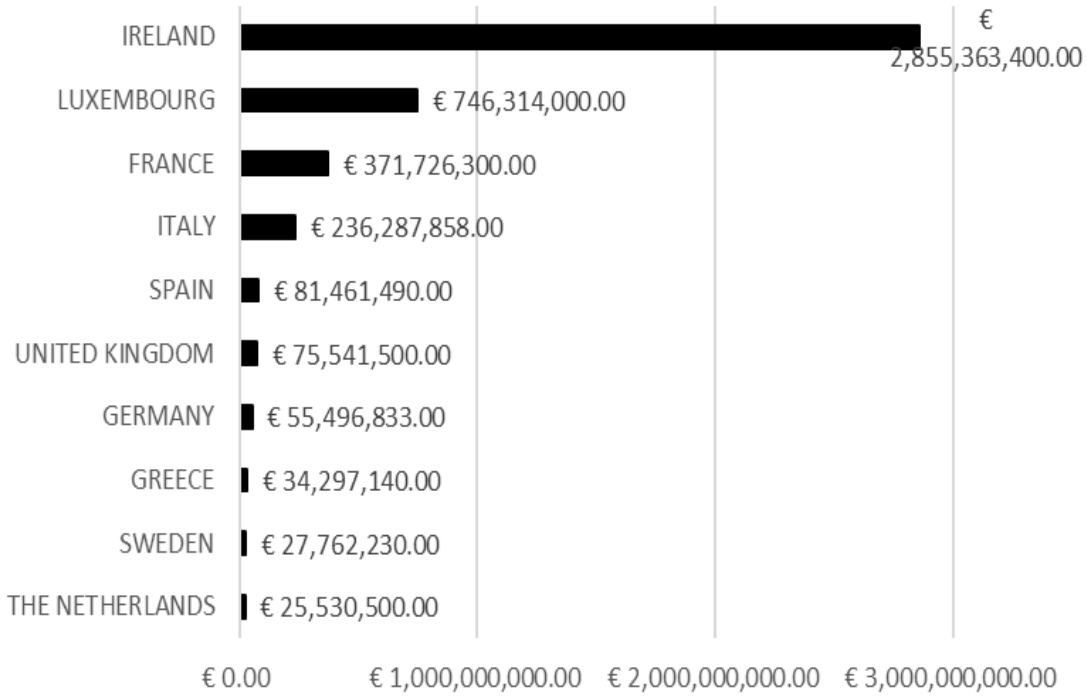


Figure 2. Top 10 DPAs by total sum of fines

(Data Source: GDPR Enforcement Tracker as of June 2024)

In terms of the number of cases and the total amount of fines, certain countries have been particularly proactive in the GDPR enforcement. As of June 2024, DPAs from Spain, Italy, Germany, and Romania have been active in dealing with data protection cases, see **Figure 1**. In terms of the total amount of non-compliance, Ireland, Luxembourg, France, and Italy stand out for imposing substantial fines, see **Figure 2**.

This further illustrates that data protection can also result in significant financial losses for those who fail to comply. Companies that use a lot of data, like tech companies, had to spend more on improving data security, training their workers, and recording their processing. Koski and Valmari (2020) used company-level data on European

ISSN 2281-5147

and US companies from 2014-2018 to reveal that the GDPR imposes financial burdens on European companies and causes a decline of profit margins for data-intensive companies.

By imposing stringent rules on the digital world, the GDPR affects how companies compete and how they share the benefits (Jones & Tonetti, 2020; Li & Feng, 2021), just like how basic rule changes affect the play of the players in a game. Early analyses argue that the GDPR can enhance competition by lowering compliance costs by using clear rules, increasing consumer trust, and fostering the uptake of new technologies. If privacy regulation is coupled with appropriate incentives, it may positively influence the development and adoption of information exchanges (Godinho de Matos & Adjerid, 2022).

3. How Personal Data Protection Affects the Market Share of Big Social Media Platforms

The present research intends to explore the actual impact of the GDPR on the EU social media market. On the social media market, platforms facilitate widespread engagement and data exchange among users (European Commission, 2021). By studying social media, we can learn how personal data protection affects market share dynamics.

3.1 Concentration Ratio: an Indicator to Measure Market Competition

The Concentration Ratio (CR_n) is commonly employed to measure market concentration[9] and assess the changes of companies' market share. To put it simply, all the sales (revenue or traffic) in a market in a region (country, region, city) can be seen as a pie, and each company in that market takes its own slice of the pie based on the proportion of its sales (revenue or traffic) to the total one in the market (See

[9] Market concentration is the market share of a certain number of companies in a given market.

ISSN 2281-5147

Figure 3). CR_n is calculated by summing the market shares of a specified number of the largest companies in a particular industry, which shows the total market share held by the n largest companies[10] in the market.

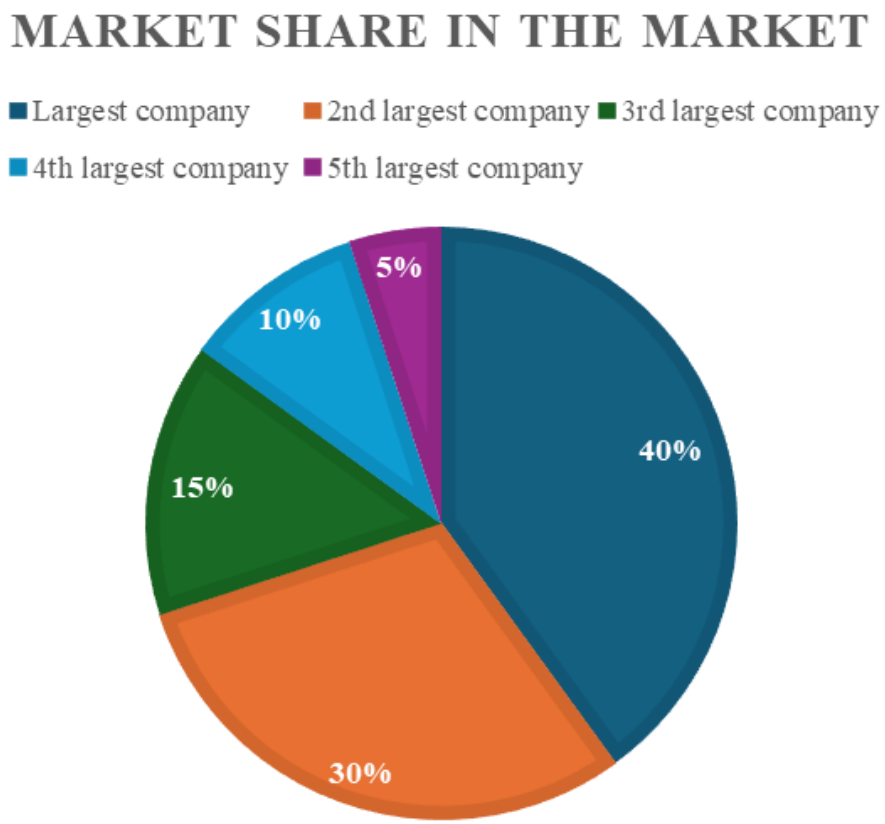


Figure 3. Example of market share and market concentration

If there are 5 companies in the market, and the company with the largest market share occupies 40 % of the market share, the second largest company occupies 30 %, the third largest company occupies 15 %, the fourth largest company occupies 10 %, and the fifth largest company occupies 5 %. Then the CR_1 is 40 %, the CR_2 is the sum of the first company's market share and the second company's market share (70 %), and

[10] n in CR_n represents the number of companies included in the concentration ratio calculation.

ISSN 2281-5147

the CR_3 is the sum of the first company's market share, the second company's market share, and the third company's market share (85 %), and so on. CR_4 is commonly used to measure the market concentration in an industry.

CR_n is like a window into the competitive landscape of a market. Higher concentration indicates that few top companies in the industry dominate the market, i.e. the market is more monopolised and competition in the market is reduced.

3.2 Market Share Dynamics in the Social Media Market Before and After the GDPR

Between 2009 and 2015, market concentration in the European social media market increased significantly before the adoption of the GDPR, see **Figure 4**:

- The CR_1 index, which shows the market share of the largest company, grew from about 30% to around 85%;
- The CR_2 index, representing the combined market share of the two largest companies, increased from around 55% to around 90%;
- The CR_4 index, accounting for the four largest companies, went from roughly 75% to around 95%.

Before the GDPR, the growing dominance of a few major companies in the social media market could be clearly observed through the rise in the CR_1 index, the CR_2 index, and the CR_4 index. While having a dominant market position does not automatically break antitrust laws, it is clear that a few key players are becoming more established in this market.

After the adoption of the GDPR (the right side of the red dotted line), the CR_1 index, the CR_2 index, and the CR_4 index gradually shift from a rising trend to a declining trend until 2022. The CR_1 index and the CR_2 index show significant decreases, and the decrease in the CR_4 index is smaller compared to that of the CR_1 index and the CR_2 index. Although other factors may also influence these changes, the glimpse shows that the GDPR can have a negative impact on market concentration, see **Figure 4**.

ISSN 2281-5147

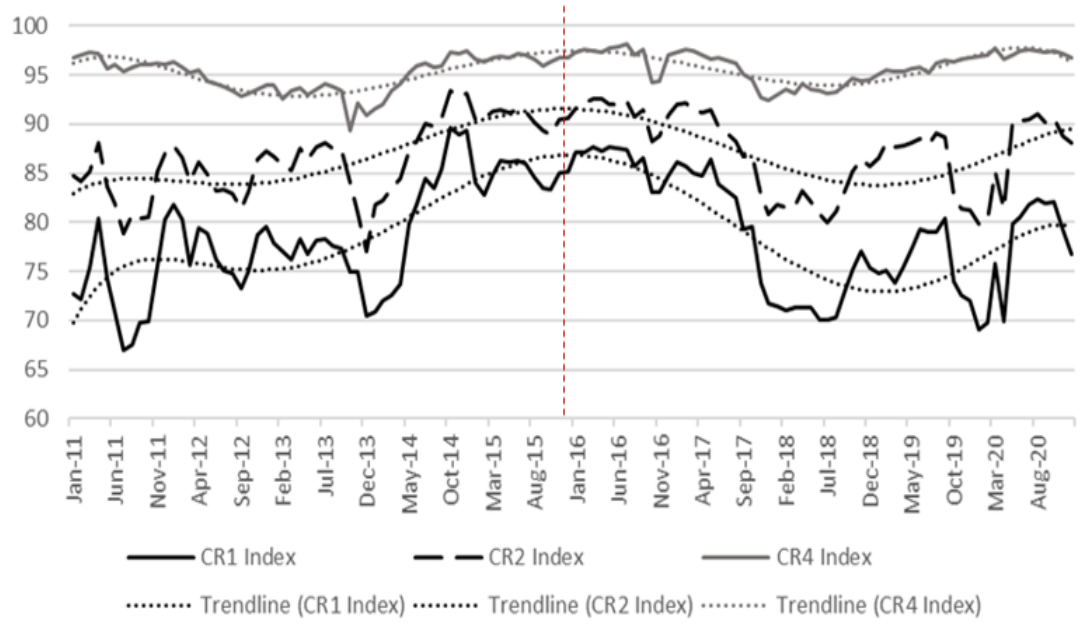


Figure 4. Concentration ratio of the social media market before and after the GDPR (Data source: StatCounter Global Stats)

3.3 Synthetic Control Methods to Estimate the Impact of the GDPR

When assessing the impact of the GDPR, the best way is to compare two groups: an EU group where the GDPR is in force (EU group) and a different group where the GDPR is not in force but has identical characteristics to the first group otherwise (twin EU). By comparing the two groups, it is possible to accurately measure the difference between the two groups in social media market concentration after the adoption of GDPR. Unfortunately, no such twin EU exists. This is where econometric creativity comes into play, specifically creating a synthetic twin EU through a tool known as the synthetic control method (Abadie et al., 2010). Although the synthetic control method[11] involves many technical parts, its main logic is very easy to understand.

[11] Synthetic control method (SCM) a statistical method to evaluate treatment effects in comparative case studies, which allows the construction of a counterfactual by selecting a weighted average of the outcome variable from a group of units similar to the treated unit.

ISSN 2281-5147

The logic of the synthetic control method can be illustrated with a simple example. Now there is a glass of mixed juice consisting of 30% apple juice, 30% orange juice, and 40% grape juice, and a test is required to examine the effect of adding mango juice to this glass of juice. It is impossible to obtain another juice with the same composition, but a similar (synthetic) juice can be created by mixing apple, orange and grape juices in the same proportions, and the effect of adding mango juice can be observed by using the synthetic juice as the twin group.

In our case, CR_4 is collected as the outcome variable measuring changes in market concentration,[12] so the CR_4 of the EU group is the "mixed juice". The different types of juices are the countries or regions that have not adopted the GDPR but are very similar to the EU in many ways, like GDP, population, technology, size of internet users, government efficiency, and level of regulation. We used 24 countries or regions[13] with some similar characteristics to the EU as a donor pool, like a pool of different juices.

Based on this pool, the statistical methods are used to select countries or regions that contribute to the CR_4 of the EU group (type of juice contained in the "mixed juice") and to find a combination of the selected countries (proportion of different juices in the "mixed juice") that matched the various characteristics and market trends of the EU before the adoption of the GDPR. Combining these selected countries[14] with the given weights will create a synthetic EU group (twin EU). After building the synthetic group, the real EU market with the GDPR is compared with the synthetic EU market without the GDPR to find the differences, and the gap between the real market and the synthetic market is the effect of the GDPR.

Figure 5 illustrates that the trend of the CR_4 index in the synthetic EU and the actual EU matches closely before the GDPR was introduced. This indicates that the

[12] To ensure the similarities between the treated unit and the units in the donor pool, the parallel trends of CR_1, CR_2, CR_3, and CR_4 in the donor pool and the treated unit are tested. CR_4 is closer to the treated units than the other indexes and is less affected by a single event targeting a particular player, so it is more appropriate as an outcome variable.

[13] The 30 countries or regions are the EU, United Arab Emirates, Argentina, Australia, Brazil, Canada, Switzerland, Chile, China, Egypt, Arab Rep., United Kingdom, Hong Kong SAR(China), Indonesia, India, Israel, Japan, Korea, Rep., Mexico, Malaysia, New Zealand, Philippines, Russian Federation, Saudi Arabia, Singapore, Thailand, Turkey, United States, Viet Nam, South Africa.

[14] In our case, the CR4 index in the EU social media market is best reproduced by the combination of Korea, Rep.(0.514), Chile(0.338), Hong Kong SAR, China (0.094) and Russian Federation (0.055).

ISSN 2281-5147

synthetic EU without GDPR regulations is a good copy of the real EU's market concentration (the CR_4 index). The disparity between the real EU CR_4 index and its synthetic unit emerged before the time of the GDPR adoption, which may be likely influenced by the European Parliament's vote for the GDPR in 2014 and the agreement on the GDPR by the European Parliament, the Council and the European Commission in 2015.

Figure 5 also reveals the CR_4 index in the synthetic EU had a very small decline, indicating a natural trend in market concentration over time. However, in the real EU, the CR_4 index dropped sharply after GDPR adoption, indicating that the GDPR had a significant negative impact on market concentration in the EU social media market.

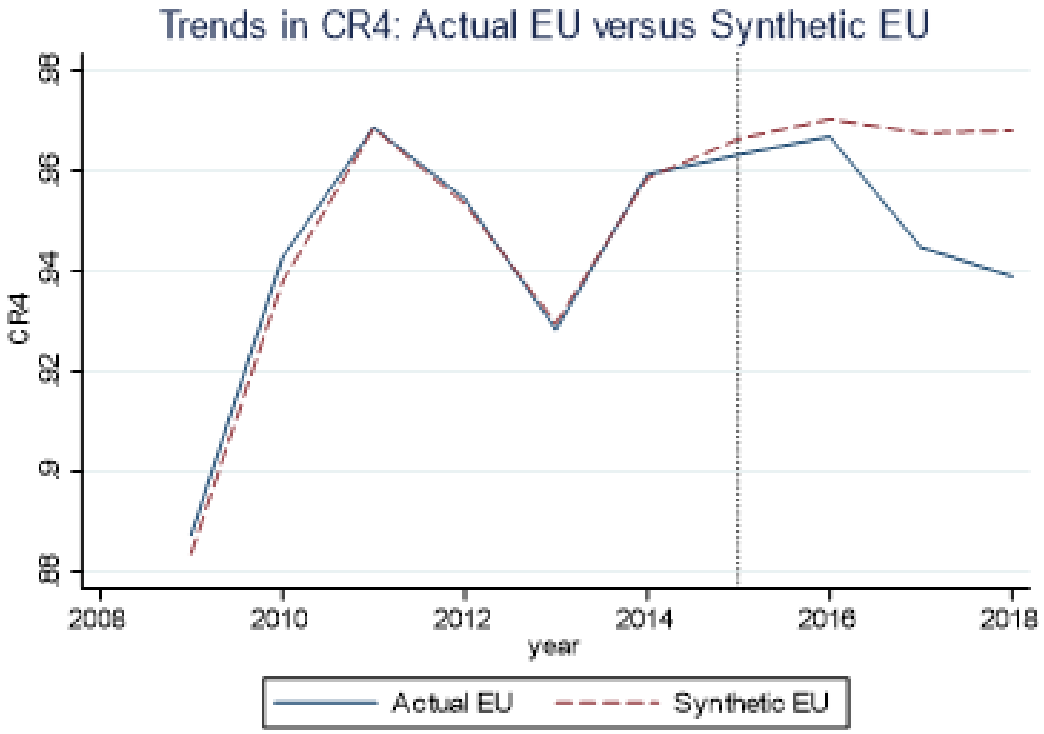


Figure 5. Trends of CR_4 between actual EU and synthetic EU

ISSN 2281-5147

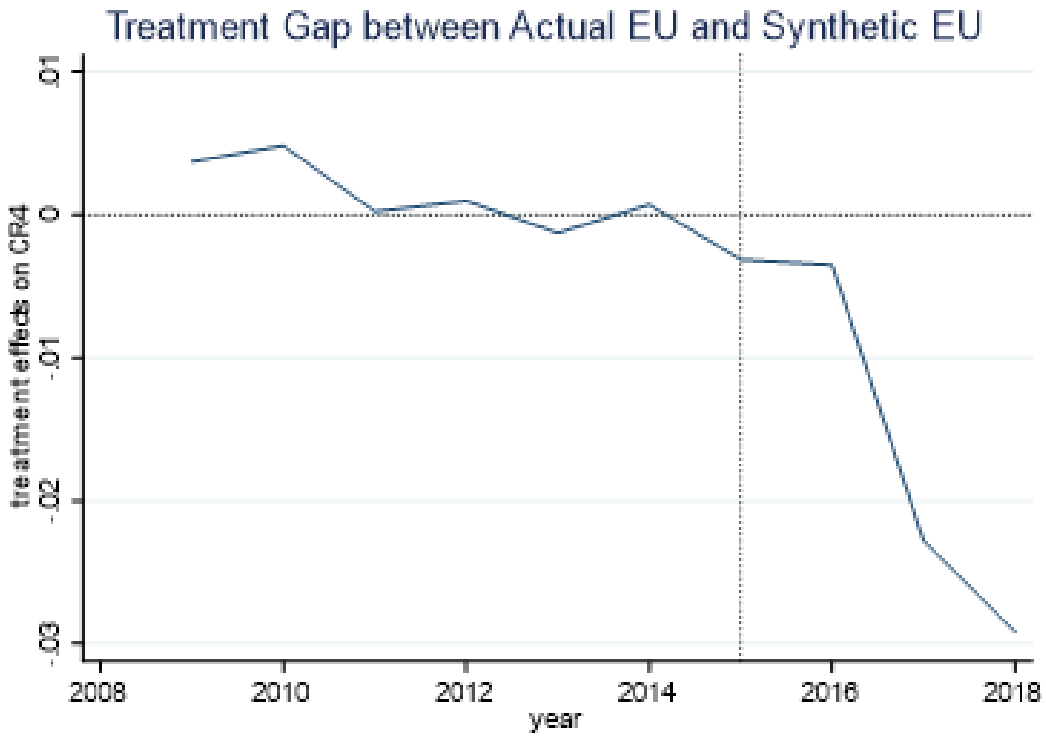


Figure 6. Treatment gap of CR_4 between actual EU and synthetic EU[15]

4. Conclusion and Future Work

Based on the above analyses and empirical evidence, we can easily find that the enhancement of personal data protection not only benefits the data subjects, but also plays a role in slowing down the market concentration, at least in the social media market. The negative effect of GDPR on social media market concentration may stem from the following reasons:

- The GDPR's transparency and accountability requirements limit social media platforms' power to misuse user data, empowering users with

[15] To test the significance of the result, I test each country and region with similarities in the sample by applying the same synthetic control method. For the distribution of the post/pre-GDPR ratios, the difference between before and after GDPR in the EU unit is about 48.165 times, a much larger difference than any of the 23 control regions. The probability of this difference happening by random chance is very low, 1/24 (about 0.0416), making the conclusion reliable.

ISSN 2281-5147

greater control and compelling platforms to weigh the costs of extensive data collection.

- Data portability provisions enable users to transfer their data between platforms, reducing data exclusivity and promoting market competition.
- Strict data protection regulations impose compliance costs on dominant platforms, leveling the competitive landscape by restricting their ability to exploit data collection advantages.
- By regulating dominant platforms' data processing, new entrants can compete more effectively without facing exclusive data constraints. Apparently, the implications of the regulation of personal data spill over into the market sphere.

This chapter attempts to provide a new perspective on the impact of the GDPR on social media market concentration in the EU, but also has some limitations. Future research could delve into the divergence of personal data regulations across different jurisdictions. Understanding these differences and their implications can shed light on the feasibility and challenges of harmonising global data governance standards. Additionally, the synthetic control evaluation has scope for further refinement in this study.

5. Selected Readings

(1) Abadie, A., Diamond, A., & Hainmueller, J. (2010). Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American Statistical Association*, *105*(490), 493–505.

(2) Buiten, M. C. (2021). Exploitative abuses in digital markets: Between competition law and data protection law. *Journal of Antitrust Enforcement*, *9*(2), 270–288.

(3) Calvano, E., & Polo, M. (2021). Market power, competition and innovation in digital markets: A survey. *Information Economics and Policy*, *54*, 100853.

ISSN 2281-5147

(4) CMS Legal Services EEIG. (2024, August 7). *List and Overview of Fines and Penalties under the EU General Data Protection Regulation (GDPR, DSGVO).*

(5) European Commission. (2021). *COMMUNICATION FROM THE COMMISSION TO THE EUROPEAN PARLIAMENT, THE COUNCIL, THE EUROPEAN ECONOMIC AND SOCIAL COMMITTEE AND THE COMMITTEE OF THE REGIONS 2030 Digital Compass: The European way for the Digital Decade.*

(6) Godinho de Matos, M., & Adjerid, I. (2022). Consumer Consent and Firm Targeting After GDPR: The Case of a Large Telecom Provider. *Management Science*, *68*(5), 3330–3378.

(7) Ibert, O., Oechslen, A., Repenning, A., & Schmidt, S. (2022). Platform ecology: A user-centric and relational conceptualization of online platforms. *Global Networks*, *22*(3), 564–579.

(8) Jones, C. I., & Tonetti, C. (2020). Nonrivalry and the Economics of Data. *American Economic Review*, *110*(9), 2819–2858.

(9) Jullien, B., Pavan, A., & Rysman, M. (2021). Two-sided markets, pricing, and network effects☆. In K. Ho, A. Hortaçsu, & A. Lizzeri (Eds.), *Handbook of Industrial Organization* (Vol. 4, pp. 485–592). Elsevier.

(10) Koski, H., & Valmari, N. (2020). *Short-term Impacts of the GDPR on Firm Performance* (Working Paper No. 77). ETLA Working Papers.

(11) Li, S., & Feng, J. (2021). Is Data Ownership Empowerment Welfare-Enhancing? *Hawaii International Conference on System Sciences 2021 (HICSS-54).*

(12) Newman, N. (2014). Search, Antitrust, and the Economics of the Control of User Data. *Yale Journal on Regulation*, *31*(2), 401–454.

(13) Saura, J. R., Ribeiro-Soriano, D., & Palacios-Marqués, D. (2021). From user-generated data to data-driven innovation: A research agenda to understand user privacy in digital markets. *International Journal of Information Management*, *60*, 102331.

(14) Sembada, A. Y., & Koay, K. Y. (2021). How perceived behavioral control affects trust to purchase in social media stores. *Journal of Business Research*, *130*, 574–582.

ISSN 2281-5147

(15) Veisdal, J. (2020). The dynamics of entry for digital platforms in two-sided markets: A multi-case study. *Electronic Markets*, *30*(3), 539–556.

ISSN 2281-5147

ISSN 2281-5147

BYOD – BRING YOUR OWN DATA. THE STRUGGLE OF RE-USING DATA IN A WORLD OF HETEROGENEOUS SYSTEMS

Tommaso Crepax*

Abstract

Data portability is often perceived as a solved problem, an aspect of digital life similar to transferring a phone number or syncing accounts across devices. However, this paper argues that the reality is far more complex—and fascinating. By rebranding data portability as the "Bring Your Own Data" (BYOD) phenomenon, this paper exposes the technical, legal, and economic challenges of making data transferable, functional, and meaningful across heterogeneous systems. Using analogies like organizing a BBQ, it analyses issues of syntax, semantics, and intensionality that encumber data exchange. The paper examines the evolution of EU regulations—GDPR, Digital Markets Act (DMA), and Data Act (DA)—and their varied approaches to data portability, from transmission to real-time access, revealing how legislative intents shift between empowering individuals and enabling market competition. It critiques the gaps in these frameworks, particularly in addressing the content and completeness of data, and explores the tensions between tight and loose integration strategies in fostering interoperability. Ultimately, this paper proposes that understanding data portability requires a multidisciplinary approach. It is not just about moving data, but about enabling control and usability in a fragmented digital ecosystem. The findings emphasize the need for thoughtful regulation and design to bridge the divide between legal ideals and technical realities, supporting a future where data flows freely and meaningfully across digital environments.

Table of Contents

* Researcher at Scuola Superiore Sant'Anna; Joint Ph.D. candidate in Artificial Intelligence (National Doctorate, University of Pisa) and Law (Vrije Universiteit Brussel), tommaso.crepax@santannapisa.it
This work is supported by the European Union's funded project Legality Attentive Data Scientists (LeADS) under Grant Agreement no. 956562.

ISSN 2281-5147

Keywords

Data Portability – Data Heretogeneity – Information Intensionality – Digital Markets Act – Data Act

1. Portability: interesting, complex, and needed

A whole Ph.D thesis on data portability? You might be thinking: "Boring, for sure." With all the buzz around artificial intelligence and the strides made in blockchain technology, does data portability really matter? And then there is the thought that it sounds easy. You have switched mobile operators and kept your number, carried data around on a pendrive, and accessed your Google accounts across multiple devices effortlessly—so, why spend three years on a boring problem that seems already solved?

ISSN 2281-5147

The piece I am about to present will highlight an often overlooked yet critical aspect: data portability is not only interesting but also remarkably challenging to implement. If we were to rebrand data portability with a catchier term, similar to "blockchain smart contracts" or "artificial intelligence," it might very well become the centerpiece of professional and non-professional conversations. This is because data portability intersects with many of the "hot topics" that captivate academics in law, technology, and economics, as well as software developers, competition authorities, legislators, data protection officers, and policy experts. In real life, we regularly engage with concepts like data governance, fundamental rights, data control, fairness, competition, personal and non-personal data protection, information system design, reference architectures, data modeling, artificial intelligence, market power, economics of data, and data security. Each of these areas is intricately linked to the seemingly mundane yet profoundly significant concept of data portability, underscoring its relevance and complexity in today's digital landscape.

For the non professionals, I am afraid to say that data portability is like politics: you might not be interested in it, but it will affect you anyways—so you should better understand it!

The perception most people have when I introduce my line of work also tends to miss the underlying complexities. Unlike simply carrying a phone number from one provider to another, or syncing Google accounts across devices, true data portability involves deeper, more intricate processes. It is about more than just access—it is about making sure that data is not only transferrable but also functional and meaningful across diverse systems and contexts, which is far from straightforward. In my Ph.D thesis I try to peel back the layers of what looks like a solved problem, revealing the technical, legal, and economic intricacies that still need untangling. In academia, we do not complicate things; we are delving into the details that are crucial for innovation and user empowerment in the digital age.

Certainly, the mere complexity and intrigue of a topic do not alone justify years of research. However, the *necessity* of data portability amplifies its importance, making it worthy of thorough investigation. Data portability is not just an academic concept; it is a crucial need in various practical contexts.

Consider your role as a consumer wanting to leave Facebook, yet feeling tethered because a decade and a half of memories—photos, conversations, and social

ISSN 2281-5147

networks—are locked away under a pseudonymous userID on a distant server. Or consider my perspective as a European citizen, troubled by the opaque handling and sale of my personal data to unknown parties, yearning to reclaim control over who can access and profit from my personal information. Think about a patient moving across borders, who relies on the continuity of care that a digital prescription from her trusted family doctor provides. Or about young software developers, feeling stuck on a platform like Amazon Web Services and seeking alternatives that align more closely with their values regarding labor practices. At a broader scale, consider the EU Governance and the European Digital Single Market, which are betting on new enterprises to drive digital economic transformation. This vision depends on the ability to freely reuse and share data currently monopolized by a few tech giants across the Atlantic.

Data portability, therefore, is not just a theoretical interest—it is a fundamental component that could reshape our digital interactions, enhance our control over personal data, and redistribute economic power in the digital age. Yes, indeed, data portability is a tool that can help make all this, and more, a reality. The problem is *how*.

2. The BBQ Dilemma—A Taste of Data Portability Challenges

Data portability, a concept gaining traction in the realms of law, technology, and economics, refers to the ability to move data seamlessly from one platform to another. However not all data that moves from one system to another is immediately usable. This is because different systems often "speak" different data "languages"—a challenge of integration.

Given the challenge of crafting a piece accessible to everyone ("Including your grandma!", cit.) I thought: why not use food as the theme? Therefore, as we dive into this exploration, I would like you to think of organizing a BBQ. It is a fun and relatable way to understand why making data portable is not as simple as just moving it from one place to another—it needs to be the right type, in the right form, at the right time, and in the right amount. Just like the perfect BBQ. With the appropriate distinctions, we can use the BBQ analogy to simplify complex technological concepts related to data governance and data handling, which might otherwise be obscured by technical

ISSN 2281-5147

jargon. By comparing technical impracticalities to a familiar social event, we can make topics like data models and formats, knowledge representation, and the concepts of classes and objects more relatable. More specifically, discussing data syntax, structures, and semantics through the lens of preparing different dishes helps clarify these ideas in a tangible way.

Imagine you are organizing a summer BBQ in Brussels, excited to prepare a delicious meal for all your international friends. You have sent out invites, asked guests to bring a dish, and you are ready to cook up. But as the guests arrive, you notice a problem—not all the dishes can be cooked on your grill, and some do not even fit the meal you had planned. This culinary chaos is a perfect way to understand the complexities of data portability.

You being an Italian assigned your French friend the task of bringing dessert and they arrived with a piece of cheese! This is what is known as a semantic issue—where "dessert" means something different to each person. Likewise, you asked another friend for pork ribs and they showed up with the cutest piglet, alive on a leash: this is a problem of data formatting and syntax. Here, the request was understood, but the format in which it was delivered—alive rather than pre-processed—was not what was intended.

3. From food to data

At the core of any data portability issue is the fact that data serves as our means to measure and represent real-world information and objects. Consider a classroom scenario where you ask N students to draw a dog. The result? You will inevitably end up with N different drawings. Each representation will share some similarities—characteristics generally recognized as canine, like the number of legs, the shape of the face, whiskers, and perhaps even the bark. However, there is no universally accepted standard detailing exactly which features must be present to definitively classify something as a dog. We recognize a dog by a collective set of characteristics, but each dog is unique in its own right. The idea of a dog, borrowing concepts and tools from Object Oriented Programming, we call it a Class; each single dog, as specifically depicted by the students, is called an Object of the Class:Dog.

ISSN 2281-5147

In a similar vein, I recently conducted an experiment during one data portability workshops where I asked participants to describe themselves using just five words. You can even try this exercise right now—pause and write down five words about yourself. The outcome will likely align with the concept: while each description will differ, much like each drawing of a dog, the nature of the characteristics chosen will also vary. Some individuals choose to focus on physical attributes like eye color or skin tone, others might describe their profession, hobbies, personal traits (like being shy or curious), or their nationality, religious belief and sexual preferences.

Here are a few answers I received during my workshop:

- Privacy valuing user, migrant, woman, funny, pizza lover
- Law, tech, European, brown eyes, curious
- Concerned, sporty, curious, atheist, engaged
- Playing piano, love pasta, work in data protection, blue eyes, black shoes
- Tired, hungry, smelly, restless, curious

This experiment illustrates the diversity and subjectivity inherent in how we define and represent data about objects, in this case: ourselves. Just as no two descriptions are exactly the same, no standardized method captures all aspects of an individual's identity perfectly. This further emphasizes the complexities involved in data portability, where not only the data itself varies but also the *dimensions and aspects considered important by different systems or contexts*. The context in which data is gathered and interpreted plays a crucial role in how it is understood. For example, the workshop was held at a privacy conference, likely influencing the type of descriptors used—perhaps focusing more on privacy-sensitive or professional aspects. However, if the same question were posed at a Star Wars convention, the responses would likely be vastly different, possibly skewing towards character traits, favorite Star Wars quotes, or affiliations within the Star Wars universe.

In the realm of software applications, the attributes that people used to describe themselves can be likened to responses to specific questions. These questions, in software terminology, we call them *data fields*. Data fields represent the *attributes* of a *class*, which in programming is a blueprint for creating specific *objects*. Classes can be: fruit, car, person, while objects can be all fruits such as apples, pears, bananas, and

ISSN 2281-5147

cars be Mercedes, Fiat, Audi, and so on. The class represents the concept, or the idea, defining what attributes (inserted in the data fields) are essential to represent the entity accurately within the software.

For example, if we are designing a class called Person, the data fields might include "name", "age", "nationality", and "hobbies". These fields dictate what information about the person needs to be gathered and how it will be structured. In the context of data portability, understanding and correctly implementing these classes and fields is crucial for ensuring that data not only moves between systems but does so in a way that the information remains coherent and retains its intended use.

4. Data heterogeneity

Classes, that is the conceptual models we use to define objects in software, are not universally applicable, but tailored to specific systems. Each software developer determines the most appropriate class to meet their application's unique requirements. Consequently, there is no one-size-fits-all "user" class that works across all applications. However, applications with similar functions—like messaging apps, email clients, or photo galleries—might share somewhat similar classes due to overlapping needs, functionalities and contexts. For example, consider our BBQ analogy: I specified certain types of food for guests to bring, rather than leaving it open to any party item. This specificity is similar to defining a class in software: you set precise requirements to meet your goals. If I had not specified at least such, we might have ended up with invitees bringing anything, from inflatable balloons to board games—fun, but not edible.

Let us apply this to a practical scenario: imagine you are developing a dating app. You need to create a "UserProfile" class for your users. What attributes would this class need? These attributes define how your app functions and how it serves its users, ensuring it meets the specific needs of the dating platform. In 5 words, what would you include to make your app effective and engaging? Jot them down.

If you have finished with the exercise, let us compare your attributes with the answers from one workshop:

ISSN 2281-5147

- Gender, sexual orientation, descriptions, personal life, age
- Location, picture, job, preference, gender
- Gender, sexual preference(s), age, location (roughly), hobbies
- Gender, ethnicity, city, likes, dislikes
- Preferences, location, gender, age, interests

These examples highlight the fundamental challenge in data portability: even with a specific prompt—like asking for only five attributes to define a user class for a dating app—the resulting classes designed by different developers will vary. Each developer might prioritize different attributes based on their understanding of what is most important for the app's functionality and user experience. This variability underscores a key issue in data portability: the lack of uniformity or, in other terms, *data heterogeneity*.

When it comes to data portability, this lack of uniformity presents significant hurdles. If data is to be portable, it must be easily transferable from one system to another *while retaining its value and functionality*. However, if every system has its own unique set of definitions and structures for what essentially should be the same class of data, transferring data becomes complex. Data that fits perfectly into one application's class structure may not fit as well—or at all—into another's.

This scenario is akin to expecting everyone at our BBQ to bring a dish that fits a specific dietary restriction, without explicitly defining what that diet entails. The results can be as varied as the interpretations of the diet itself, making it difficult to ensure that every dish will be suitable for every guest. In the world of data, this leads to integration challenges, requiring additional transformation or even leading to data loss during the transfer process.

Here are described, in simple terms, the most common data heterogeneity problems:

- Syntax or schematic issue: A guest brings a unique regional dish that looks intriguing but is completely unfamiliar to you. How do you cook it on your grill? Similarly, data syntax—or format—differences mean that even if data is transferred, it might not be in a usable form without some adjustments.

ISSN 2281-5147

- Semantic issue: You asked for a dessert, envisioning pies or cakes, but a French friend brought cheese. In the world of data, this is akin to semantic issues—where the meaning of information varies across systems. What qualifies as "dessert" or "user data" in one system might be broader or narrower in another.

There is however, a third problem. Imagine that, finally, someone brings the side dish that you asked for: provided in a format that fits the grill, it is exactly the dish you asked for, but…it is just a small bowl! Not nearly enough for all your guests. This reflects data *content* issues, where the *volume or completeness* of the data transferred is not adequate for the new system's needs.

To better grasp the *content* issue, there is a beautiful concept borrowed from philosophy and logics called *intensionality*—normally opposed to *extensionality*. Intensionality refers to the essential attributes that define a concept—attributes that are crucial to its identity. If these attributes are absent, the concept itself fundamentally changes. For instance, the definition of a dog includes specific characteristics such as having four legs, fur, and barking. If you were to imagine a dog with wings, this would challenge the conventional definition and identity of a dog—it would not fit our archetype or *intensional* understanding of what a dog is. Intensionality involves those non-substitutable characteristics that are critical to a concept's identity. In the context of data and software, this concept is crucial when considering how data is structured and defined across different systems, especially when dealing with data portability. Ensuring that the essential attributes of data remain consistent and meaningful across different platforms is yet another key challenge, similar to preserving the intensional properties of philosophical concepts.

5. Enough with data. What does "porting" mean?

The general definition of portability, as found in most dictionaries, refers to the quality of being easily carried or moved. Commonly, the attributes that contribute to an object's portability include its mobility (its ability to move), its carry-ability (how easily it can be transported), and the convenience with which these actions—moving and transporting—can be carried out. However, how easily something can be moved or transported depends on various factors. Some items naturally have features that

ISSN 2281-5147

facilitate movement, while others do not. For example, a mountain is immovable and untransportable. But what about water? Water can be moved and transported in some quantities, but if you try to carry it in a pasta strainer, you will find it impossible due to the container's unsuitability.

Using a water bottle, on the other hand, makes transporting water straightforward. So, portability of some contents might depend on the carrier. Yet, consider a 10-liter water tank: while it is designed to be moved and be carried, it would not be considered portable if a 90-year-old had to transport it. Thus, portability also depends on who is doing the moving. Now think about a gun, which is small and light enough to be easily carried by an elderly person. Can one bring to the office? In most cases, no—there are legal restrictions that prevent such items from being brought into certain spaces.

This illustrates that the ease of transportation and carrying is context-dependent, influenced by factors like the person transporting the item, the start and end points of the journey, and the specific conditions under which the transportation occurs, including legal constraints. These elements all significantly affect the practicality of moving and carrying an object.

Since we can generally agree that easy movability and carry-ability are the core characteristics of portability, and considering our discussion on intensionality, we can identify these qualities as the indispensable attributes of the "portability class". Now, it is time to explore whether the existing laws on portability align with this understanding.

6. EU Portability laws

6.1 Personal data portability

To address data portability challenges, the EU adopted (2016) and implemented (2018) the General Data Protection Regulation (GDPR), which was initially conceived in response to privacy issues posed by social networks, and so introduced a critical new right: the right to data portability. This right empowers all individuals to take control of their digital personal information. Essentially, it allows you to receive your personal data—like photos, conversations, and posts—from platforms such as

ISSN 2281-5147

Facebook in a way that machines understand, and transfer it to another service provider, or to request that Facebook directly transfer this data to another provider.

The GDPR mandates that this data be received in a structured, machine-readable, and commonly-used format to facilitate transfer. This shows that the creators of the GDPR were well aware of the syntactic data heterogeneity issues—that data can be formatted very differently across platforms—hence the requirement for data to be in such specifically generic format. However, here is where we encounter a significant gap: the regulation, while precise about the format, does not address the semantics (the meaning and context of the data) or the content (the completeness and detail of the data). If you have familiarized with the technical concepts explained earlier, you are probably already spotting a problem there.

Consider a picture you uploaded on Facebook: if the data model of that photo had 30 attributes, under GDPR, it could be transferred with as few as five, as long as the format is structured, commonly used, and machine-readable. So, while the data's format and perhaps even the semantics of logs or metadata might comply with GDPR standards, the content might be insufficient if important attributes needed at the destination are omitted. This situation highlights a critical shortfall of GDPR: while it facilitates the transfer of data, it does not ensure that all the necessary information—crucial for the data's utility in its new location—is transferred *effectively*.

6.2 Portability of data from "Gatekeepers"

Six years after the GDPR came into force, the goal of data portability remains largely unfulfilled. Many users, perhaps like yourself, were not even aware of its existence and, without consumer demand, no market for alternative services developed. Yet, the need to port data has grown increasingly critical to achieving the European Commission's Data Strategy and create a Digital single Market. Recognizing this importance, more regulations have been introduced, notably the Digital Markets Act (DMA) of 2022 and the Data Act of 2023. These laws build on the GDPR's concept of data portability, but introduce some key modifications.

The DMA specifically targets large platforms, such as Meta,with their social networks Facebook, Instagram and messaging service WhatsApp. They are referred to as "gatekeepers" and the DMA is to rebalance the market asymmetries they created by putting into law that "with great power comes great responsibility –and further

ISSN 2281-5147

obligations." Under the DMA, gatekeepers that control personal or company data must provide *access* to it. This access must be in a machine-readable, commonly-used and structured format (similar to GDPR), and be provided in real-time and continuously, addressing a timeliness issue that the GDPR did not consider. However, a critical element is still missing that is, the necessary, un-substitutable minimal content of the ported data. This is to say that the minimum content required, reflecting our earlier discussion on intensionality, is once again not specified.

Additionally, there is an interesting shift in terminology from the GDPR to the DMA. The GDPR required data controllers to "transmit" data, while the DMA requires gatekeepers to provide "access" to data. Consider the difference between having food delivered to your home (transmission) and dining in at a restaurant (access). While the end result—eating prepared food—may be the same, the process and experience are quite different. In digital terms, the design of a data portability system that allows for such an exchange of information differs significantly depending on the direction of the flow of data, whereby considerations like data security, authorization mechanisms, logging, latency, and more come into play.

Furthermore, if we consider the intensional characteristics of data portability as discussed with the GDPR, and apply the restaurant analogy, is the data really being moved? Does it need to be carried? If we think it does, then by such definitions, even a mountain could be considered portable! This raises fundamental questions about what data portability truly means and how it should be implemented to effectively serve both users and the market.

Connected products' data portability

If you have devices like a smart fridge, smart washing machine, a car connected to the internet, or a smart speaker like Amazon Alexa at home, you are a part of the vast network known as the Internet of Things (IoT). These connected products communicate and share vast amounts of data about their operation and usage. Since you contribute to generating this data, the Data Act (DA) is designed to ensure you can access and utilize this information, derived from your interactions with these products and their associated services.

Consider a scenario where a traditional fridge breaks down. Previously, a handyman would need to inspect only the physical hardware to diagnose and fix the issue.

ISSN 2281-5147

However, with smart appliances, faults could be software-related, necessitating access to operational data to understand what is wrong. Additionally, this data can power other services, like an app that monitors your household's energy consumption by accessing data from your various smart devices.

Under the DA, the entity holding this data—whether it is the manufacturer or another party—must make it accessible to you. This requirement echoes the approach of the Digital Markets Act (DMA), but with a notable twist: making data available upon request is akin to directing you to where your meal is prepared in a restaurant, rather than delivering it directly to your home—there is no actual movement or carry-ability involved.

There are four key points to note about the DA:

1. The IoT data must include relevant metadata,[1] which is essential for interpreting and using the data effectively. This inclusion addresses the challenges of both semantic (meaning and context) and syntactic (format and structure) data heterogeneity.

2. The data must be easily accessible, directly mentioning "easiness".

3. The quality of the data provided must match what is available to the data holder. However, this does not imply equal quantity. The data must be in a format that is structured, commonly-used, machine-readable…and comprehensive!

4. The mention of comprehensiveness might be the first hint at the required content of the shared data. This suggests that the data fields collected from a device, such as a fridge, must include all necessary attributes to make them actionable by another user or system, like a smart meter app or a technician fixing the appliance. However, the DA is particularly focused on enabling access to the *raw data* collected by smart sensors in real time and continuously, as it aims to ensure that the data can be effectively utilized in practical secondary applications. But as raw data has not been processed and formatted, the formatting issue takes the back seat.

Ultimately, it can be argued that not even the DA has a generalizable answer to the problem of content, or intensionality, in the data models, as the "portable" (meaning,

ISSN 2281-5147

accessible) connected data is all, and the same raw data being collected at the source in real time.

7. Applying technical notions to analyze laws

Now that we have ventured into the world of computer science, let us be Legality Attentive Data Scientists (LeADS) that is, let us delve into a meta-analysis of data portability as outlined in the three regulations using the concepts of Class, Object, and Data Field. If we were to conceptualize a Class for Data Portability—essentially capturing the essence of what data portability entails—what would be the essential fields and attributes that define it?

Starting with the GDPR: to model the Data Portability class, a critical data field we require is the "format". This field must meet specific conditions: it needs to be machine-readable, commonly used, and structured. Additionally, the model must facilitate the transmission of data by the data controller and its reception by the data subject—recalling our food analogy, this situation is akin to home delivery: you order the BBQ, and it is brought directly to your door.

Now, let us examine the DMA. You might expect the DMA's approach to data portability to mirror that of the GDPR, right? Thus, the fundamental data fields should remain unchanged as they encapsulate the necessary requirements. However, what you find is that while the format remains the same, the methods of transmission and reception are replaced by the concept of access at the controller's location. Essentially, the gatekeeper (akin to the restaurant in our analogy) allows you to come in and pick up your BBQ.

Lastly, under the Data Act, the focus shifts to the data holder making information available to you. Here again, we see an adjustment in the class's fields rather than just the attributes.

What does this signify? Typically, once a class is defined (in this case, the concept of data portability), within the objects (GDPRportability, DMAportability, DAportability), the data fields are expected to remain consistent while the attributes might vary. However, if the fields themselves are changing, this indicates a fundamental change in the class. Consequently, if GDPRportability version is

ISSN 2281-5147

considered true data portability, it is logical to conclude that the versions under the DMA and DA may not be—given their divergent approaches to how data is accessed and handled. This analysis suggests a broader, more complex landscape of data portability where the core idea may shift based on legislative context and technological needs.

What we have learned about the application of classes and objects, complete with data fields (which set the model) and attributes (which provide specific answers in an object of that model), is that they can effectively represent just about anything. This conceptual framework has proven particularly helpful when analyzing the concept of data portability. In our LeADS-style examination, we have treated data portability as a class within various legal frameworks, utilizing the normative descriptions provided by each to identify the essential data fields that define this class.

Our findings reveal that in different legislative acts, not only are the specific attributes of data portability varied—as one might typically expect—but the data fields themselves also differ. This indicates that the very concept of data portability is not uniformly understood across different regulations. The paradox here is profound: the laws designed to resolve issues related to intensionality (the essential characteristics that define a concept's identity) are themselves plagued by intensionality issues.

8. Effects of laws to systems and technological design

When I wrapped up one workshop, someone professionally involved in implementing the Digital Markets Act (DMA) approached me with a crucial question: "So what?", they asked. They pointed out that whether through transmission and reception or simply providing access, consumers ultimately gain access to their data in both scenarios. So, what is all the fuss about?

It is a valid observation, but there is a subtle, yet profound difference. The essence of the right is not merely about accessing data; it is about the ability to move data from one place to another in order to enable switching providers. Imagine if we were discussing money instead of data: in such case, you would understand immediately the significant difference between transferring your funds from one bank to another versus merely having one bank allow another access to view your funds. It is about control—how, when, and by whom it can be exercised.

ISSN 2281-5147

There is also a less practical, but fundamental difference. Under the GDPR, the concept of data portability is rooted in the idea that individuals should have control over their information. This control allows individuals not just to access, but to physically relocate their data, asserting control and authority over its use, and interrupting other's control if they so wished.

However, in the DMA and the DA, this control is conceptualized differently. In DMA and DA the rationale of data portability is enabling data to move around—actually: be accessed and used—in the internal market, while the interest of the individual to control data is secondary with respect to third parties to access the data. This shift might seem minor, but it alters the dynamic of control and underscores a different interpretation of what it means to "port" data, as well as a shift from porting that is beneficial to the individual to porting that is beneficial to the market, or society.

Finally, it is entirely legitimate for different regulations to define data portability in their own ways—just as different software systems might have their own definitions and requirements. There is not a one-size-fits-all "Universal Data Portability Class"; each regulation can and does establish its own parameters, much like individual software solutions tailored to specific needs.

This diversity however, while flexible, introduces complexities similar to those encountered in software integration, particularly concerning data heterogeneity. Systems engineers and software developers must understand these distinctions deeply. They need to decide how to architect their systems: Should their system be capable of sending information at a user's request in a universally compatible format? Or should it facilitate a system where other systems can make such requests? The answers to these questions are crucial, shaping how effectively these systems can serve their intended purposes and comply with varying regulatory expectations. And these systems' designs, as they are the means through which data portability rights (by the way, a *fundamental right* under EU law) will be exercised, will foster individualistic or utilitaristic views of informational self-determination.

9. Dependencies and Competition

The vision of a Digital Single Market for the European Union is formed on the seamless flow, sharing, and reusability of data. However, as we saw, the reality is

ISSN 2281-5147

complicated by significant data heterogeneity issues that demand a strategic level of coordination. This coordination can manifest in two primary ways: data coordination can happen at the source, in which case data shared and pooled adheres to a standardized format, using a unified vocabulary, and is appropriate and timely enough for reuse across different systems. The most famous case of data standardization is perhaps that of health data, where specific formats (FHIR from HL7), semantics (ICD-11 from World Health Organisation) and content are required to participants in the health data space. This approach represents a tight-coupling integration strategy, which is more centralized and ensures consistency and standardization from the onset. Conversely, in a lack of coordination scenario, the burden of adaptation falls on the data recipients, who must contend with data in whatever form it arrives, often leading to compatibility issues. This represents a loose-coupling integration strategy, which is decentralized and varies greatly in effectiveness.

These two approaches sit at opposite ends of a spectrum that spans from tightly integrated to increasingly looser integration strategies. While, theoretically, establishing a new digital market from scratch might simplify the decision on which strategy to follow, the practical landscape is much more complex. Currently, the vast majority of data is controlled by a few major platforms, formatted primarily to meet their specific needs. Prior to regulations like the GDPR the DMA and the DA, which mandate to different levels data sharing, these platforms had little to no incentives to share their data, let alone making them interoperable with other systems. In fact, their strategies often aimed to maintain a *de facto* monopoly by limiting data interoperability.

Addressing these challenges now is complex. With a handful of dominant data sources and potentially millions diverse receivers, choosing between tight and loose coupling strategies hinges on practical feasibility. Historically, loose coupling has proved less effective, suggesting a need for moving towards tighter integration. However, this raises critical questions about governance:

- Issue of decision authority: Who determines the formats and content of shared data? Leaving this solely in the hands of the major platforms is problematic. Firstly, it benefits these platforms as they continue their operations without needing to adjust their systems, thus maintaining market dominance. Secondly, depending on the technologies used for data sharing, these platforms might

ISSN 2281-5147

gain undue competitive advantage by accessing information about the data receivers, especially if those receivers are also competitors.

- Issue of dependence and competition: If major platforms dictate data formats without restrictions, every data receiver becomes wholly dependent on these formats. This could lead to a situation where a sudden change in format by the data sources could disrupt or even halt the operations of numerous businesses and organizations that rely on this data. Moreover, even the market based on a specific data source might be molded dependently on the model decided by the private actors.

In summary, while advancing towards a more coordinated approach appears necessary, it also intensifies the need for equitable governance in the digital data marketplace, ensuring that no single entity holds too much power over the entire ecosystem.

10. Conclusions

In conclusion, data portability might seem like a straightforward concept—after all, many of us switch mobile providers or use cloud services without a second thought. However, the reality is far more complex and its significance extends across various fields including law, technology, and economics, marking its fundamental role in our digital society.

The BBQ analogy serves well to illustrate the matter: just as a dish that does not fit the grill or match the meal plan can disrupt a gathering, data that is not immediately usable when transferred between different systems due to compatibility issues interferes with consumers' freedom and disrupts the digital market. It is not just about moving data; it is about ensuring it remains useful and meaningful in its new context.

Moreover, regulations like the GDPR, DMA, and the Data Act have been stepping stones towards better data portability, but there is still a lot to think-and-do about. These efforts show the necessity for a rounded approach that addresses the mechanics of data transfer, as well as the meanings and completeness of the data itself.

Ultimately, enhancing data portability will involve more than just technological fixes; it requires a holistic strategy that integrates legal, economic, and technical

ISSN 2281-5147

perspectives. A Legality Attentive Data Scientist approach, which sees legal issues through technical lenses, can be such useful tool to discover problems hiding between the bordering folds of law and technology. This approach will not only boost user control over their data but also foster competition and drive innovation in the digital marketplace.

ISSN 2281-5147

ISSN 2281-5147

SELF-SOVEREIGN IDENTITY: THE REVOLUTION IN DIGITAL IDENTITY

Cristian Lepore*

Abstract

Digital identity is important for businesses and governments to grow. When apps or websites ask us to create a new digital identity or log in using a big platform, we do not know what happens to our data. That is why experts and governments are working on creating a safe and trustworthy digital identity. This identity would let anyone file taxes, rent a car, or prove their financial income easily and privately. This new digital identity is called Self-Sovereign Identity (SSI). In our work, we propose an SSI-based model to evaluate different identity options and we then prove our model value on the European identity framework.

Table of Contents

*Cristian is a cybersecurity fellow and ESR at the University Paul Sabatier. He is the primary maintainer of the SecTeal compiler and the original architect of the AuthclicK application. In 2020, Cristian focused his activity on formal models and digital identity. In 2021, he assessed the self-sovereign identity European framework (eIDAS) to figure out possible challenges and solutions. Previously, Cristian worked in the industry as a cryptographic engineer and infrastructure analyst. In the LeADS project, he will gain the critical thinking to lead the next wave of innovation and push his research forward

.This work is supported by the European Union's funded project Legality Attentive Data Scientists (LeADS) under Grant Agreement no. 956562.

ISSN 2281-5147

Keywords

Digital identity - Self-sovereign identity – Security – eIDAS - European ID

1. Introduction

Digital identity is crucial for businesses and governments because it helps build trust with customers and citizens (Camp, 2004). In our connected world, we often need to create new online accounts or log into different websites and apps. This raises important questions about how our personal information is managed and kept safe (Ansaroudi et al., 2023). It's completely normal to be concerned about the risk of our data being misused or stolen.

To tackle these issues, experts in the industry and governments are teaming up to develop secure and trustworthy digital identities for people. This new approach, known as Self-Sovereign Identity (SSI), represents a major step forward in how we think about identity (Satybaldy et al., 2020). With this approach, individuals can better protect their personal information and have more control over how their data is managed.

With Self-Sovereign Identity, people can easily and privately show things like tax statements, rent cars, or prove their income without giving away extra personal information (et al Alvaro Martin, 2019). For instance, when applying for a rental car, a person can present their driver's license and insurance information without needing to disclose their home address or date of birth. Similarly, when applying for a loan, they can share their verified income statement directly from their financial institution

ISSN 2281-5147

without exposing unnecessary personal details, like their full social security number or banking history. This approach empowers individuals to control their data and share only what is necessary, enhancing both privacy and security.

Currently, there are no clear rules to guide experts in creating and distinguishing Self-Sovereign Identity solutions, making it challenging to assess the impact of new technologies. For example, does a mobile wallet truly empower individuals to control their identities? What specific mechanisms allow users to have control over their personal data and identity? What factors determine whether an SSI system is truly secure?

To answer these questions, we began with a simple explanation of Self-Sovereign Identity, supported by practical examples (Section 3), and outlined the European Union's efforts in developing a secure identity system for citizens (Section 4). This was supplemented with easily understandable examples, along with a proposed approach to evaluate Self-Sovereign Identity (Section 5). Additionally, we created an image that illustrates our approach and described its components using everyday objects. In the future, clear rules would be highly beneficial for evaluating solutions proposed by private companies and governments.

2. What is digital identity

The Internet is like a massive web connecting computers and smartphones around the globe, enabling people to communicate, share information, and explore contents. From catching up on the latest news and watching entertaining videos to connecting with friends on social media, the Internet has transformed the way we live and interact. Nowadays, it is hard to imagine a day without searching for answers to questions or shopping online from the comfort of our homes.

In its early days, the Internet was a playground for a small community of academics and researchers who trusted one another (Johnson Jeyakumar et al., 2022). Because of this, security was not a major concern; they believed that everyone online had good intentions. However, as the Internet grew and became a central part of everyday life, its landscape changed dramatically. Today, nearly everyone owns a smartphone and relies on the Internet for everything from work and communication to entertainment.

ISSN 2281-5147

Unfortunately, this increase in Internet use has also led to serious security issues. Problems like online fraud, identity theft, and hacking of social media accounts are now common threats that can affect anyone, from individuals to large businesses ('A Brief History of the Internet', 2023). These issues can lead to stolen personal information, financial loss, and a feeling of vulnerability in the online world. In summary, while the Internet offers incredible opportunities for connection and information, it also poses risks that we must navigate carefully. To address these challenges, the concept of digital identity has emerged as a potential solution.

Digital identity is a term that refers to how you present yourself in the online world (Davie et al., 2019). It encompasses everything from basic personal information – like your name, date of birth, and email address – to the way you behave and interact with others on the internet. Think of it as your online persona, which follows you around whenever you visit websites or use apps on your phone (*Digital Identity in the ICT Ecosystem*, 2023).

To better frame the concept, let's describe the (online) identity in a way that's easy to understand, using the example of setting up an Instagram account, a platform where people connect by sharing images, videos, and stories from their lives. Imagine you decide to join Instagram to share your experiences with friends and family. The first step is to create an account, which is like establishing your identity in the digital world (Commission, 2023). You start by filling out some personal information, which helps define who you are online. This includes picking a unique username, entering your full name, and providing your email address. Next, you create a secure password to protect your account. Now your profile is beginning to take shape. This profile serves as your digital identity on Instagram. You can upload a profile picture that represents you – maybe a fun snapshot from a recent trip or a casual selfie. Additionally, you can write a short bio that shares a bit about yourself, like "Adventure seeker and photography enthusiast." This bio helps others understand your interests immediately. Once your profile is set up, it becomes your digital business card, showcasing who you are and what you love. You start posting content – photos of your travels, snapshots of your daily life, or even videos of special moments. Each piece of content you share, along with captions and hashtags, contributes to building your online identity. For example, if you post a picture from a recent mountain hiking trip, your friends can react to it by liking or commenting on the photo. This

ISSN 2281-5147

interaction not only enhances your online identity but also fosters a sense of community, connecting you with others who share similar interests. Your Instagram profile is more than just a collection of photos; it represents your personality, passions, and experiences in the digital landscape. Over time, as you engage with others and share more content, your digital identity evolves, reflecting the unique story of who you are and how you connect with the world.

In addition to regular posts, you also use Instagram Stories, which allow you to share more casual and fleeting moments. Stories disappear after 24 hours and offer a glimpse into different aspects of your life, like having coffee with friends or enjoying a breathtaking view. These daily interactions enrich your digital identity, making you more accessible and relatable. As you start following other users, you build a social network. Each time you interact with their content – whether through comments or direct messages – you contribute to a mutually connected environment. If you follow many travel accounts, Instagram will suggest similar profiles and show you relevant content, personalizing your experience on the platform.

In summary, your experience on Instagram extends beyond sharing photos; it reflects your identity, passions, and relationships. Every post, comment, and interaction help shape the overall image of who you are in the digital world, creating bonds and connections that go beyond the screen.

As of today, in many situations, people's identities are managed by the government or other authorized organizations. This means that governments have the power to control how a person's identity is used in society. This situation can feel like a form of "hostage-taking," because individuals are dependent on the rules set by these authorities to access services, travel, or participate in certain activities. For example, governments decide who gets a passport, who is allowed to vote, and who can access healthcare, all based on verified identity information. In this way, people's identities become a tool that the government uses to control access to rights and opportunities.

Many people today have little control over how their personal information is handled online. Decisions about what happens with their data are often made by companies or organizations without clearly explaining it to the public. This lack of transparency has led to the development of a new approach to managing personal identity information, designed to give individuals more power over their own data. This concept allows people to take charge of their identity information, deciding what to

ISSN 2281-5147

share, with whom, and under what conditions. By doing so, they gain more control and privacy in the digital world. One of the most advanced models in this area is called Self-Sovereign Identity (SSI) (Laatikainen et al., 2021). SSI is built on the idea that individuals, not intermediaries, should own and manage their personal data. It empowers users to securely store their identity information and share it only, when necessary, without needing a central authority to approve or manage their actions (Soltani et al., 2021).

3. Self-Sovereign Identity

Self-Sovereign Identity (SSI) is a new way of thinking about how people manage and control their personal information in the digital world. Traditionally, our personal data – like our name, address, or online profiles – is stored and controlled by large organizations like social media companies, banks, or government institutions (Ehrlich et al., 2021). We trust these intermediaries to keep our data safe, but they often have access to more information than necessary and could be vulnerable to breaches or misuse. SSI turns this model upside down. With SSI, you own your digital identity, just like you own your passport or ID card in the real world. You get to decide who sees your information, what details you want to share, and for how long (Ruff, 2018).

Let's take the example of Giulia, a European citizen who needs to apply for a tourist visa for an international trip. Traditionally, Giulia would have to collect and send several physical or digital documents – like her passport, bank statements, and proof of residence – via email or mail to the embassy. This process can be time-consuming, potentially risky, and involves sharing more personal information than necessary.

In an SSI system, things work differently and much more smoothly. Giulia has a unique digital identity, which is stored in a digital wallet (an application) on her phone or computer. In this digital wallet, Giulia holds verified credentials such as her passport information, proof of her financial stability from her bank, and government certifications that prove her nationality.

When Giulia applies for a visa, instead of sending all her documents, she simply shares the relevant details directly from her wallet. For example, she can allow the embassy to verify her nationality and passport details, as well as her financial status, without showing unnecessary information like her home address or full banking history. Giulia

ISSN 2281-5147

decides exactly what information to share and for how long the embassy can access it. Once the visa process is complete, she can easily revoke access to her data, ensuring that her personal information is not unnecessarily exposed for longer than needed.

The security of this system is much higher because the information Giulia shares has already been verified by trusted authorities like her government or bank, significantly reducing the chances of fraud or document forgery. Also, since everything is handled through her digital wallet, the need to send documents back and forth or navigate cumbersome bureaucratic procedures is eliminated. The process is quicker, more private, and more secure.

But SSI is not limited to visa applications. Giulia could use the same wallet to prove her identity when opening a new bank account, sign a job contract, or demonstrate her qualifications when applying for a new position. The system allows her to control her data, share only what's necessary, and ensure her privacy is respected throughout different interactions – all through a single, easy-to-use digital platform.

In Europe, there are several initiatives aimed at improving digital identity systems, with governments, universities, and businesses working together to make it happen (Sharif et al., 2022). These programs are designed to give citizens more security and control over their personal information, while also making digital services more transparent and reliable for everyone.

4. The European study case

In the 2010s, the European Union (EU) took a leading role in exploring how a regional online identity system could benefit its citizens. This led to the creation of eIDAS (Electronic Identification, Authentication and Trust Services), a groundbreaking initiative aimed at providing all Europeans with a secure digital identity (Susanna, 2022). The idea behind eIDAS is to give people a reliable way to prove their identity online, much like using a passport or ID card in the physical world. However, the journey to achieving this has been far from simple.

The eIDAS system officially came into effect in 2016 (Commission, 2016). Initially, each EU country had a lot of control over how they managed their citizens' digital

ISSN 2281-5147

identities. While this seemed like a flexible approach, it resulted in uneven progress across the region. By the late 2010s, fewer than half of Europeans had access to a usable electronic identity, which limited the effectiveness of the system (Sharif et al., 2022).

It was in 2020 that the EU recognized the need for a more unified approach. This realization led to a major shift: instead of each country working independently, the EU began pushing for a single, standardized digital identity system across all member states. This system would be based on Self-Sovereign Identity (SSI).

A revision of the original regulation, in February 2022, led to the introduction of the Architecture Reference Framework (ARF) (Commission, 2023), a blueprint for how this unified digital identity should work. The first draft of this framework came out in February 2023, and discussions continue as it evolves.

As of today, several technical and legal documents guide the development of this European digital identity system. The key component is a "digital wallet" that will serve as a personal online ID for European citizens. It is expected to be fully operational by the end of 2026.

When finished, the system will offer a major convenience to citizens across Europe. For example, a person could use their digital identity to access services or request official documents (like civil registration records) online, even while living or traveling abroad, without having to physically return to their home country. This marks a big step toward seamless digital integration within the EU, making life simpler for millions of people across the region.

Let's explore how the European digital identity system, introduced by eIDAS, works and how it can benefit people in everyday situations. Marco, an Italian citizen, is eager to continue his studies and has found an interesting course at a university in Spain. To enroll, he uses the European digital identity system provided by eIDAS, which simplifies the entire process.

Instead of filling out lengthy forms and dealing with paperwork, Marco logs onto the Spanish university's website. He notices an option to use his European digital identity for the enrollment process. This system allows people across the EU to securely verify

ISSN 2281-5147

their identity and share necessary personal data. Using his smartphone, Marco opens his digital wallet – an app that holds his ID and important documents. He selects his identity and begins the authentication process, which uses biometric recognition (like his fingerprint) to confirm it is really him.

Once verified, Marco is presented with a list of personal information the university needs for enrollment, such as his name, date of birth, and details of his high school diploma. With just a few taps, he selects the relevant information and gives his consent to share it with the university. The system automatically fills out the enrollment form for him, saving Marco from manually entering his details, making the process faster and more efficient.

Additionally, Marco needs to submit supporting documents, such as proof of residency and recommendation letters. Conveniently, these documents are already stored in his digital wallet. Instead of scanning and uploading them separately, Marco can easily attach the required files from his digital wallet directly to the application.

After submitting everything, Marco receives instant confirmation both via email and on his digital wallet app. The system also lets him track his application status in real time, keeping him informed of any updates. If the university needs further details, they can request the information through the platform. Marco can respond quickly, knowing that his personal data remains safe and secure, thanks to the privacy protections built into the eIDAS system.

Once his enrollment is accepted, Marco is notified and can finalize his registration online. Again, using his digital identity, he signs any required documents, without the need for printing or mailing anything. This entire process – from authentication to document submission and signature – is secure, convenient, and timesaving, allowing Marco to focus on preparing for his studies in Spain.

Thanks to eIDAS, the entire enrollment process is smooth and efficient. Marco has not only simplified the university enrollment procedure but also gained greater control over his personal data.

ISSN 2281-5147

4.1 Limitations

While the eIDAS system offers many advantages for Marco, there are several issues that need to be addressed.

One of the problems is the slow pace at which some countries have adopted the eIDAS system. Some countries have fallen behind, and this discrepancy can cause confusion in the use of cross-border digital identities (*Study to Support the Impact Assessment for the Revision of the eIDAS Regulation | Shaping Europe's Digital Future*, 2021). For example, Marco wants to use his Italian digital ID card to open a bank account in Spain, but he finds that the Spanish bank has not yet integrated the system for European recognition. In fact, many banks in Spain do not accept digital identities from other countries. As a result, Marco is forced to use paper documents and go through a lengthy and complicated registration process.

In other cases, both countries may support the European digital identity system, but there are issues with the certification of identity providers. For example, Maria, a German citizen, has just learned about the Lissi digital identity service (*Interact with European Digital Identity Wallets According to eIDAS 2.*). She decides to register to take advantage of the possibility to access other online services, not only in Germany but also in other European countries. After a few months, Maria applies for a scholarship in France. She wants to use her German identity card, but the French online service cannot accept Maria's document. This is because Lissi does not meet the security requirements required by France. Maria is confused, as she thought her new digital identity would allow her to use French services as well. Now, not only does Maria need paper documents, but she also has to travel to France in person to sign the documents.

To solve these issues, a single certification recognized at the European level is necessary. This would allow a digital identity service to be accepted in another part of Europe, thereby increasing citizens' trust. However, from this point of view, eIDAS leaves countries too much room for maneuver, slowing down the harmonization process.

Finally, the adoption of the system by the public and companies is still limited. Many citizens and businesses still harbor doubts and uncertainties about digital identities, mainly due to a lack of understanding and concerns about privacy. Imagine Maria

ISSN 2281-5147

wants to use an online service that requires authentication through a digital identity. When registering, she is offered the option of using eIDAS but is skeptical, fearing that her information could be compromised. Similarly, Sophie, the owner of a local business, wants to digitize her system but is reluctant to use eIDAS for the same reasons. Without proper training and information on the benefits of eIDAS, both Maria and Sophie decide not to use it, continuing to prefer traditional methods. This lack of trust hinders the adoption of eIDAS, limiting opportunities for citizens and businesses to access more modern and efficient services.

In conclusion, while the eIDAS system presents significant advantages for users like Marco and Maria, its effectiveness is hindered by notable shortcomings, particularly the lack of harmonization among EU member states.

5. Building the future of identity

Implementing eIDAS and self-sovereign identity requires significant effort, both technically and legally, to ensure that all the different online accounts we use, such as social media and banking services, can work together seamlessly. Additionally, each European country will need to recognize identities from other countries. For example, a French account must be recognized in Germany and vice versa. Each country will propose its own wallet based on different technical implementations. At this point, we still do not know how many wallets will coexist, potentially more than 27. This means that various implementations of national identity will also coexist.

Suppose you want to link your Instagram profile to a European identity, making it easier and safer to access various online services with a single identity recognized everywhere. You might choose to use a French digital wallet. This wallet could not only contain your official documents, such as your ID card or driver's license, but also your social media accounts and other services. This wallet must be recognized by other countries, and we still do not know the best existing identity solution.

In this context, our goal is to create a common set of rules to assess which of the many existing solutions is truly self-sovereign. This common set of rules can also be used to facilitate the integration of existing accounts with new online accounts.

ISSN 2281-5147

- *What we do*

To explain our work, we use a simple analogy by referring to a concept familiar to us. Imagine a messy desk full of pens, papers, and other objects. If you receive a phone call and need to jot down notes quickly, finding a pen amidst the chaos becomes complicated. Now, picture the advantages of an organized desk. You can instantly find what you need. Imagine that everything is sorted and stored in labeled boxes. That streamlines your work. To make the labels easily distinguishable, you might choose to use simple geometric symbols. For example, a triangle could indicate the box for pens, a square could represent printer cartridges, and so on.

This way of cataloging items becomes critically important in the digital world, where technologies evolve rapidly.

- *Our approach*

The various online accounts we use, such as banks, corporate email, etc., metaphorically represent the objects on our desk. First, we organize these accounts and the related technologies into a model called Trust Over IP. This model, developed a few years ago by a non-profit organization, aims to guide experts in designing new technologies.

Imagine being an influencer who wants to create an account on Instagram. Now, imagine you can link a digital version of your passport or ID card to your Instagram account. This would allow anyone to verify that the account truly belongs to you. Once your account is linked to this digital identity, your name becomes proof of the account's verification. The information is securely stored and cannot be altered or forged. In short, mapping an Instagram account within the Trust Over IP framework means linking the account's digital identity to a verified structure.

With our work so far, what we have is a very precise description of an account. We could describe an Instagram account using labels. For example, the first step would be called "Creating a Digital Identity," to which we would assign a specific label. The collection of all labels would describe our online account.

ISSN 2281-5147

At this point, we want to provide a universal description of our accounts. That is, we want to be able to describe every existing account in a simple and clear way for everyone. For instance, we'd like all websites to easily indicate the procedure for creating an account, or for deleting it, etc. To do this, we have described the objects using a common language for everyone. A sort of universal language. This language is called ontology and can be used to describe Self-Sovereign Identity (SSI). Therefore, our ontology becomes a universal description of our digital accounts. The ontology becomes a kind of guideline for creating, deleting, and modifying our online accounts, regardless whether it is a bank account or an Instagram account.

Finally, the ontology can be associated with a definition of SSI. At this regard, we elaborated a new definition of SSI from past works. The result is a list of properties that focus on individual's privacy and protection of information. Figure 1 shows an overview of our research. The elements are represented with rectangles and squares in the figure. The right side represents the different users' digital accounts. Trust Over IP is represented in the center as a rectangular box. Our descriptive language is on the left. The arrows indicate the processes of associating elements between the rectangles.

- *The outcome*

The output of our approach is to be able to evaluate any digital identity system. Citizen in Europe will be aware of what nation will propose the best identity based on control of information.

ISSN 2281-5147

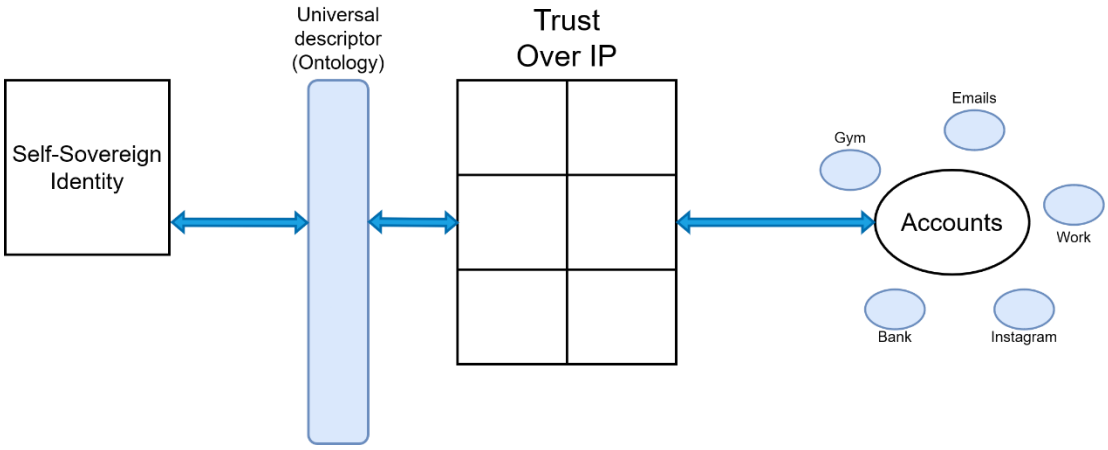


Figure 1. A schematic overview of the research work.

6. Conclusion

In conclusion, the adoption of a self-sovereign digital identity system, such as Self-Sovereign Identity (SSI), represents a crucial step towards a future where individuals will have full control over their personal data. The example of the European Union with the eIDAS system demonstrates that the integration of such technologies can simplify processes like authentication, access to public services, and the management of personal information, ensuring greater security and privacy. However, challenges remain, including the lack of standards and uniformity across various existing online accounts. This applies to both private accounts and solutions proposed at the European level by different countries. This slows down the adoption of new technologies by countries.

To build a more secure and connected future, it is essential to continue working on harmonizing regulations and educating citizens about the benefits of digital identities. Our approach is aimed at evaluating existing identity systems. It supports the collective effort of governments, businesses, and users to create a truly autonomous and universal digital system capable of ensuring privacy and data security in all online interactions.

ISSN 2281-5147

References

A Brief History of the Internet. (2023, September 29). *Internet Society*. https://www.internetsociety.org/internet/history-internet/brief-history-internet/

Ansaroudi, Z. E., Carbone, R., Sciarretta, G., & Ranise, S. (2023). *Control is Nothing Without Trust a First Look into Digital Identity Wallet Trends*. 113–132.

Banabilah, S., Aloqaily, M., Alsayed, E., Malik, N., & Jararweh, Y. (2022). Federated learning review: Fundamentals, enabling technologies, and future applications. *Information Processing & Management*, *59*(6), 103061.

Camp, J. (2004). Digital identity. *IEEE Technology and Society Magazine*, *23*(3), 34–41.

Commission, E. (n.d.). *DIRECTIVE OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL. Regulation Proposal 2021/0136 (COD)*.

Commission, E. (2023). *ARF - architecture reference framework. January 2023*.

Davie, M., Gisolfi, D., Hardman, D., Jordan, J., O'Donnell, D., & Reed, D. (2019). The trust over ip stack. *IEEE Communications Standards Magazine*, *3*(4), 46–51.

Digital identity in the ICT ecosystem: An overview. (2023, October 23). ITU. https://www.itu.int:443/en/publications/ITU-D/Pages/publications.aspx

Ehrlich, T., Richter, D., Meisel, M., & Anke, J. (2021). Self-Sovereign Identity als Grundlage für universell einsetzbare digitale Identitäten. *HMD Praxis Der Wirtschaftsinformatik*, *58*(2), 247–270. https://doi.org/10.1365/s40702-021-00711-5

et al Alvaro Martin, A. I. S. (2019). Digital Identity: The current state of affairs. *BBVA Research*, 1–46.

Interact with European Digital Identity Wallets according to eIDAS 2. (n.d.). Retrieved 16 October 2024, from https://www.lissi.id/

Johnson Jeyakumar, I. H., Chadwick, D. W., & Kubach, M. (2022). A novel approach to establish trust in verifiable credential issuers in Self-sovereign identity ecosystems using TRAIN. *Open Identity Summit 2022*.

ISSN 2281-5147

Laatikainen, G., Kolehmainen, T., & Abrahamsson, P. (2021). *Self-sovereign identity ecosystems: Benefits and challenges*. Scandinavian Conference on Information Systems.

Ruff, T. (2018). *7 Myths of Self-Sovereign Identity*. https://medium.com/@timothy.ruff/67aea7416b1

Satybaldy, A., Ferdous, M. S., & Nowostawski, M. (2024). A Taxonomy of Challenges for Self-Sovereign Identity Systems. *IEEE Access*.

Sharif, A., Ranzi, M., Carbone, R., Sciarretta, G., Marino, F. A., & Ranise, S. (2022). The eIDAS Regulation: A Survey of Technological Trends for European Electronic Identity Schemes. *Applied Sciences*, *12*(24), 12679.

Soltani, R., Nguyen, U. T., & An, A. (2021). A survey of self-sovereign identity ecosystem. *Security and Communication Networks*, *2021*, 1–26.

Study to support the impact assessment for the revision of the eIDAS regulation | Shaping Europe's digital future. (2021, June 9). https://digital-strategy.ec.europa.eu/en/library/study-support-impact-assessment-revision-eidas-regulation

Susanna, T. (2022). *Revision of the eIDAS Regulation: Findings on its implementation and application*.

Young, K. (2010, May 27). *The Identity Spectrum*. Identity Woman. https://identitywoman.net/the-identity-spectrum/

ISSN 2281-5147

ISSN 2281-5147

EVALUATION AND HARMONIZATION OF DATA QUALITY CRITERIA: INSIGHTS FROM EXPERT INTERVIEWS FOR LEGAL APPLICATION

Louis Sahi*

Abstract

This article presents a framework for assessing data quality, highlighting its growing importance in today's data-driven organisations. With the emergence of regulatory frameworks such as the GDPR and the EU's Open Data Regulation, the need for robust data quality standards has never been more important. The study begins by addressing the inconsistencies in current data quality criteria (DQCs) and proposes a unified list, derived from an extensive literature review. By aligning these standards with the broader context of data processing, including governance and lifecycle management, the research aims to create a coherent approach to data quality. Expert interviews were conducted with data management and legal professionals to validate the framework. This involvement not only consolidates the DQCs, but also ensures their compliance with EU regulations. The findings underline the need for collaborative data processing (CDP) in decentralised environments, such as the European Common Data Spaces, and highlight the importance of trust, legal compliance and reliability of shared data. Ultimately, this research contributes to bridging the gap between academic methodologies and practical industrial applications of data quality assessment, fostering a more secure and efficient data landscape.

*Louis Sahi, IRIT, Université de Toulouse, UT3, Toulouse, France, sahilouis@gmail.com
Louis graduated in Cybersecurity (Master) from Côte d'Ivoire. He has a bachelor's degree in Network and Computer Sciences and experience in the industry in Cote d'Ivoire and Morocco. This work is supported by the European Union's funded project Legality Attentive Data Scientists (LeADS) under Grant Agreement no. 956562.

ISSN 2281-5147

Table of Contents

ISSN 2281-5147

Keywords

Data Quality Criteria – Collaborative Data Processing – Trust– Reliability – Compliance

1. Introduction: Unlocking the Potential of Data, Ensuring Quality for Competitive Advantage

In today's economy, data is a valuable asset that is rapidly being integrated into business processes across all sectors. For data-driven organisations, data isn't just useful, it's essential for developing innovative strategies and products that ensure competitive advantage (Hupperz et al. 2021). By collecting and analysing data, these organisations are able to make more informed decisions and respond efficiently to market changes (Fabijan et al. 2017). However, effective use of data remains a challenge, often due to the quality of the data itself.

The Cost of Poor Data Quality

Despite significant investments in data infrastructure, many organisations face problems due to poor data quality. This is more than just a technical issue: data quality affects productivity, decision-making and even financial performance (Haug, Zachariassen, et van Liempd 2011). As Gartner's Ted Friedman highlighted in 2018, organisations that go digital risk a crisis of trust in data, which can reduce business value and harm financial outcomes (Moore 2018). Reliable, high-quality data is critical to maintaining trust in an organisation's information systems, particularly in light of regulatory requirements around data management.

ISSN 2281-5147

How to Define Data Quality?

The question of what makes data 'data quality' remains complex. To address this, we conducted a systematic review (Sahi et al. 2023) to explore academic perspectives on data quality criteria (DQC). Our findings show that there is no universal set of data quality standards, with different authors emphasizing different criteria and sometimes using inconsistent terms and definitions. From this literature review, we finally identified 30 essential DQCs and proposed a single definition for each, with the aim of standardising the assessment of data quality across systems. While this list is a valuable foundation, it needs to be validated in practice to bridge the gap between theory and practice. Furthermore, data quality standards need to be aligned with evolving regulations, such as the GDPR and the EU Open Data Regulation, highlighting the need for a compliance-focused approach to data quality.

Working With Data Management Experts

To ensure that our criteria are applicable in industry, we worked with data management experts, including legal experts, to review and assess the relevance of each DQC within the European regulatory framework. This collaboration provided a multi-faceted understanding of data quality and strengthened the practical relevance of our framework. Overall, this research highlights the importance of high quality data as the backbone of successful data-driven organisations. A systematic approach to data quality, informed by both academic research and practical insights, is essential for organisations seeking to remain competitive and compliant in an increasingly data-centric world.

Structure of the Article

The rest of this article is structured as follows. Section 2 outlines the motivation and contributions of the expert interviews. Section 3 describes interview preparation and participant selection. Section 4 presents the background and data management expertise of each interviewee. Section 5 summarises the feedback, definitions and

ISSN 2281-5147

relevance of the DQCs. Finally, section 6 provides a comprehensive analysis of the findings.

2. Motivation & Contribution: Ensuring Data Quality for Effective Data-Driven Decision Making and Compliance

2.1 Motivation: Contributing to Compliance in the European Data Space

With the European Union's initiative to establish *European Common Data Spaces*, ensuring the quality of shared data has become increasingly important. In a data space, data is managed at its source and shared only when needed, involving different stakeholders such as data providers, intermediaries and users. This collaborative model, known as collaborative data processing (CDP), spans the entire data lifecycle and emphasises community-driven interactions between users and systems (Gan et al. 2017).

However, CDP introduces specific challenges, such as:

1. **Compliance with Legal Requirements**: Data quality assurance strategies need to be aligned with regulatory frameworks, such as GDPR, to ensure secure and responsible data sharing.
2. **Trust in Decentralized Governance**: Building trust within a distributed system is essential for reliable data use and sharing.
3. **Reliability of Distributed Systems**: Ensuring that data remains accurate and consistent across a distributed infrastructure is critical.

2.2 The Challenge: Lack of Standardization in Data Quality Criteria

Many surveys and research studies have proposed various Data Quality Criteria (DQCs), each focusing on specific domains such as health information, information security and business performance. However, these studies vary in scope and context, resulting in different sets of DQCs - each with unique terminology, interpretations

ISSN 2281-5147

and criteria definitions. Our previous review highlighted a major issue: *there is no universal set of DQCs that can be applied across all domains*. In response, we conducted a systematic review (Sahi et al. 2023) to propose a comprehensive set of DQCs that can be applied across domains. However, bridging the gap between academic research and industry practice requires validation by data management experts from different sectors. In addition, collaborative data processing brings its own challenges. Ensuring data reliability in distributed systems, establishing trust within decentralised governance, and complying with legal requirements are essential for successful collaborative data processing (Jonathan et al. 2017).

2.3 Long term objective: Towards automated data quality assessment

A key goal of our work is to enable automated data quality assessment through standardised and unified DQCs. Automated assessment uses algorithms to evaluate data quality, eliminating the need for manual oversight. However, without consistent and universal DQCs, this vision remains out of reach. Previous research has classified DQCs based on data dimensions (Foote 2022), but few have explored the contextual aspects of data processing. Our framework aims to address this gap by including broader aspects such as data lifecycle, governance and regulatory requirements. This approach will provide a solid foundation for future automation efforts and improve data reliability, trustworthiness, and compliance across multiple domains.

2.4 Long Approach: Gathering Expert Insights to Refine DQCs

These challenges raise important questions about how DQCs can be formalised and enforced within a regulatory framework. While regulatory requirements vary by data type and use, no current framework analyses DQCs and data management regulations, such as GDPR or Open Data, together. This study fills this gap by developing a framework that aligns DQCs with regulatory requirements, with a focus on the evolving European data environment.

To address these challenges, we worked with data management experts with expertise in regulatory data governance. Through interviews, we gathered insights into relevant DQCs and refined our framework to align with key EU regulations. This study presents the findings of these regulatory data experts and identifies DQCs that incorporate critical regulatory requirements, ensuring that data quality in collaborative

ISSN 2281-5147

spaces meets both operational and regulatory standards. This research is a step towards a universal data quality framework that will serve as a foundation for future efforts to automate data quality assessment and compliance in Europe's collaborative data ecosystems.

3. Methodology: Consolidating Data Quality Insights from Legal Data Experts

To gather expert opinion on data quality criteria (DQCs), we conducted semi-structured interviews with data professionals specialising in data law and compliance. This process allowed us to consolidate their insights into a validated set of criteria. Here's how we approached this research:

1. **Formulating Relevant Questions**
2. **Selecting Qualified Data Professionals**
3. **Conducting the Interviews**
4. **Analyzing Results**

3.1 Step 1: Developing Key Questions

Our aim was to validate and refine a list of 30 DQCs, focusing on the relevance, thedefinition and the impact on trust, reliability and compliance of each criterion. To conduct effective interviews, we designed open-ended, semi-structured questions to encourage thoughtful, experience-based feedback. Key questions included:

- What aspects of data management have you explored?
- Have you encountered legal challenges in data management, like privacy or compliance issues?
- What kind of data do you work with, and how is it processed?
- What defines a "quality" dataset in your view?
- For each DQC, we asked:

ISSN 2281-5147

- How would you define this criterion?
- How should it be evaluated?
- Does it enhance trust, reliability, or legal compliance?

3.2 Step 2: Selecting Data Experts with Legal Backgrounds

The interviews were hosted by a partner organisation involved in the LeADS project. The Security and Technology Policy Director of the host organisation helped us to select participants based on their expertise and relevance to the study. Five professionals from different European organisations with expertise in data management, privacy and compliance were chosen:

- **Interviewee 1:** A Senior Legal Counsel focused on ensuring global data protection compliance within a major financial organization.
- **Interviewee 2:** A privacy expert in a semiconductor company, overseeing the processing of telemetry data to ensure regulatory adherence.
- **Interviewee 3:** Chief Officer of Privacy Protection in a consumer goods company, advising on GDPR compliance and privacy strategies.
- **Interviewee 4:** An academic specializing in data quality and governance, with a focus on regulatory challenges in AI and data integrity.

3.3 Step 3: Conducting the Interviews

The interviews were conducted via Webex between February 7 and 22, 2024. Each session, which lasted between 30 minutes and one hour, was recorded and supported by handwritten notes. Transcriptions were made to assist in the analysis, while maintaining strict confidentiality.

3.4 Step 4: Analysis and Results

The insights gathered from these professionals have been systematically analysed to identify key themes, validate the DQCs and align them with the practical compliance

ISSN 2281-5147

and governance requirements of the European regulatory landscape. This feedback is invaluable in refining a DQC framework that supports effective and compliant data management across multiple sectors. By consolidating expert feedback, this study advances the development of universally applicable data quality standards that prioritise trust, reliability and compliance.

4. Insights and Findings from Legal and Privacy Data Experts on Data Quality

4.1 Insights from Interviewee 1: Senior Legal Counsel in Financial Privacy

Our first expert, with a background in financial privacy, emphasised that high quality data is precisely tailored to its purpose and must be rigorously governed and protected. This respondent identified critical attributes such as granularity, relevance and security, as well as the importance of governing data through tagging and taxonomy for reliable use. Key DQCs identified include:

1. **Appropriate amount of data:** Data must contain the necessary attributes for accurate use, ensuring no excessive details that might clutter its intended purpose.
2. **Governance:** Effective governance is essential to ensure high-quality data, including protocols for maintaining integrity and usability.
3. **Understandability:** Data should be clearly labeled and tagged to facilitate accurate interpretation across contexts.
4. **Consistency:** Data should remain uniform across platforms to be correctly interpreted by all users.
5. **Currency:** The data is up-to-date.
6. **Timeliness:** Data must be available, accessible, and usable within required timeframes, ensuring timely actions and decisions.

ISSN 2281-5147

7. **Uniqueness:** Avoiding redundancy and duplication, ensuring each data element is unique and specific.
8. **Ease of manipulation:** High-quality data should be reusable and adaptable for various purposes without compromising its integrity.
9. **Free of error:** Data must be accurate and reliable, without distortions that might mislead its users.
10. **Integrity:** Data must be secured against unauthorized access, ensuring only designated individuals can access and modify it.
11. **Interpretability:** Data should be interpretable in ways that allow for multiple perspectives and uses.

These DQCs emphasize the importance of governance and clarity in data to support financial operations while safeguarding privacy and integrity.

4.2 Insights from Interviewee 2: Privacy and Legal Compliance Expert in Semiconductor Manufacturing

The second expert, from a semiconductor manufacturing background, highlighted the need for standardised data references across teams to avoid misunderstandings and compliance risks. This expert emphasised the value of metadata and advocated adaptable data frameworks that can accommodate evolving privacy regulations. Key DQCs from his perspective include:

1. **Understandability:** Inconsistent data referencing across teams can lead to misunderstandings and compliance risks. Metadata plays a key role in ensuring clarity and consistency, helping teams to use data correctly across different functions.
2. **Authorization:** Only authorized individuals should handle sensitive data, with adherence to legal bases and privacy regulations
3. **Objectivity:** Data should remain impartial and collected consistently to avoid bias, ensuring equal treatment across individuals and contexts.

ISSN 2281-5147

4. **Relevancy:** Data relevancy is crucial for organizations seeking to derive value from their data assets. By ensuring that data outputs align with business objectives, fit the context, and meet customer needs, organizations can enhance decision-making processes and drive successful outcomes (Micheli et al. 2020).
5. **Value-added:** Data should deliver tangible benefits; for example, data from current products should inform improvements in future designs, providing real value.
6. **Communication:** Data must be clear, timely, and accessible only to authorized users, preventing security and privacy breaches.

This feedback emphasizes data consistency, privacy, and relevance within a technologically complex environment where accuracy and compliance are paramount.

4.3 Insights from Interviewee 3: Chief Officer on Privacy Protection

Focusing on data protection in a company that produces a wide range of consumer goods, this expert highlighted the importance of balancing data utility with privacy obligations, particularly in the context of AI applications that require high accuracy. She highlighted some of the challenges posed by Europe's stringent data protection regulations, particularly in relation to the handling of sensitive data. Key DQCs identified by this expert include:

1. **Accuracy:** Data must be correct and reliable, free from errors that could mislead decision-making.
2. **Accessibility:** Data should be easy to locate and accessible only to authorized personnel, ensuring streamlined retrieval processes.
3. **Authorization:** Access rights should be flexible, allowing tailored levels of access based on roles within the organization.
4. **Relevancy:** Particularly in AI, it's crucial to distinguish between relevant data and "noise" to ensure models are built on accurate information.
5. **Objectivity:** Data collection and processing must be unbiased and neutral, avoiding discrimination based on personal characteristics.

ISSN 2281-5147

6. **Ease of manipulation:** The usability of data depends on its format and inherent restrictions; for example, personal data should not be reused for unrelated purposes without consent.
7. **Traceability:** Tracking changes to data is essential, requiring individual accountability and monitoring to ensure transparency and security.

These criteria reflect a strong commitment to both data utility and the safeguarding of personal privacy.

4.4 Insights from Interviewee 4: Researcher in Data Protection Principles

Our final expert, a privacy and data governance researcher, offered a philosophical perspective on the relationship between privacy and data quality. He stressed that privacy includes the right to be unobserved and stressed that individuals do not always want their data to be accurate. His proposed DQCs include:

1. **Safety:** Poor quality data processing of sensitive information can lead to significant risks. A thorough risk analysis is essential to mitigate potential harm, especially under frameworks like GDPR.
2. **Free of error:** This criterion should apply exclusively to factual data, emphasizing the need for objective validation methods.
3. **Reliability:** Ensuring accuracy in data processing and the procedures that support this reliability is critical.
4. **Accuracy:** Data must represent its meaning accurately without bias, maintaining the privacy of individuals while delivering reliable outputs for decision-making.
5. **Value added:** This principle balances legal compliance with the costs of data processing, helping data controllers determine when it is worth processing certain data.

Together, these insights illustrate the intricate balance between data quality, privacy rights, and the ethical considerations inherent in data governance.

ISSN 2281-5147

5. Summary of the Study

5.1 Overview of Analysis

This study explores the perspectives of legal data professionals on key DQCs in light of EU data management regulations. Through the insights gathered from four experienced professionals, we identified 28 key comments highlighting 20 critical DQCs. Notably, 95% of these criteria (19 in total) are consistent with our previously established list. Several DQCs, including **accuracy, authorisation, ease of manipulation, freedom from error, objectivity, relevance, understandability and value added**, were highlighted multiple times, reflecting their importance in the data management landscape..

The responses contributed to a comprehensive classification of DQCs across a generic data lifecycle, confirming the thoroughness of our DQC list and its relevance to current data management practices.

This research aims to enhance the way data controllers should prioritize data quality management and delineate responsibilities to ensure high-quality data outputs. Furthermore, it highlights the connection between data quality management and legal compliance, as a lack of awareness in this domain can lead to significant disadvantages for organizations. The insights provided by our interviewees emphasize the necessity of integrating key EU regulatory points into DQCs.

5.2 Classification of DQCs within a Generic Data Lifecycle

Our primary objective is to assess the trust, reliability and legal compliance of collaborative data processing by DQCs. This analysis places significant emphasis on the legal aspects of data handling in collaborative ecosystems. It facilitates the classification of DQCs within the data lifecycle, enables a clearer assessment of data quality within information systems, and supports the development of automated data quality assessment systems.

Figure 1: Classification of Data Quality Criteria in the data lifecycle

Discussions with the data experts revealed that their responsibilities and expertise in data management correspond to essential steps in the data lifecycle (Shah, Peristeras,

ISSN 2281-5147

et Magnisalis 2021), which include 1) **Collection, 2) Preparation, 3) Analysis, 4) Sharing, and 5) Reuse.** These steps form the backbone of effective data management practices. The insights shared by the experts allowed us to categorize and structure DQCs according to these lifecycle stages, as illustrated in Figure 1.

6. Conclusion

Data quality is a multi-faceted concept that encompasses regulatory compliance, confidence in governance and the reliability of data processing. Achieving these goals requires a thorough understanding of data quality criteria (DQCs). In this study, we have proposed a standardised and unified list of DQCs derived from a comprehensive literature review, aiming to bridge the gap between academic methodologies and practical industrial applications.

As the regulatory landscape for data management continues to evolve in Europe, there is a growing interest in privacy and data protection legislation. This dynamic environment highlights the need for a balanced approach that integrates data quality, system reliability and regulatory compliance. Our findings suggest that trust in data processing can only be established through this fairness.

This paper presents a consolidated framework for assessing data quality, based on insights from open and semi-directive interviews with European data experts from multinational companies. Through these discussions, we refined our initial list of 30 DQCs, identifying the most relevant criteria that align with key points of European data management regulations. We also mapped these DQCs to different stages of the data lifecycle, providing a roadmap for building a trusted, collaborative data processing ecosystem.

7. Future Direction

Looking ahead, our focus will shift to implementing a decentralised, blockchain-based solution for sharing DQCs throughout the data lifecycle. This innovative approach aims to increase the transparency and traceability of data processing activities, enabling all stakeholders to effectively assess data quality. By leveraging such

ISSN 2281-5147

technologies, we can support distributed systems governed by a collaborative governance ecosystem that fosters trust among all data stakeholders.

Through these efforts, we hope to advance the field of data quality management and contribute to a more secure and reliable data processing environment that meets both regulatory requirements and the expectations of all stakeholders involved.

8. Selected Readings

Fabijan, Aleksander, Pavel Dmitriev, Helena Holmström Olsson, et Jan Bosch. 2017. « The Evolution of Continuous Experimentation in Software Product Development: From Data to a Data-Driven Organization at Scale ». P. 770-80 in *2017 IEEE/ACM 39th International Conference on Software Engineering (ICSE)*.

Foote, Keith D. 2022. « Data Quality Dimensions ». *DATAVERSITY*. Consulté 4 octobre 2023 (https://www.dataversity.net/data-quality-dimensions/).

Gan, Wensheng, Jerry Chun-Wei Lin, Han-Chieh Chao, et Justin Zhan. 2017. « Data Mining in Distributed Environment: A Survey ». *WIREs Data Mining and Knowledge Discovery* 7(6):e1216. doi: 10.1002/widm.1216.

Haug, Anders, Frederik Zachariassen, et Dennis van Liempd. 2011. « The Costs of Poor Data Quality ». *Journal of Industrial Engineering and Management (JIEM)* 4(2):168-93. doi: 10.3926/jiem.2011.v4n2.p168-193.

Hupperz, Marius, Inan Gür, Frederik Möller, et Boris Otto. 2021. *What is a Data-Driven Organization?*

Jonathan, Albert, Muhammed Uluyol, Abhishek Chandra, et Jon Weissman. 2017. « Ensuring reliability in geo-distributed edge cloud ». P. 127-32 in *2017 Resilience Week (RWS)*.

Kahn, Beverly K., Diane M. Strong, et Richard Y. Wang. 2002. « Information quality benchmarks: product and service performance ». *Communications of the ACM* 45(4):184-92. doi: 10.1145/505248.506007.

ISSN 2281-5147

Laranjeiro, Nuno, Seyma Nur Soydemir, et Jorge Bernardino. 2015. « A Survey on Data Quality: Classifying Poor Data ». P. 179-88 in *2015 IEEE 21st Pacific Rim International Symposium on Dependable Computing (PRDC)*.

Micheli, Marina, Marisa Ponti, Max Craglia, et Anna Berti Suman. 2020. « Emerging Models of Data Governance in the Age of Datafication ». *Big Data & Society* 7(2):2053951720948087. doi: 10.1177/2053951720948087.

Moore, Susan. 2018. « How To Create A Business Case For Data Quality Improvement ». *Gartner*. Consulté 28 juillet 2023 (https://www.gartner.com/smarterwithgartner/how-to-create-a-business-case-for-data-quality-improvement).

Sahi, Louis, Romain Laborde, Mohamed-Ali Kandi, Michelle Sibilla, Giorgia Macilotti, Benzekri Abdelmalek, et Afonso Ferreira. 2023. « Towards Reliable Collaborative Data Processing Ecosystems: Survey on Data Quality Criteria ». P. 2456-64 in. IEEE Computer Society.

Shah, Syed Iftikhar Hussain, Vassilios Peristeras, et Ioannis Magnisalis. 2021. « DaLiF: a data lifecycle framework for data-driven governments ». *Journal of Big Data* 8(1):89. doi: 10.1186/s40537-021-00481-3.

ISSN 2281-5147

ISSN 2281-5147

EXTRACTING DATA VALUE THROUGH DATA GOVERNANCE

Armend Duzha*

Abstract

Harvesting value from data requires an organization-wide approach. Data governance plays an essential role in a heterogenous environment with multiple entities and complex digital infrastructures, enabling organisations to gain a competitive advantage. This research examines a new approach for data governance developed to extract data value respecting the ever-delicate balance between transparency and privacy. In addition, it provides an overview of the key innovations brought in by novel technologies such as Artificial Intelligence, Federated Learning and Blockchain, and how these can be integrated in a data governance program.

Table of Contents

* Armend is a Marie Skłodowska-Curie fellow working as an Early Stage Researcher at University of Piraeus, Greece within the Legality Attentive Data Scientists (LeADS) project funded under the EU's Horizon 2020 Research and Innovation Framework) under Grant Agreement no. 956562.
aduzha@unipi.gr

ISSN 2281-5147

Keywords

Data value – Data governance – Federated learning – Blockchain – Internet of Things

1. Introduction

Many organizations consider data as one of the most important assets (Kitchin, 2021). By leveraging data processing, they transform data into valuable information and meaningful insight. This can involve performing calculations, applying statistical analyses, or using machine learning (ML) algorithms. The use of internet-connected smart devices, the so-called Internet of Things (IoT), and the adoption of Artificial Intelligence (AI) in social life and daily activities, have significantly enhanced the need for data and the general complexity of digital systems. Personal data is collected from various sources (see Figure 1) such as sensors, mobile applications, social networks, and digital footprint left by online activities, in continuous and extensive ways, resulting in a vast data flow for any product and service in use. This has created a new wave of applications that allows organisations to offer user-centric services in many different sectors such as smart city and mobility, healthcare and well-being, smart manufacturing, and finance. For example, consumers can use IoT to monitor their home security and overall health parameters, while businesses can monitor in real-time their supply chain, track energy spending, and engage in predictive maintenance of their machines.

ISSN 2281-5147

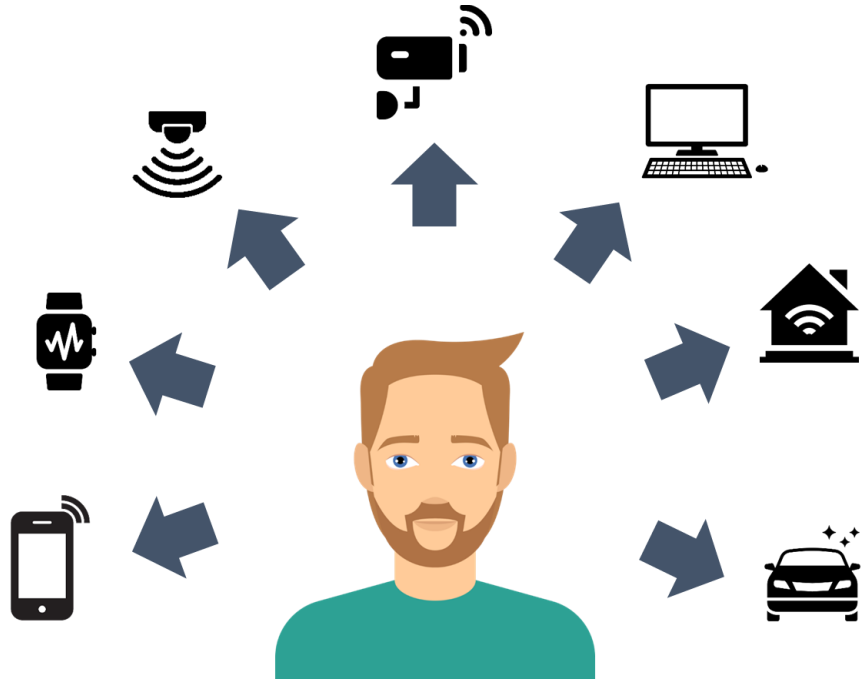


Figure 1: An abstract representation of the Internet of Things

In such a heterogenous environment, data governance plays a crucial role in defining, implementing, and monitoring the context, responsibilities, tools and stakeholders involved throughout the data lifecycle. Therefore, establishing policies, processes and procedures around data and subsequently enacting those to compile and use such data for effective management and decision-making is extremely important. Data governance not only enhances existing products and services but also supports appropriate adjustments during the design and development of new ones.

2. What is Data Governance and why is it important?

The increasing popularity of data governance is closely related with the growing recognition of data value (Zygmuntowski et al, 2021). Before the emergence of AI, data governance primarily focused on the control and management and it constituted a task that was performed mostly by private and large companies. Today, the concept has evolved to encompass "authority" and "control" over the entire data lifecycle with the objective of increasing the data value while minimizing associated risks and costs. In this context, data governance refers to the system of decisions and accountabilities

ISSN 2281-5147

that regulate and guide information-related processes, ensuring adherence to pre-defined policies (Janssen et al, 2020). It outlines permissible activities, assign responsibilities to different entities and actors throughout the value chain, and determine the data to be used, when it can be used, by whom, and for what purposes (Nielsen, 2017).

In the context of AI-based systems, data governance takes a broader role as the system of processes and infrastructures that enable organizations to align AI-enabled technologies with their strategies, objectives, and business values while maximizing the value of data (Mantymaki et al, 2022). Some studies propose the concept of "data governance by design", which facilitate the design of effective data governance frameworks for organizations (Khatri & Brown, 2010).

The emergence of distributed systems, where multiple infrastructures and systems are interconnected, has led to collaborative data processing that involves multiple entities and actors situated in different organisations (Domingue et al, 2019). In this context, decentralised data governance is defined as a set of policies, procedures, and principles that govern the data flows processed by various entities. The decentralised data governance represents a community-based approach for storing, managing, and sharing data, in contrast to a centralized one, where a single entity governs the decision-making throughout the data lifecycle (Greer et al, 2022). Moreover, decentralised data governance ensures compliance with legal, ethical, regulatory, and data protection requirements, specifically the constrains imposed on data processing activities. However, further research to address the challenges posed by decentralisation and distribution are necessary to guarantee that transparency and privacy are not compromised.

3. The Role of Artificial Intelligence

Artificial Intelligence or commonly referred to as AI is one of the dominating trends that affects most industries today, and its impact on data governance is profound. The AI's ability to analyse large datasets fast and efficiently in real-time enables organizations to streamline their data governance practices. Hence, AI solutions can accomplish several tasks such as identification, clustering, classification, and tagging of data, significantly decreasing manual operations.

ISSN 2281-5147

AI-based solutions are adaptable to new data types and sources without requiring extensive re-configuration. Such flexibility is essential in an evolving landscape where both data formats and regulatory requirements keep changing. AI deepens data processing operations by automating existing workflows, making it an indispensable resource in the quest for flexible and scalable data governance.

AI integration into data processing has progressed data governance considerably, thus elevating these tools into intelligent systems capable of autonomous analysis, learning, prediction, and action. With the use of AI, organizations can overcome the complexities of existing digital systems, ensuring their data governance strategies remain effective and responsive to the needs at hand.

As AI systems become increasingly sophisticated, so do the risks associated with data privacy. The same capabilities that make AI powerful also pose significant threats to privacy preservation. To address these challenges, emerging technologies such as Federated Learning and Blockchain are being adopted by the industry.

4. The Power of Federated Learning

Federated Learning (FL) is a recently developed approach for collaborative data processing that is secure and private (Abreha et al, 2022). Let's consider an example from real world to explain how does it work. Google uses FL to build better models for next-word prediction and voice recognition. The company uses the pool of devices (e.g., phones, tables, PCs, watches, and speakers) where Google Assistant is being used. In such a decentralized environment (see Figure 2), user data is stored locally, averting sensitive information from being disclosed with other entities. Instead, each user trains on-device an instance of the ML model, and submits the differences in the parameters to the central server, once the training is completed. The various updates from all the users are then aggregated at the central server, which in return produces an updated global model and distributes it to all users. This iterative process continues until the global model reaches an acceptable level of accuracy or satisfies other criteria defined by Google.

ISSN 2281-5147

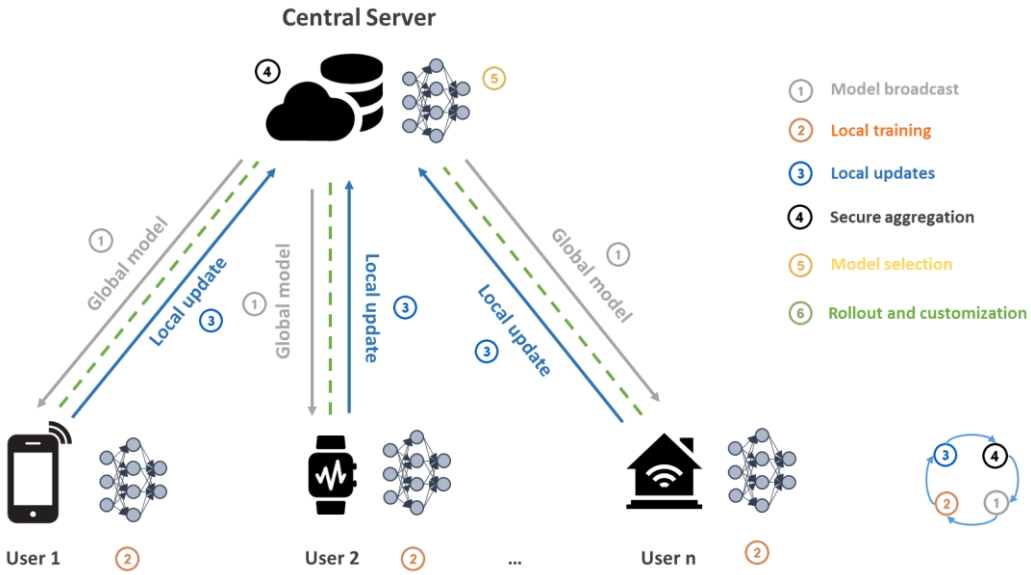


Figure 2: Federated Learning

This inherent structure gives FL the capability to bring together multiple entities in the data processing without the need to disclose any raw data. FL plays a pivotal role in building trust between the involved entities since they can verify at any time that no personal data is exposed, which is crucial for preserving their privacy (Foy et al, 2022).

In terms of resilience, FL enhances the robustness of the system in use in many ways. For instance, it ensures efficient use of data even when network connectivity is poor or one of the users is offline (Bonawitz et al, 2019). Most of the computations happens locally, requiring only intermittent network access to send aggregated model updates. Additionally, FL builds resilience against device failures and data corruption. In our example, if one or more devices goes offline or have corrupted data, the FL process will continue with minimal disruption because it depends on many other devices that continue to work normally, performing local computations. This redundancy makes FL models extremely reliable, ensuring their functionality in any adverse circumstances (Alsamhi et al, 2024).

FL also promotes inclusivity in data processing. By enabling data to remain decentralized, FL allows many data sources as contributors to AI models training without compromising privacy. This can lead to more efficient and potentially more generalized models, since they are trained on a wider variety of data that could not be

ISSN 2281-5147

possible in a centralized setting. Additionally, FL can abide to regulatory requirements, such as data minimisation and storage limitation principles of the GDPR, by limiting sharing of personal data, making it suitable for industries like healthcare, finance, and telecommunications where data privacy is key.

In summary, FL is a powerful methodology for collaborative data processing that enhances privacy, reliability, and transparency. It guarantees that sensitive data are not accessed by third parties, fosters trust among participants, and enables efficient data usage even with limited connectivity. This reliability coupled with FL models' resilience against device failures and data corruption makes it a compelling choice for modern applications.

5. Blockchain: The Backbone of Trust and Transparency

Blockchain, a revolutionary concept that is broadly connected to the use of digital cryptocurrencies like Bitcoin, has turned out to be a powerful tool across different industries ranging from finance to healthcare and beyond (Pilkington, 2016). At its core, it is an immutable distributed ledger that enables data to be secure and tamper-proof, thus forming the core foundation upon which trust can be established.

Its most significant value lies in the ability to produce an irreversible documentation of all transactions and events. Blockchain ensures that once the information is recorded, it cannot be altered or deleted (Zwitter & Hazenberg, 2020). This immutability has profound implications for transparency and trust, since it ensures higher security in data exchanges. Moreover, the very nature of decentralization means that no single entity controls the entire network, instead distributes power across all participants within the blockchain ecosystem, which adds to collaboration and reduce the risk of data misuse.

One of the key benefits of such a decentralised system is that end-users – especially consumers, but also companies – would have much more transparency and control over how their data is used, reclaiming power from big tech and pharma companies that centralise large datasets for competitive benefit.

Healthcare systems in every country and region are struggling with the problem of data siloes, meaning that patients and healthcare providers have an incomplete view

ISSN 2281-5147

of medical histories. In 2016, Johns Hopkins University published research showing that the third leading cause of death in the US was medical errors resulting from poorly coordinated care, such as planned actions not completed as intended or errors of omission in patient records.

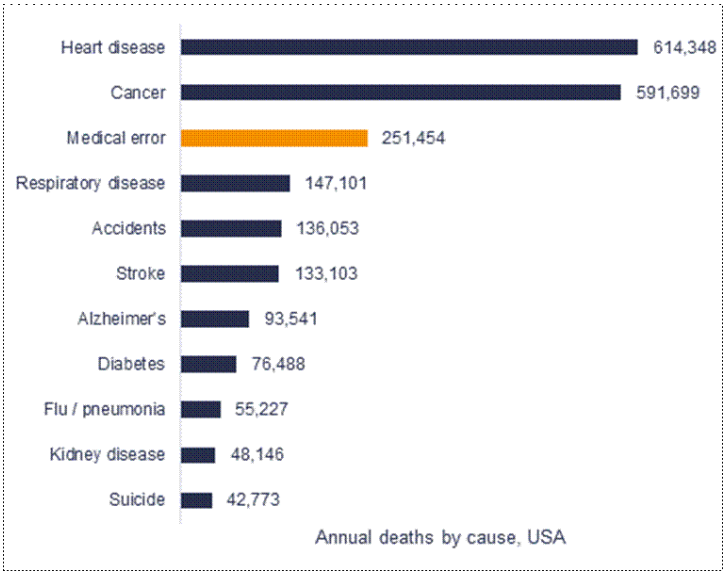


Figure 3: John Hopkins research on medical errors as a proportion of annual deaths in the US, 2016

One potential solution to this problem is creating a blockchain-based system for electronic medical records (EMRs) that can be linked into existing electronic medical record software and act as an overarching, single view of a patient's records. It is crucial to emphasize that actual patient data does not go on the blockchain, but that each new record appended to the blockchain, whether a physician's note, a prescription or a lab result, is translated into a unique hash function – a small string of letters and numbers. Every hash function is unique, and can only be decoded if the person who owns the data – in this case, the patient – gives their consent. In this scenario, every time there is an amendment to a patient medical record, and every time the patient consents to share part of their medical record, it is logged on the blockchain as a unique transaction.

Therefore, in embracing blockchain, organizations are stepping forward towards creating a more trustworthy digital environment. Altogether, blockchain offers unprecedent opportunities to enhance data quality in regards to its integrity, security,

ISSN 2281-5147

and transparency. Due to the distribution of power and record-keeping permanency, it encourages trust among participants in any given sector.

6. Ethical, Legal and Regulatory guidelines

The Ethical, Legal and Regulatory Framework (ELRF) refers to the set of comprehensive rules and practices that should be applied while using AI within organizations. This framework aims to ensure that cutting edge technologies do not violate fundamental rights and values, do not impose bias or discrimination, and are compliant with existing laws, regulations, and ethical standards.

The existing regulatory measures of the European Union include a large number of regulations that tend to facilitate data access and re-use and ensure responsible AI practices. Among others, key regulations include, the *Data Act (DA)* that aims to enhance data re-use by strengthening data sharing and interoperability, the *Data Governance Act (DGA)* that establishes data exchange mechanisms for safe and efficient sharing of data across sectors, and the *AI Act (AIA)*, the European legal framework for AI, which draws standards for just and fair deployment of AI systems. These initiatives are intended to release data currently confined within silos (Patel, 2019) and effectively enforce the data protection framework outlined in the *General Data Protection Regulation (GDPR)*.

More specifically, the AIA seeks to play a fundamental role in the fight against inequality, unfair treatment and infringement of right to privacy by AI systems. By safeguarding individual rights, it aims to create the much-needed trust in these technologies, which is a prerequisite towards its widespread acceptance. The AIA thus introduces a risk-based approach categorizing AI systems into different levels of risk:

- *Minimal risk*: systems that pose low or no risk to rights and safety,
- *Limited risk*: systems requiring transparency obligations,
- *High risk*: systems that significantly impact individual rights and safety, subject to stringent regulations,
- *Unacceptable risk*: systems that are prohibited due to their potential of causing harm that is so severe that is unacceptable.

ISSN 2281-5147

To this end, organizations are required to adapt their strategies to navigate the risk-based framework, ensuring that adopted measures align with the level of potential harm associated with specific AI applications.

The ELRF provides a crucial foundation for the responsible use of AI and Big Data. It ensures that technological advancements initiated by organizations are aligned with societal values and legal requirements. This not only safeguards individual rights but also fosters confidence in AI technologies, paving the way for their broader acceptance and integration into various sectors.

7. The Decentralised Data Governance

Figure 4 presents the high-level architecture of the Decentralised Data Governance framework, which is an evolution of the architecture proposed in the first submission made to AIAI 2022 conference[1]. It has been subsequently improved by integrating the Federated Learning as core component in the data processing layer. Moving outward, it stresses the importance of security, privacy and access control on one side, and ethical, legal and regulatory compliance on the other. These are combined with a blockchain-enabled transactions tracking mechanism, ensuring transparency throughout the system.

[1] Armend Duzha and Dimosthenis Kyriazis, "A Novel Approach for Data Processing and Management in Edge Computing", in 18th International Conference on Artificial Intelligence Applications and Innovations (AIAI 2022), Crete, Greece, June 2022.

ISSN 2281-5147

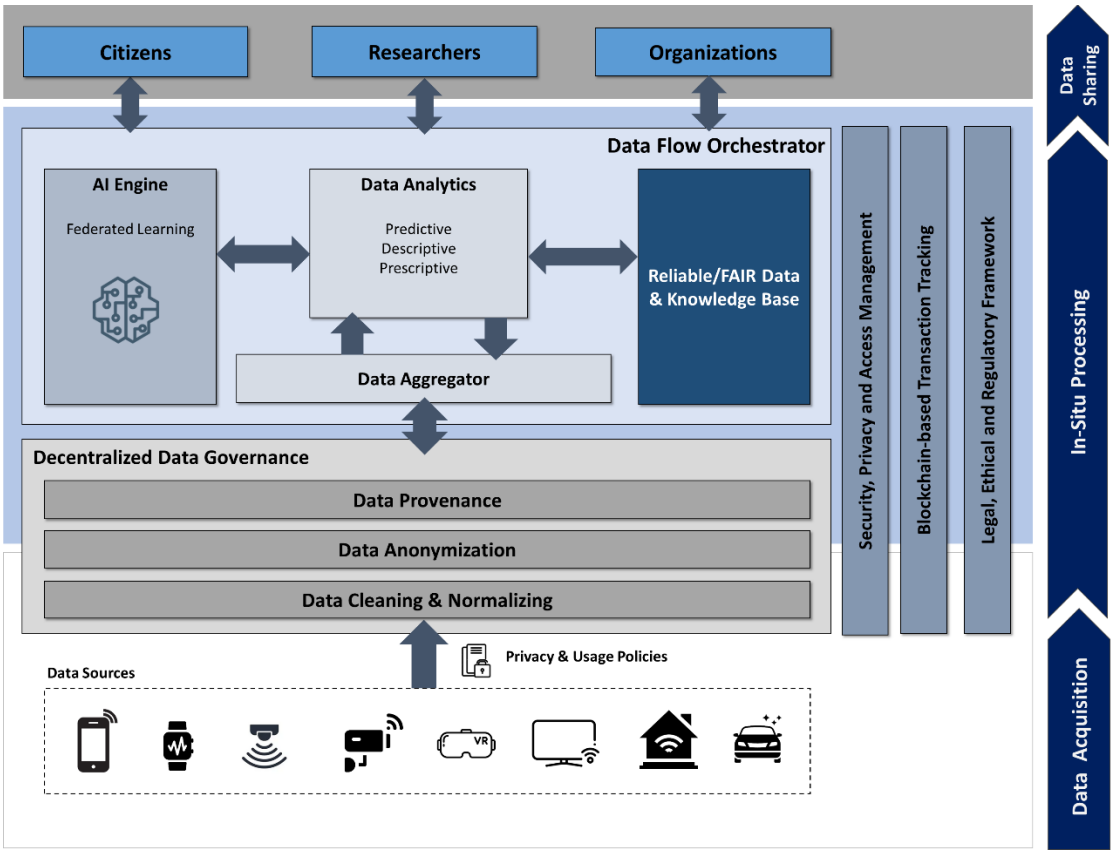


Figure 4: Decentralised Data Governance Framework

The proposed framework consists of three layers: the data layers, the processing layer, and the user layer. The *data layer* is responsible for the preparation of the data to be consumed by the processing layer. This includes cleaning and normalisation so that data is of good quality, as well as an anonymization process before any data activities commences. It follows a decentralised storage approach where each resource remains autonomous. The *processing layer* consists of the AI models and analytics pipelines that process the data derived from the data layer to deliver valuable insights that support informed decisions. Finally, the *user layer* allows to share the processed results and new AI models parameters with data consumers (citizens, researchers and organisations).

The developed framework follows a user-centric approach, empowering and supporting data consumers, be they citizens, researchers, or organisations. It is flexible and adaptable to the needs of the various sectors and aims to maximise the value of data, while minimizing the cost and risks associated with data processing.

ISSN 2281-5147

7.1 Implementation challenges

When initiating a data governance program, organizations may face several challenges that need to be addressed for its successful implementation. Here are some key challenges:

- **Change management and adoption**: persuading business stakeholders to define a data governance program can be a significant obstacle. It requires organisational change efforts that usually involves training, education, and promoting a cultural shift towards data governance practices. Effective communication strategies and stakeholder engagement are crucial for driving adoption and achieving the desired outcomes.

- **Financing**: securing necessary funds can be problematic as it may require determining funding levels for tools to be adopted as part of the program, addressing resource limitations, and understanding how to deliver tangible value. While traditional costs associated with data, such as storage, are relatively quantifiable, assessing its value to the organisation is more complex. The implementation of a decentralised data governance approaches often come with substantial costs and present complex integration and operational activities (Petzold et al, 2020) 0. Moreover, its financing could often be affected by annual variations, making it more challenging to adequately plan it in advance over time.

- **Resource and time constraints**: Data governance activities, such as data collection and definition of data assets, are resource and time-consuming, despite the ability of AI tools to learn data structure and format directly from database structures. Moreover, many organizations nowadays operate in hybrid environments, incorporating both on-premises and cloud-based infrastructures. The distributed data governance requires solutions across multiple platforms and infrastructures, providing consistent governance practices, data policies, and compliance measures.

- **Learning curve, skills and commitment**: Acquiring the necessary skills and the learning curve associated with data governance practices pose time investments and may hinder business adoption, presenting serious commitment challenges.

ISSN 2281-5147

By overcoming these challenges, organizations can unlock the full potential of data governance and realize its benefits in terms of improved data management, enhanced decision-making, and compliance with regulatory requirements.

7.2 Benefits for organisations and individuals

The decentralised data governance can be adopted by several organisation, and more specifically: (i) *private organisations* for planning, designing and implementing data-driven solutions to align their vision and business objectives with their customers' needs; (ii) *public organizations* for organizing, planning, and monitoring the timely provision of appropriate and enhanced policies leading to efficient decision-making and services provided to the citizens; and (iii) *researchers* involved in data processing activities, to facilitate data discovery, interoperability and re-use.

Moreover, data governance can boost the ability of public and private organisations to exploit and monetize their data assets (Ofulue and Benyoucef, 2022), while at the same time developing novel services or enhancing the existing ones.

Finally, data governance changes the approach of assessing the cost related to data, shifting the focus from "How much does processing data cost?" to "How much should our organization invest in data governance?". By adopting a distributed data governance approach, organisations can distribute the implementation risk across the whole data lifecycle. Unlike in-house data processing, distributed data governance offers a flexible support structure that can be adopted to the project's needs, timelines, and requirements.

8. Conclusions and future work

The decentralised data governance represents a necessary step towards responsible and ethical harnessing the value of data. The proposed approach underscores the importance of clear principles, policies, and robust privacy-preservation techniques such as Federated Learning and Blockchain, putting control over how personal data is used into the hands of individuals and organisations. Key legal and normative considerations, such as adherence to the AI Act (AIA) and the General Data Protection Regulation (GDPR), will ensure that data processing within the data lifecycle is conducted legally and ethically.

ISSN 2281-5147

Looking ahead, the prospects that data governance can truly revolutionize the data economy is great. By facilitating secure and transparent data processing, decentralised data governance can drive innovations in different sectors, including healthcare, smart city and mobility, smart manufacturing, and finance. Its potential is huge, although some challenges still remain – respecting the ever-delicate balance between transparency and privacy and making sure that data processing does not jeopardise individual rights. The framework must continuously evolve in view of such problems, adapting to new types and sources of data, emerging technologies, or regulatory changes.

By raising awareness of data value and governance principles, the benefits of the data governance can be realized while safeguarding privacy and data protection. This decentralised approach will help build trust and support for data-driven innovations, ultimately fostering a more responsible and secure data ecosystem.

9. Selected Readings

R. Kitchin, "The Data Revolution: A Critical Analysis of Big Data, Open Data and Data Infrastructures", SAGE Publications, 2021.

J. J. Zygmuntowski, L. Zoboli, and P. F. Nemitz, "Embedding European values in data governance: a case for public data commons", Internet Policy Review, Vol. 10, No. 3, September 2021. DOI: 10.14763/2021.3.1572

M. Janssen et al. "Data governance: Organizing data for trustworthy Artificial Intelligence", Government Information Quarterly, Vol. 37, No. 3, pp. 101493, July 2020. DOI: 10.1016/j.giq.2020.101493.

O. B. Nielsen, "A comprehensive review of Data Governance literature", Selected Papers of the IRIS, Vol. 3, No. 8, 2017. Available at: https://aisel.aisnet.org/iris2017/3.

M. Mantymaki et al. "Defining organizational AI governance", AI and Ethics, February 2022. DOI: 10.1007/s43681-022-00143-x.

V. Khatri and C. V. Brown, "Designing Data Governance", Communications, Vol. 53, No. 1, pp. 148–152, January 2010.

ISSN 2281-5147

J. Domingue, A. Third, and M. Ramachandran, "The FARI-TRADE Framework for Assessing Decentralized Data Solutions," in Companion Proceedings of the 2019 World Wide Web Conference (WWW '19), ACM, New York, NY, USA 2019.

S. L. Greer et al., "Centralizing and Decentralizing Governance in the COVID-19 Pandemic: The Politics of Credit and Blame", Health Policy, Vol. 126, No. 5, May 2022.

H. G. Abreha, M. Hayajneh, and M. A. Serhani, "Federated Learning in Edge Computing: A systematic survey," Sensors, Vol. 22, No. 2, January 2022.

M. Foy, D. Martyn, D. Daly, A. Byrne, C. Aguneche, and R. Brennan, "Blockchain-based governance models for COVID-19 digital health certificates: A legal, technical, ethical and security requirements analysis", Procedia Computer Science, No. 198, pp. 662–669, 2022.

S. H. Alsamhi et al., "Federated Learning meets Blockchain in decentralized data sharing: healthcare use case", IEEE Internet of Things Journal, Vol. 11, No. 11, pp. 19602-19615, April 2024.

Marc Pilkington, "Blockchain technology: principles and applications", Research Handbook on Digital Transformations, pp. 225–253. Edward Elgar Publishing, 2016.

M. Foy, D. Martyn, D. Daly, A. Byrne, C. Aguneche, and R. Brennan, "Blockchain-based governance models for COVID-19 digital health certificates: A legal, technical, ethical and security requirements analysis", Procedia Computer Science, No. 198, pp. 662–669, 2022.

A. Zwitter and J. Hazenberg, "Decentralized network governance: Blockchain technology and the future of regulation", Frontiers in Blockchain, Vol. 3, 2020. DOI: 10.3389/fbloc.2020.00012.

R. Latif, M. U. Ahmed, Sh. Tahir, S. Latif, W. Iqbal and A. Ahmad, "A novel trust management model for edge computing", Complex & Intelligent Systems, Vol. 8, pp. 3747–3763, 2022. DOI: 10.1007/s40747-021-00518-3.

B. Petzold, M. Roggendorf, K. Rowshankish, and C. Sporleder, "Designing data governance that delivers value", McKinsey Digital, June 2020.

ISSN 2281-5147

J. Ofulue and M. Benyoucef, "Data monetization: insights from a technology-enabled literature review and research agenda, Management Review Quarterly, 2022.

ISSN 2281-5147

ISSN 2281-5147

PERSONAL HEALTH INFORMATION MANAGEMENT SYSTEMS (PHIMS) FOR USER EMPOWERMENT: A COMPREHENSIVE OVERVIEW

Christos Magkos*

Abstract

The management of continuously increasing personal health data, in the digital information era, is becoming more and more relevant in modern healthcare. Through integrating raw data in digital platforms, personal health information management systems (PHIMS) could provide a method for the storage, management, and regulation of personal health data access. We examine how PHIMS can empower users to take control of their own healthcare by combining diverse health information sources such as health monitoring devices and electronic health records into a single easily accessible system.

Table of Contents

* Christos joined the LeADS project in January 2022 as an early-stage researcher at University of Piraeus Research Centre (UPRC), Greece. He completed a BSc in Pharmacology and Molecular Genetics in King's College London and then proceeded to study genetics with a focus on computational genetics in University College London for his MSc. His interests lie in ethics and bioethics, data analysis and modelling. Upon undertaking both dry and wet-lab research projects, investigating pathological pain and evolutionary models of cancer, he became increasingly interested in the ethics, regulation and personalisation of data collection in the context of research and online business.
magkos.christos@gmail.com

This work is supported by the European Union's funded project Legality Attentive Data Scientists (LeADS) under Grant Agreement no. 956562.

ISSN 2281-5147

ISSN 2281-5147

Keywords

Health records - Personal data management – PIMS - User empowerment - Personalized healthcare

1. Introduction

1.1 Issues arising from the inflation of medical information

Scientific research paper publications are becoming more available while medical devices and electronic health records amass large amounts of information that requires processing to efficiently be utilized. As not only medical knowledge but also personal health data increases exponentially, the use of this absurd amount of data in an efficient way is one of the most pertinent issues in modern healthcare.

So how can individuals take control of their personal data efficiently? How can physicians and researchers access individual healthcare data to provide them with clinically actionable recommendations while at the same time respecting the individual's right to privacy? How can we gather and organize massive amounts of data in a method that can be explained to patients and be used in daily healthcare practice?

One key issue that can arise, is the method through which an individual can take control of their own data, choose how this data will be used and who has access to said data. Personal Health Information Management Systems (PHIMS) represent a novel approach to health data management in the age of digital information

ISSN 2281-5147

1.2 Our proposal for individualized personal data management

PHIMS can function as a digital folder, serving as integrated repositories where an individual can access all their personal health information through the internet. PHIMS empower individuals to store, manage, and control access to their health data, which can include any type of medical records, medical and biomedical results, and live access, real time measurements from hospital logs, live measuring devices and doctor and self-reported data.

Patient autonomy is a significant pillar of medical ethics and healthcare demands that patients are not only informed in their treatment and monitoring but also play the key deciding role in their own health. Patients engaging in their own healthcare leads to better outcomes for themselves. Furthermore, whether an outcome is beneficial or not, depends on the patient's own view of what is beneficial or detrimental to themselves. Personal involvement and complete information in healthcare leads to higher compliance to treatment and increased patient satisfaction and fulfillment. In that sense, PHIMS can empower users in taking control of their digital health fingerprint by enhancing their ability to be more informed and active in their health decisions and enabling them to have a much more personalized healthcare as we will describe further on.

Hence the principal use of PHIMS is to integrate health information from various sources into a single, easily accessible and intuitive platform. This includes electronic health records, test results, medication use and notes from healthcare providers and doctors alike. A significant source of medical data which PHIMS can integrate data from, is health monitoring digital devices that often have access to the internet, the so called "Internet of Things" (IoT) devices. This can include fitness trackers, smart scales, blood glucose measuring devices and all sorts of biometric data measurements such as heart beats per minute, daily steps, sleep quality variables and even mood tracking.

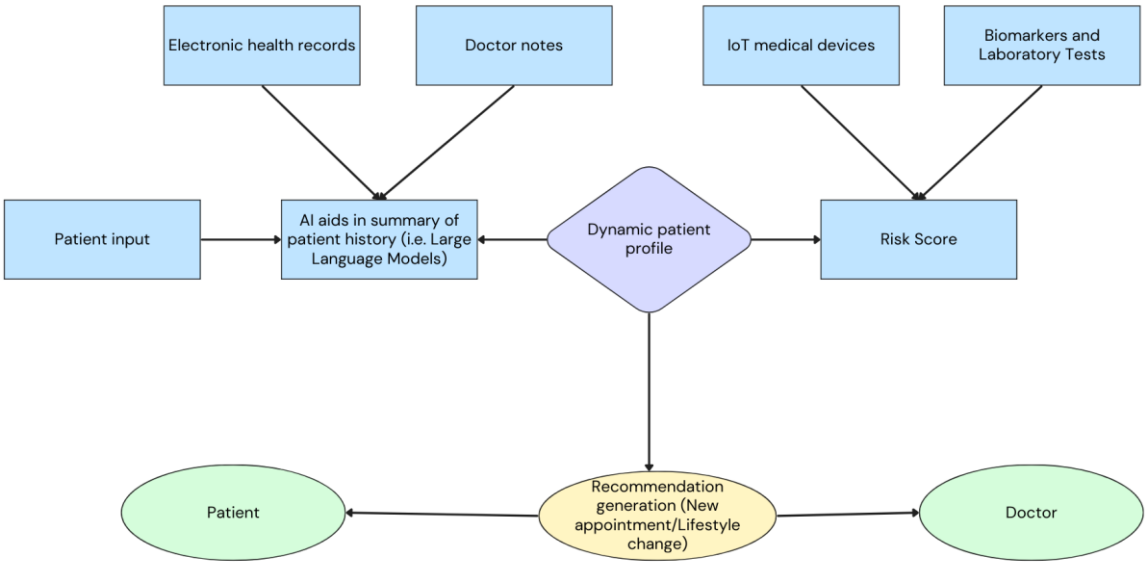
While personalized health tracker apps have been developed, they are rarely integrated with multiple sources of information and accessible to healthcare practitioners. They aim to resolve a single issue such as booking appointments, providing easily accessible mental health resources or monitoring blood glucose for instance. Furthermore, they

ISSN 2281-5147

are not equally easy to use for all population groups while public confusion arises regarding both the usage of and transparency of data accessibility.

Medical apps are being increasingly used but due to the limitations in their scope and use, we attempt to establish frameworks, guidelines and pipelines to define what PHIMS should encompass and how they can be used efficiently, using state of the art data analysis and medical devices and ethically without excluding specific populations.

We present below a flow chart of information flow and data processing through the PHIMS platform that we envision:



2. What value would PHIMS offer to a patient?

2.1 User Empowerment Through Control and Access

PHIMS are being developed due to a need to offer users control over their own medical data. Ensuring autonomy is essential according to medical ethics and PHIMS

ISSN 2281-5147

can be a tool that enables it through data ownership and control. It is up to individuals to determine who has access to their information, monitor the means through which their data is being used, through systems that are built upon security and privacy.

This control extends to sharing information with healthcare providers and practitioners such as doctors, as well as hospitals in order to enable a more comprehensive, targeted and coordinated care. Doctors often have to deal with fragmented and disorganized data, hence through granting doctors access to complete, organized, up-to-date health records and patient histories, PHIMS can aid in the optimisation of patient treatment. At the same time, the user, in this case the patient, is choosing who the information is shared with, as well as what information is shared through the platform to avoid unwarranted sensitive data use and can play a much more active role in the decision-making while avoiding predation of their personal information. Patients may wish their doctor/hospital to be granted access without granting access to insurance companies and pharmaceutical companies they are not affiliated with for instance. At the same time they could select which specific hospitals and doctors have access to what information and which third parties do not.

With all pertinent health data at their fingertips, individuals can form increasingly informed decisions regarding their health as we shift further away from the paternalistic model of healthcare where the doctor was a figure of authority and the patient followed already made decisions. Such a data-driven approach, when accompanied with the correct supervision, can provide better management and hence outlooks for chronic conditions, which are becoming more and more relevant in aging populations, as well as rapid responses to health complications, and prognostic tools that allow for proactive healthcare. Finally, the generation of personalized health recommendations can be enforced, informed by a patient's specific data profile. This is a step towards personalized medicine, a more granular and individualized approach to personal healthcare, better catered to the patient's needs and more accurate and efficient.

2.2 Integration with Health Monitoring Devices

PHIMS can be used as a platform through which to access health information in a form that the patient can comprehend easily as well as integrating existing medical

ISSN 2281-5147

knowledge with inputs from different health monitoring devices, expanding knowledge on health trends and incorporating them in a more holistic treatment.

IoT device networks can play an important role in modern healthcare as the devices used on an everyday basis, have access to the internet. The main benefit of IoT devices in healthcare is their unique ability to allow continuous data collection in real-time while also allowing for live sharing of the information with pertinent healthcare providers.

PHIMS can leverage an ecosystem like this in order to gather extensive health data, providing insights into daily activities and health lifestyles and personalized longitudinal health trends. For instance, a fitness tracker can monitor daily steps and sleep patterns, while a smart scale can track weight and body composition. This continuous flow of data enables more precise and personalized health recommendations. Longitudinal data tracking helps with avoiding inaccuracies due to temporary fluctuations of biological markers due to chance or measurement errors as multiple measurements are taken, and a sense of personalized tracking can be established rather than simply following general medical guidelines, establishing individual trends and seeing what helps the specific person respond.

Key Takeaways:

PHIMS can give patients control over their own personal health data, granting them the ability to determine access to it and how it can potentially be used. More personalized treatments and well-informed health decisions can therefore be made enabling patients to play a more active role in their healthcare. PHIMS may provide real-time data collecting and personalized health recommendations through integration with health monitoring equipment, improving chronic condition management.

3. Practical Applications of Personal Health Management Systems

3.1 Personalized Healthcare and Individualized Treatment

ISSN 2281-5147

PHIMS is in line with the general trend that seems to guide to an extent the future of medicine: personalized healthcare. Every individual's health data can be considered unique, and PHIMS can personally adjust data-driven recommendations and hence guide personalized treatments based on this personal specificity. Personalized medicine approaches healthcare holistically through the consideration of genetic profiling, lifestyle factors, environmental exposures, and family history, for a more effective and accurate planning of health interventions.

Through analysis of this comprehensive medical history, users can receive more comprehensive and clinically actionable outcomes. Using predictive algorithms based on the risk of certain factors, known causes of diseases, and existing "prior" medical knowledge, the accumulation of personal and general medical knowledge can be treated as an optimization problem, meaning that many different lifestyle parameters are being adjusted to optimize personal health preservation and improvement. The final goal always being either recommendations or treatment adjustments, in conjunction with medical supervision.

3.2 PHIMS used to derive Recommendations

What do recommendations entail? Recommendations may encompass dietary changes, exercise plans, and proactive, preventive adaptations regarding a patient's daily lifestyle, tailored to the individual's health profile. A typical example of a disease which requires individualized treatment is diabetes, which is a chronic disorder, allowing for longitudinal data tracking and also displays great interpersonal differences.

Someone with a family history or with high risk of developing diabetes could receive recommendations to monitor blood sugar levels and adopt a ketogenic diet preemptively. Additionally, for current patients, healthcare recommendations for their subtype of diabetes and its outcomes can be applied, and used to generate further information for the future of disease treatment by comprehending the disease at a more granular level and discovering further subtypes.

Through incorporating genetic data to the PHIMS predictive algorithms, we can further improve and customize treatment. Healthcare professionals can then suggest preventive measures and personalize/adjust therapies live. For example, medicine selection can be influenced by pharmacogenomics, which describes the fashion

ISSN 2281-5147

through which a person's genes impact their reaction to medication, in order to reduce side effects and increase effectiveness.

3.3 Advantages for healthcare providers

PHIMS provides healthcare professionals with a more thorough and uniform understanding of a patient's medical history. The detailed data enables the planning of treatment, chronic disease management, and can aid in improving the accuracy of diagnosis.

Complete medical records will, of course, enable health professionals to make better-informed decisions. Clinicians and researchers can utilize historical data in order to observe patterns and evaluate the efficacy of treatment methodologies longitudinallyin order to provide a more holistic approachfor the early detection of possible health issues and more accurate diagnosis.

Continuous health surveillance in patients with chronic diseases in particular is essential for disease monitoring. The PHIMS platform would allow for continuous observation and health parameters tracking, enabling the healthcare professional to follow up the course of the disease and adjust or observe disease management. In this way, it can also support continuous collection of health data concerning risk factors and possible complications.

PHIMS significantly increase the coordination of patient care among various providers. If any given set of health records is available to all those providing care for a patient, there would be effective collaboration amongst professionals. This will eventually result in better outcomes for patients since chances of unnecessary testing, conflicting treatments, and lapses in communication will be decreased.

This collaboration among practitioners would necessitate however digital literacy by doctors and willingness to coordinate care as well as investment in relevant digital infrastructures.

3.4 Public health research

PHIMS could not only provide advantages for individuals but also provide guidance and data to public health agents and larger scale health studies. PHIMS data, after aggregation and anonymization can offer significant insights into population health, disease outbreaks, and treatment efficacy.

ISSN 2281-5147

Through amassing and analyzing large amounts of data gathered by PHIMS, researchers can use PHIMS to identify novel treatment targets and enhance prognosis models. This data-driven research methodology can improve our understanding of a range of medical diseases and hasten the discovery of new medicines. Public health organizations can better respond to public health emergencies and track health trends by utilizing data from PHIMS. For instance, real-time data on flu symptoms could be used to track viral spread. PHIMS's raw data can be transformed into useful insights using epidemiology algorithms and predictive modelling. Researchers can identify higher-risk populations, use analytics for forecasting disease outbreaks, and attempt to create focused interventions by examining trends and patterns. The allocation of resources and public health interventions can both be enhanced by this proactive strategy or at least be evidence-based to a certain extent.

Key Takeaways: By modifying recommendations based on personal health information that takes into account genetic profiles, lifestyle choices, and family health history, PHIMS allow for personalized healthcare. This personal healthcare, when aggregated, can aid in the investigation into public health by providing anonymized datasets for analysis and support the healthcare professionals with detailed patient information.

4. Approach and application: How the framework we propose can help patients

In the following section, we display some examples of healthcare scenarios and use cases for PHIMS in order to gain a better understanding of potential benefits and practical applications of this system from an individual's perspective.

4.1 Handling Chronic Illnesses

The case of chronic illness is a prime example for PHIMS usage. Chronic illness allows researchers to perform longitudinal tracking. It is highly individualized, as no patient is the same, and treatment requires continuous adjustment and monitoring. We believe PHIMS can provide an intelligent solution as a system for tracking disease progression and integrating multi-source data. Below we provide a proposed framework for the management of diabetes as an example.

ISSN 2281-5147

4.1.1 A proposed Diabetes Management Pipeline

Disease tracking

The care of diabetics requires frequent medical visits, regular medication and its monitoring, dietary adjustments, and ongoing blood glucose testing. Through information derived from insulin pumps, meal logging and continuous glucose monitors or CGMs, PHIMS can assist this procedure.

Data Integration: When a diabetic patient wears a CGM, blood glucose measurements are automatically sent to their PHIMS. Their insulin pump and smart dietary app, which monitors carbohydrate intake, are both connected to the system.

Personalized Alerts: If a patient's blood sugar level is abnormally high or low, the PHIMS may evaluate this data in real-time and deliver personalized alerts to both the patient and their healthcare professional, making quick modification of one's diet and medication possible.

Patient History Analytics: As a patient's PHIMS accumulates information longitudinaly, thorough health profiles are generated. This can aid in enabling physicians to determine and extrapolate patterns while modifying treatment regimens accordingly. For instance, dietary or pharmaceutical adjustments may be recommended if the PHIMS registers a trend of elevated glucose readings following meals.

Activity Monitoring: Heart rate monitoring, sleep patterns, and daily steps are all automatically transmitted to the PHIMS using a user's fitness trackers or IoT devices such as smart scales recording body composition and tracking bodyweight. All this information can then be used to monitor activity levels, important for tracking chronic disease progression.

Personalized recommendations: PHIMS can offer personalized fitness recommendations based on longitudinal data. The system could, for instance, recommend and assess new exercise routines and adjust minimum daily step goals upon detection in reduction of physical activity. All these lifestyle and wellness adjustments are non-invasive interventions that improve life quality and longevity in diabetics, but require constant effort and reminders for compliance.

ISSN 2281-5147

Health Metric Dashboard: The user has access to a dashboard that shows their health metrics across time, providing information regarding fitness and identifying lagging health parameters

4.1.2 Integrated Healthcare for Senior Citizens

Elderly patients may oftentimes require care from various different healthcare professionals or specialist teams and it can be difficult to track multiple prescriptions, therapies as well as their extensive medical history. PHIMS can improve results and facilitate better care coordination.

The PHIMS of an older patient would compile information from a range of healthcare professionals, such as specialists, primary care doctors, and pharmacists, aiding them in the coordination and intercommunication of their healthcare needs by accumulating pertinent data in a single repository. Particularly with older patients, who are more likely to present with more comorbidities as well, meticulously tracking and documenting health records is essential and often very difficult given the current decentralized nature of healthcare.

In terms of drug management, the patient can receive reminders from the system which keeps track of drug schedules. Additionally, it can notify medical professionals and carers of any possible drug interactions or missing doses.

Finally, PHIMS can help manage remotely the tracking of comorbidities, which are other diseases interacting with the main disease affecting the patient, very common in older patients. PHIMS can incorporate data from home monitoring devices for individuals with chronic diseases like heart disease or hypertension to track comorbidities and the disorder at the same time. Healthcare professionals can keep an eye on this data remotely and take appropriate actions such as early interventions and lifestyle adjustments.

4.2 Technological advancements and their implementation in personal healthcare

The capabilities of PHIMS will keep growing as technology advances, providing increasingly sophisticated tools for health management. The following developments and trends are anticipated to influence future development of PHIMS.

ISSN 2281-5147

4.2.1 The role of machine learning

Machine learning will be essential for future PHIMS's development. Large-scale health data can be analyzed, which can then be used to spot macroscopic trends to then forecast individual patient outcomes, and provide more precise recommendations.

Predictive analytics can be provided based on machine learning algorithms and as the field advances algorithms assess an increasing number of dimensions. The goal in this case will be to develop models that are able to identify possible health problems long before they worsen and prevent them. For instance, AI can spot early biomarkers of life threatening or life deteriorating conditions like cancer or heart disease by examining patterns in a patient's medical records and medical devices.

4.2.2 Customizing treatment

Personalized medicine is undergoing a revolution. We suggest the integration of genomic data with PHIMS in early stages of risk stratification. Knowing a person's genetic predispositions enables more individualized and efficient medical treatment and along with longitudinal tracking, disease can be monitored and factors affecting both the stressors and the diathesis of the disease can be evaluated. Polygenic risk scores are utilized to guide radiotherapy and cancer interventions as well as certain heart and neurological conditions. However, oftentimes personalized approaches guided by genetics do not present with clinical applications or are not followed in the clinic due to the complexity of implementation. However, larger data and more tracking can actually help identify the cases where personalized approaches are actually effective and aid in the stratification and optimization of therapy for the rest of the population.

Genetic Risk Assessment, which is an estimation of the chance a patient develops a disease based on their DNA can be incorporated in PHIMS, which can assist in determining a person's risk for developing specific diseases. Early interventions and preventative efforts can be guided by this knowledge, given the consent of the individual to be tested.

ISSN 2281-5147

Finally, implementations of pharmacogenomics. Based on a patient's genetic profile, PHIMS can be used as an assistant to healthcare practitioners in the selection of drugs that are more likely to be effective while having fewer side effects to the specific patient.

4.3 Remote patient monitoring and telehealth

PHIMS development can take advantage of the current growth of telehealth and remote patient monitoring. Telehealth refers to the use of communication technologies such as mobile apps, phone and video calls between patient and healthcare providers in order to enable remote healthcare consultations

PHIMS can also assist with virtual consultations, by giving medical professionals access to patient data during telehealth appointments. While difficult to match the effectiveness of remote care and in-person visits, telehealth is a promising tool, expected to equalize to a certain extent the healthcare of those in more remote areas with reduced access to healthcare as well as those with reduced mobility.

4.4 Healthcare communication and integration

Continuous Monitoring: PHIMS could be used to receive real-time data from patient monitoring devices (remotely or in hospital settings), granting medical professionals the ability to monitor patients' health and consider taking appropriate action. Such monitoring can be of particular interest in both post-operative care and the chronic illness management.

Improved Communication/National and International Standards:

Enhancing the compatibility among diverse health information systems is crucial for the extensive implementation of PHIMS. Data interchange will be smoother through standardized communication protocols and data formats development.

Interoperability can be facilitated by the adoption of standardized frameworks like the Fast Healthcare Interoperability Resources (FHIR). These standards facilitate more effective data transmission and communication between various systems.

ISSN 2281-5147

Integration with the Healthcare Ecosystem: PHIMS will progressively interface with electronic health records (EHRs), and public health databases, among other elements of the healthcare ecosystem. Both public health monitoring and care coordination will be improved by this combination.

4.5 A practical example of PHIMS in day-by-day: The case of John

John is a 58-year-old affected by diabetes. How could PHIMS help John in his day-by-day management of his disease?

Day 1: John wakes up and his blood glucose monitor communicates with his PHIMS, sending an alert to his phone regarding his low blood sugar as he has not yet eaten. PHIMS renders a recommendation regarding his breakfast nutritional content. A reminder is sent to him regarding daily exercise, and suggests a morning walk for him to remain fit and manage his body weight. During his morning walk his heart-rate and steps are recorded and his daily exercise recommendations for the following days are adjusted accordingly. Before lunch, his blood glucose is low and the PHIMS reminds him to take his insulin before his next meal. His blood glucose variation after his meal is recorded and the PHIMS registers potential recommendations for John regarding his carbohydrate intake and recommendations towards his doctor regarding his insulin posology.

Day 2: John has a virtual doctor's appointment. His PHIMS sends a reminder to John and asks him whether he wishes to share his latest data and recommendations with his doctor. When he accepts, recent test results, patient history and medication lists are made available to the doctor. Prior to his consultation, John is stressed regarding his health outcomes. As his heart rate rises, John is notified by his PHIMS and realizes he forgot to take his heart medication.

Day 7: At the end of the week John decides to check his health dashboard. His daily health metrics are summarized and his new fitness and health goals are determined by the app. General wellness and fitness recommendations for the following days are also shared with him. John continues improving his activity levels and improves the stability of his blood sugar levels thanks to general recommendations and real time data. Doctor's recommendations based on his health trends are also shared with him depending on the disposal of his doctor.

ISSN 2281-5147

Key Takeaways: PHIMS, which integrate data from several sources and offer personalized alerts, are especially helpful for controlling chronic conditions like diabetes. PHIMS enhances care coordination for senior persons by gathering data from various healthcare providers. The customisation of treatment plans and remote patient monitoring are made possible by technological breakthroughs like machine learning and genetic data integration, which also improve overall healthcare communication and integration

5. Implementation challenges

5.1 Providing Usability and Accessibility

PHIMS must be usable and accessible to everybody, regardless of age, digital literacy, location or financial status, in order to be really revolutionary. In the digital age, ensuring diversity is crucial to preventing health inequalities.

User-friendliness should be considered when designing PHIMS. Features and interfaces need to be simple to use and intuitive. It is important that guidelines regarding universal design principles that ensure equality of access are developed for all possible users. Efficient use of PHIMS can be facilitated by offering educational resources and support, particularly to older adults and individuals with limited technological proficiency.

In order to guarantee that people from all socioeconomic levels may access and benefit from PHIMS, efforts must be made to close the digital gap, the gap between those having access and knowledge of information technologies and those who do not. This can entail supplying reasonably priced devices, enhancing internet connectivity in underprivileged regions, and delivering programmes to raise digital literacy. PHIMS should be culturally aware and support many languages in order to guarantee accessibility. Translation services, culturally appropriate health information, and consideration of various health-related beliefs and behaviors are all included in this. User trust and engagement can both be improved by such inclusivity.

ISSN 2281-5147

5.2 Security and Privacy

Robust privacy safekeeping and security frameworks are necessary due to the nature of health data. Health data is deemed sensitive data, meaning it requires special handling and increased security precautions compared to other types of personal data. In the context of protecting user data, PHIMS must ensure robust, cutting edge and constantly advancing security technologies are used and all data handling performed strictly adheres to the relevant regulations.

There are very specific laws and regulations defining particularly the standards for health data protection. The Health Insurance Portability and Accountability Act (HIPAA) in the US and the General Data Protection Regulation (GDPR) in Europe are commonly consulted but local regulations also exist and are to be respected, particularly when the software is used in multiple states. Adherence to these regulations is meant to guarantee the responsible and secure handling of user data, however these regulations constantly evolve along with technologies, as the landscape of the data economy is dynamic and changes at rapid velocities.

In order to ensure protection against data breaches, where third parties can unlawfully or unethically obtain sensitive data and unauthorized access to patient derived information, strict security measures need to be imposed. Examples of these measures, such as encryption (the transformation of sensitive data into a format that is unreadable), multi-factor authentication (requiring multiple different types of verification for access to the data), and frequent security audits (audits to identify vulnerabilities in the security of the data protection system) are already being utilized for medical data and need to be constantly updated.

Additionally, users must be educated and informed regarding the best ways to safeguard their personal data, like creating strong passwords and spotting phishing scams. Insufficient user knowledge of digital "self-defense" is a key limitation of data self-management, however, it can be prevented to an extent with proper security measures and user awareness initiatives.

When sharing information for public health and research, PHIMS can use data anonymization techniques to protect privacy. These methods guarantee that data cannot be linked to specific users by eliminating personally identifiable information, while they nevertheless yield insightful information for analysis.

ISSN 2281-5147

Key Takeaways: Ensuring the usability and accessibility of PHIMS for all individuals, regardless of age or socioeconomic status, is crucial. This can include designing user-friendly interfaces, providing educational resources and tutorials, and addressing the digital divide between generations. Security and privacy can be of major concern, and require strict adherence to regulations like HIPAA and GDPR, and the implementation of advanced and constantly evolving security measures to protect user data from both breaches and unauthorized access.

6. Summary

Although PHIMS can display many advantages, there are more than a few obstacles that are required to be overcome in order to fully realize their widespread use. Integrating data from several sources and guaranteeing interoperability across diverse systems is one of the major issues. Standardizing communication protocols and data formats can help to ensure smooth data integration and interchange.

It is important to give serious thought to legal issues pertaining to consent, data privacy, and the use of health data in research. To solve these issues, user consent and transparent policies and practices are crucial.

Innovations in technology will continue to influence PHIMS in the future. The capabilities of PHIMS will be improved by advances in data analytics, machine learning, and artificial intelligence, offering even more predictive and personalized healthcare solutions.

Another issue is ensuring public engagement.Promoting user involvement and adoption is essential to PHIMS's success. This entails showing the usefulness of PHIMS in enhancing health outcomes in addition to creating user-friendly interfaces. Adoption and awareness can be raised with the aid of successful marketing and education initiatives.

PHIMS development, implementation, and maintenance can come at a high cost. It is crucial to guarantee these systems' financial viability. This could entail looking into business models, like government funding, partnerships with healthcare providers, or subscription services.

ISSN 2281-5147

Systems for managing personal health information are a major advancement in the digitization of healthcare. PHIMS may help people take charge of their health, improves the effectiveness and efficiency of healthcare delivery, and facilitates personalized care by combining health data into a single, easily accessible platform. To fully utilize PHIMS, it will be essential to guarantee security, privacy, and accessibility while resolving issues with data integration and ethical considerations. PHIMS will become more and more important in influencing how healthcare is shaped going forward, helping to make it more inclusive, data-driven, and personalized. While initiatives for personal information management in healthcare have been made, they face multiple challenges as mentioned above and hence we attempt to establish a framework defining the necessary specifications for the development of medical information management systems as they require further improvements for truly widespread adoption and to reduce digital gap derived discrimination.

7. Selected Readings

Andrea Civan et al. "Personal health information management: consumers' perspectives". eng. In: AMIA Annual Symposium proceedings. AMIA Symposium (2006), pp. 156– 160.

Anton Hasselgren et al. "Blockchain in healthcare and health sciences–A scoping review". en. In: International Journal of Medical Informatics 134 (2020-02), p. 104040. DOI: 10.1016/j.ijmedinf.2019.104040. URL: https://linkinghub.elsevier.com/retrieve/pii/S138650561930526X (visited on 2022-06- 12).

David W. Bates and Asaf Bitton. "The future of health information technology in the patient-centered medical home". eng. In: Health Affairs (Project Hope) 29.4 (2010-04), pp. 614–621. DOI: 10.1377/hlthaff.2010.0007.

Heleen Janssen and Jatinder Singh. "Personal Information Management Systems". en. In: Internet Policy Review 11.2 (2022-04). DOI: 10 . 14763 / 2022 . 2 . 1659. URL: https : / / policyreview . info / glossary / personal - information - management-systems (visited on 2022-06-12).

ISSN 2281-5147

Heleen Janssen et al. "Personal information management systems: a user-centric privacy utopia?" en. In: Internet Policy Review 9.4 (2020-12). DOI: 10.14763/2020.4.1536. URL: https://policyreview.info/articles/analysis/personalinformation-management-systems-user-centric-privacy-utopia (visited on 2022-06-12).

Malgorzata Kolotylo-Kulkarni et al. "Personal Health Information Management Among Older Adults: Scoping Review". eng. In: Journal of Medical Internet Research 23.6 (2021- 06), e25236. DOI: 10.2196/25236.

Massimo Attoresi et al. EDPS TechDispatch: personal information management systems. Issue 3, 2020. Issue 3, 2020. English. OCLC: 1242739955. 2020. URL: https://data. europa.eu/doi/10.2804/096824 (visited on 2022-06-12).

Mirko Zichichi et al. On the Efficiency of Decentralized File Storage for Personal Information Management Systems. 2020-07

Pravin Pawar et al. "eHealthChain–a blockchain-based personal health information management system". en. In: Annals of Telecommunications 77.1-2 (2022-02), pp. 33–45. DOI: 10.1007/s12243- 021- 00868- 6. URL: https://link.springer. com/10.1007/s12243-021-00868-6 (visited on 2022-06-12).

Wanda Pratt et al. "Personal health information management". en. In: Communications of the ACM 49.1 (2006-01), pp. 51–55. DOI: 10 . 1145 / 1107458 . 1107490. URL: https://dl.acm.org/doi/10.1145/1107458.1107490 (visited on 2022- 06-13).

Xiao-Fei Teng et al. "Wearable Medical Systems for p-Health". In: IEEE Reviews in Biomedical Engineering 1 (2008), pp. 62–74. DOI: 10.1109/RBME.2008.2008248. URL: http://ieeexplore.ieee.org/document/4711366/ (visited on 2022-06- 12).

Yongmin Kim Eung-Hun Kim. "Application and evaluation of personal health information management system". In: ().

ISSN 2281-5147

ISSN 2281-5147

WHAT AI IS STEALING ! DATA PRIVACY RISKS IN AI

Soumia Zohra El Mestari*

Abstract

Even if we may not realize it, AI's presence in our lives is increasing at a great pace. Most technological services we use nowadays are driven by AI, and that could be good news since AI's aims to improve the quality of the services. Unfortunately, to work well, AI greedily feeds on user data: AI models collect, process, and store a great deal about us, which is a problem if such sensitive information is leaked. This chapter discusses that this risk of AI's leaking personal data is not only hypothetical and suggests how to mitigate it.

Table of Contents

* PhD student at the Sociotechnical Cybersecurity (IrisC Group) SnT, University of Luxembourg, Soumia.elmestari@uni.lu

Soumia Zohra El Mestari is a PhD student at the university of Luxembourg supervised by Pr. Gabriele Lenzini . Her research interests are in Machine Learning, Trust and Transparency in data-driven tools and Privacy-Preserving machine learning. Prior to joining the Sociotechnical Cybersecurity Interdisciplinary research group, IRiSC, headed by Prof. Gabriele Lenzini, Soumia worked as a machine learning engineer and data analyst. Currently she is pursuing her PhD in IRiSC funded by the interdisciplinary EU project Legality Attentive Data Scientists (LeADS). This work is supported by the European Union's funded project Legality Attentive Data Scientists (LeADS) under Grant Agreement no. 956562.

ISSN 2281-5147

Keywords

Privacy Preserving Machine Learning – Machine Learning – Membership inference attack – Artificial Intelligence – Privacy Enhancing Technologies

1. Introduction

Artificial intelligence (AI) is revolutionising how we perform certain tasks by making it more tech-dependent. Today, AI tools can perform sophisticated tasks more efficiently than we can. These tasks include, for example, video and audio processing [4], natural language understanding [1], summarising and generating content [8, 11], and decision making. However, today, most people are not even aware of the number of AI-based tools they use on a daily basis; since once the technology is spread and used by everyone, it is no longer referred to as AI but rather seen as a mainstream tool, for example, receiving new social media content based on your interest is seen as a "the normal way" by which any social media feed operates. This integration of AI in life often makes users forget that AI is involved in the first place.

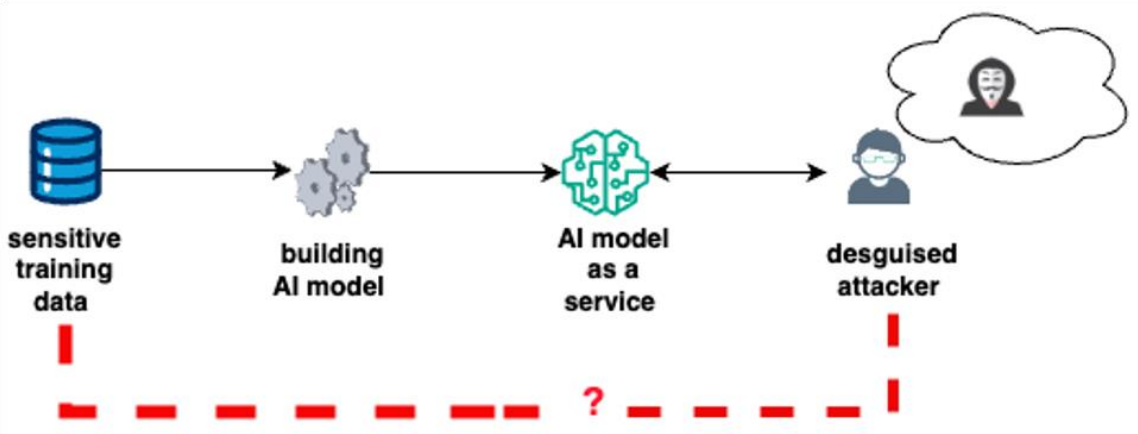
AI tools have the ability to enhance their performance by learning from feed- back and data collected from their environment. Thus, the word "intelligence" comes from this particularity of learning from data without explicit instructions on how to solve the tasks at hand. In this context, when we mention data, we mean huge volumes of

ISSN 2281-5147

images, texts, videos, audio, search feeds, in some cases, health records, social media preferences and anything that can be recorded and stored by electronic means.

The closer the purpose of the AI tool is to humans, the more sensitive the data becomes. For example, an artificial intelligence model that helps doctors make diagnoses needs to be built using patients' health records which are considered highly sensitive. Similarly, targeting a specific range of users for online ads calls for the collection of many users' online traces, including their search history. From another angle, the more data we feed to these tools, the better they become. This data-greedy nature can be seen in the viral ChatGPT-3, which has been trained with roughly 45 terabytes (TB) of text, nearly a trillion words [2]; and despite the huge amount of data collected to build ChatGPT, this AI model continues to collect more data to improve its performance on a regular basis.

This wide adoption has been a game changer in many fields, promising increased efficiency and enhanced cost-over-convenience ratios. However, like any other technology, AI has drawbacks: its efficiency and proficiency come with a heavy privacy bill. AI-tools feed on data, requiring more input data to deliver more accurate predictions, but this data hunger is a threat to our privacy.



2. All emerging technologies raise privacy issues

The growing need for more data to build efficient AI models made regulators and ethicists run the marathon in an attempt to prevent the misuse of data in AI by setting out data protection regulations that establish what should be done so that everyone's data are used properly and fairly. However, the evolution of artificially intelligent techniques is faster than the ability to regulate them, which opens a huge gap in data

ISSN 2281-5147

protection regulations when AI is involved in the process. To explain, although current data protection laws attempt to address these issues, the way AI tools process data is more complex, and regulations are not always flexible to address these complexities. Furthermore, regulations emphasise principles like fairness and privacy without a clear guidance on how to achieve them technically.

There is an inherent opacity in the way AI models operate and specifically in their learning processes, so understanding what steps are taken by these tools to produce certain decisions is challenging [10]. This opacity complicates attempts to detect and prevent data leaks. Under the hood, scientists still face the struggle to justify or even explain the learning patterns that govern the decision-making of a given AI model, which is not good news for lawyers and policy makers.

A significant distinction between AI and traditional analytics technologies is the ability to automate many tasks that humans used to have control over, such as data storage, processing, and maintenance. Thus, AI can interfere by modifying and automating the current applications in which data are used and consequently affects the privacy implications of these applications with little to no accountability to be redeemed. For example, using CCTV cameras for public surveillance is fairly common in contemporary society. This role was previously performed by security staff and in the first implementations of CCTV cameras, the task of refining and analysing the clips of videos was still performed mostly by a human staff and generally done in case of a security-threatening event. However, nowadays, when this technology is coupled with facial recognition software, a camera network could become a much more invasive privacy tool. Thus, the issue is in the way data are processed and the opacity that covers this processing; traditionally, when these tasks were mainly performed by humans, the risks and leakages were simpler to detect, and the responsible for the harm was easier to point out and hold accountable. However, with modern tools that include different layers of AI and little knowledge about processing details, it is harder to spot privacy issues, prevent them, or identify who is responsible for ensuring that they are held accountable.

Furthermore, building AI models to mimic human behaviour, such as voice- based conversational abilities, can give the illusion that these tools have human characteristics. Consequently, we find that users would deal with it as if it were a human being [9], thus giving away sensitive personal data without realising that these personal data may be processed in ways that may not be in line with one's perception of privacy.

ISSN 2281-5147

2.1 How do these challenges look like at the engineering level?

Under the hood, AI models can unexpectedly leak data. First of all, AI models offer results and outputs based on user inputs, for instance, the social media feed is full of ads for products that match the interests of the user since the underlying AI models for this purpose have already received a considerate flow of data that include, but not limited to, browsing history, previous interactions (Likes, comments, or watching video clips) on products or topics that are in line with the user's preferences, adding to that even the messages he may have exchanged with other persons that express these needs or even the interest of these people who he frequently interacts with assuming that socially close people generally share similar interests. All these inputs are sent to servers where AI models are deployed, in other words, data are stored on servers that users ignore their locations with no tangible guarantees to prove that entities who govern these servers will not use the data for other purposes or even monetise them.

In addition to the amounts of data that users have to disclose to the AI services to obtain results like predictions, recommendations, etc. AI is also built from data that are generally collected under a set of terms that do not include the free disclosure of these data to the public. In the best scenarios, user consent is requested for the purpose of building or maintaining AI models not to put the data in public for anyone to see and process. The bad news is that AI models can still leak these data in such a way that anyone using an online service that makes use of AI can push the AI model to leak the data that were used to build it. To picture the seriousness of this potential risk, one may agree on the usage of his medical data to build models that may help the diagnosis of similar pathologies he suffers from in an attempt to contribute to rapid recovery and improving medical services for a larger public. The data sharing terms here include only using the data to build an AI that helps diagnosis, however, when this AI is used to leak information to the large public, this may cause serious social, economic, and legal implications. This leakage can be on different scales from revealing that a certain person participated in a given study [7] to an actual recovery of their entire data [12] [3].

If the challenges mentioned above do not shake our awareness about the risks we face, then the following will certainly do!

ISSN 2281-5147

2.2 The privacy enemies may get away with it!

From a different perspective, identifying privacy issues in machine learning sys- tems can be challenging. In most scenarios, the threat is not detected until a significant data breach occurs! More critically, detecting a potential attacker who shows no signs of malicious behaviour remains a difficult problem. In other words, many attacks that may be conducted against the AI models cause one major problem: They can go undetected, and it is hard to hold the enemy in the loop (i.e., the attacker) accountable for the privacy breach they caused.

The attacker will use the system the same as any benign user who uses it, yet still leak the AI model data, which not only enlarges the pool of potential attackers but also decreases the chances of proving their malicious behaviours, thus holding them accountable and preventing their actions.

3. Are there any solutions offered to mitigate these risks ?

Privacy Enhancing Technologies (PETs) are frequently suggested as means of protecting personal data and achieving general trustworthiness according to cur- rent EU regulations on data protection and AI, this trustworthiness is important to ensure a safe usage of data for the best benefit of society. This set of tools is generally promoted as a means of achieving PPAI (Privacy Preserving AI), also known as PPML (Privacy Preserving Machine Learning)[1]. PETs offer privacy guarantees that depend on how they are applied. Different PETs offer different privacy guarantees and defend against different privacy risks, and there is no Privacy Enhancing Technology (PET) that can solve all privacy issues for a given AI system. Thus an off the shelf usage of these tools is not sufficient to render a privacy invasive AI tool into a perfect privacy-friendly AI. Switching the vision by placing these PETs under a legal lens makes the situation more confusing. The appropriate measures to be used to ensure legal compliance for an AI tool must be built on solid grounds, including an analysis of the whole data flow against the legal and technical guarantees that this flow must ensure.

[1]Machine Learning is a subfield of AI, both terms AI and machine learning can sometimes be used interchangeably

ISSN 2281-5147

4. Research Questions, Findings and Limitations

The first aspect of this project involved exploring the relationship between the requirements outlined in the EU data protection regulations and the actual privacy risks.

This includes creating detailed threat models[2] that take into account the stakeholders involved, the infrastructure where AI is deployed, and various stages of the process before and after implementing PETs.

This kind of analysis takes into account the trust relationships between the different stakeholders along with the guarantees that PETs are designed to offer a direct method of establishing a comparison between the desired privacy guarantees and those actually achieved [6].

This model may look complex, but building an AI tool and deploying it as a system include the participation of many entities depending on the system design, these entities may have different trust assumptions among themselves. For example, the user can trust that the entity deploying the AI model will actually process the data in a correct way to give the actual desired output to the user without outputing a wrong result. The same user may not trust this same entity to keep a copy of their data, in this particular example the PETs used must be really studied to satisfy the privacy guarantees of each entity without compromising neither the correctness of the output nor the privacy guarantees of the other entities.

It is important to recognise that while PET offer important protections, they also have limitations and disadvantages that need to be clearly communicated and considered when aiming for legal compliance in this field.

One of the key findings of our analysis was the shortcomings of current regulations in addressing certain complex AI scenarios, such as when AI models are repurposed, a practice known as transfer learning. This practice can potentially undermine the principle of allowing users to have control over their data. Transfer learning involves changing the purpose of an AI tool to perform another task. This can be achieved even without reusing the original data, making it difficult to detect or prevent data leakage. Legal mechanisms like informed consent struggle to keep up with the various

[2] In security analysis a threat model is a practice that studies a given system by identifying the parties that must be threatened and the potential threatening parties along with the threat points which symbolises the points of interactions between stakeholders that may constitute a risk on one another.

ISSN 2281-5147

potential transfer learn- ing purposes that can arise after a AI model has been built, posing challenges in informing users comprehensively about the data processing objectives. The legal examination of this issue delves into uncharted territories and contentious issues, including the issue of AI model ownership and whether people can claim co-ownership of a model developed with their data. This analysis highlights the inadequacies of existing EU legal tools to address complex AI issues and the limited adaptability of data protection regulations in addressing technical AI challenges [5].

The problem of safeguarding the privacy of the data has an interdisciplinary nature. AI is now being used by organisations and large tech companies, placing a great responsibility on engineers and decision makers to comply with data protection regulations. However, tailoring the technical implementations to the legal provisions faces many drawbacks.

In this type of techno-legal issues, one of the main challenges lies in the terminology used and how the casual use of terms like "anonymisation" can be misleading. In the EU data protection regulations, data that have been anonymised do not meet the criteria to be classified as personal data, which means that the processing of these data is not subject to the same restrictions under the GDPR (General Data Protection Regulation).To this end, a study to explore the idea of anonymisation, which goes beyond mere technical aspects, is crucial. Data anonymisation is like putting on a disguise for sensitive information. Imagine that a data-holding entity has a list of names, social security numbers, and addresses stored in a database. Anonymisation ensures that even if someone gets hold of these data, they will not be able to directly connect them back to specific individuals. The goal is to protect people's privacy while still allowing useful data to be shared and analysed. Unfortunately, the oversimpli- fied link between the legal definition of anonymisation and the technical tools called 'anonymisation algorithms' often leads to their limited use. The term 'anonymisation' as defined in the regulations can be mapped to many PETs including a kind of PETs that is also known under the name of 'anonymisation techniques,' which creates confusion for engineers who may confuse the legal definition with the technical one assuming that the regulation refers only to the set of tools known as 'anonymisation algorithms.' The tricky point here is that in certain scenarios the anonymisation algorithms are insufficient to satisfy the legal provisions and thus result in an underestimation of privacy risks and of the PETs guarantees by both the regulators and the engineers. This superficial approach may hinder effective data processing and

ISSN 2281-5147

also cause confusion within the tech community. Such misunderstandings can harm stakeholders, especially those meant to be protected by regulations.

To put the reader in the view, we describe the leakage risks and the challenges in detecting the risk and defend against it via a simplified example of a secret spy game.

4.1 The Secret Spy Game: Membership Inference Attacks

Imagine your favourite puzzle: AI systems are like that, solving complex problems. But sometimes they accidentally reveal secrets to attackers. These attacks are like invisible ninjas. They do not shout, 'Hey, I am attacking!' Instead, they blend in with regular users. Membership inference attacks are no exception. Imagine playing a super cool spy game. But instead of chasing bad guys, you are trying to figure out secrets about a secret club. Here is how it works:

The Secret Club: Imagine that there is a secret club (let's call it AI Model'). This club knows how to do cool things like recognise cats, dogs, and even unicorns in pictures! But the club has a hidden secret: It was trained using special pictures (such as a secret recipe).

The Spy (Attacker): You are a spy! Your mission: find out if a specific picture was part of the secret training. For example, you want to know if a picture of your cat was used to train the club.

The Clues (Model Output): The club gives you clues. When you show it a picture, it says, 'I am pretty sure this is a cat' or 'Maybe it is a dog?' These clues are like secret messages from the club.

The Sneaky Trick (Membership Inference Attack): You use these clues to guess whether the picture was part of the secret training. If you are right, you have cracked the code! You know whether your cat's picture was in the secret training of the club.

Why does it Matters: Imagine if the club were trained on medical records. Identifying which records were used, you could guess someone's health condition! It is like saying, 'Hey, this person's medical information was part of the secret training; maybe they have a unicorn allergy!' The Challenge The spy game is tough because you do not get to see the secret training pictures directly. You only get the club's hints. Remember, membership-inference attacks are like playing detective with AI models

ISSN 2281-5147

4.2 Catching the spy in the jungle

Spotting your attackers before they leak the data of your model to the large public is difficult and can be impossible in some cases, so imagine looking for them in complex scenarios like the scenario of repurposing AI models or when federated learning[3] is involved ?

The repurposing use case is studied in the context of those large AI tools known as language models. Language models like ChatGPT are making their way into our daily lives in an invasive way. The amount of privacy leakage that these models have proved is alarming. For example, attackers can trick ChatGPT or any other AI tool that generates text to reveal sensitive information about someone's data that were used to build this model by asking it precise questions about this individual such as to reveal their phone number. Thus, putting them under another test in a more complex setting such that of transfer learning have shown interesting results, and despite the general belief that the practice of re-purposing in its technicalities helps in preserving the privacy of data and making the mission of the spies (the attackers) more difficult, when the AI model is a language tool the game balances change.

These AI language tools can become a spy helper disguised as your confident writing assistant.

4.3 Not only that the spy may be a member of the privacy team!

One of the PETs that has a good reputation in the privacy team is a technique known as Federated Learning. In Federated Learning, user data never leave the user's device, and AI tools are created so that multiple users contribute to the construction of the building blocks of AI tools under the governance of one entity called the aggregator. The privacy angel called the aggregator does not have access to the data of the users and its job is to assemble the building blocks that are sent by all users to build the AI model; these building blocks are, however, built by the client from their data.

One of our most interesting studies showed how this aggregator may modify the way users build their AI building blocks to further use those building blocks to extract users' data, the power of this entity being an aggregator is not only underestimated, but when the entity is clearly doing a malicious behaviour it goes undetected! Our

[3] Federated learning is a technique to build ai models by using data from different entities without having to merge all their data in one place, so the training happens in a collaborative way where each data holder does a portion of the processing

ISSN 2281-5147

results suggest that aggregators can spot which users to target and perfectly push them to reveal their data without the users spotting this behaviour until it is already too late and data have been revealed.

Our study also explored the different defence mechanisms and the limita- tions of each. The performant defence strategy included periodically testing aggregators, and once privacy-invasive behaviour is detected, users should opt out of the collaborative learning process (federated learning process).

4.4 Exploring the horizons!

Despite the wave of research efforts in tailoring the privacy risks of AI, the limitations and challenges in the field are severe.

Privacy-preserving AI methods often involve adding noise or altering data to protect privacy. But this can affect the AI tools' performance. Think of it as baking cookies: If you add too much flour (privacy protection), the cookies might taste bland (low accuracy). If you add too little, they might fall apart (privacy breach). Researchers are working hard to find the right balance between privacy and accuracy, but it is a delicate dance.

In addition, implementing privacy-enhancing techniques requires expertise. It is like assembling a puzzle with many pieces. Developers need to understand how to set privacy parameters, choose the right tools, and ensure that the model does not accidentally leak sensitive information. It's a bit like building a sand-castle: You need to know where to place each grain of sand to keep it sturdy and safe. In summary, privacy-preserving ML is like protecting a secret recipe. You want to share the delicious cookies (ML predictions) without revealing all the ingredients (private data). Finding that sweet spot between privacy and accuracy is the challenge!

5. Conclusion

This project aims to study the privacy risks of using AI without being aware of its risks. Allowing your data to circulate without being aware of the different ways your data may be exposed, manipulated, and shared. Our findings show that the risks are hard to detect and can go without being noticed. In addition to that, and from a technological perspective, privacy enhancing techniques are still immature to be used in an efficient way, thus the researchers are still trying to enhance the privacy enhancing versions of AI to achieve the same service quality results without

ISSN 2281-5147

compromising the privacy of the users. Furthermore, regulations need to gain more flexibility to capture all risks and provide users with the necessary legal protection they need.

References

[1] Md Ali, Nawab Yousuf, Md Rahman, Jyotismita Chaki, Nilanjan Dey, KC Santosh, et al. Machine translation using deep learning for universal networking language based on their structure. *International Journal of Machine Learning and Cybernetics*, 12(8):2365–2376, 2021.

[2] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Ka- plan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33:1877–1901, 2020.

[3] Nicholas Carlini, Florian Tram`er, Eric Wallace, Matthew Jagielski, Ariel Herbert-Voss, Katherine Lee, Adam Roberts, Tom Brown, Dawn Song, U´lfar Erlingsson, Alina Oprea, and Colin Raffel. Extracting training data from large language models. In *30th USENIX Security Symposium (USENIX Security 21)*, pages 2633–2650. USENIX Association, August 2021.

[4] Junyi Chai, Hao Zeng, Anming Li, and Eric WT Ngai. Deep learning in computer vision: A critical review of emerging techniques and application scenarios. *Machine Learning with Applications*, 6:100134, 2021.

[5] Soumia Zohra El Mestari, Fatma Su¨meyra Do˘gan, and Wilhelmina Maria Botes. Technical and legal aspects relating to the (re)use of health data when repurposing machine learning models in the eu. In Stefan Schiffner, S´ebastien Ziegler, and Meiko Jensen, editors, *Privacy Sympo- sium 2023*, pages 33–48, Cham, 2023. Springer International Publishing.

[6] Soumia Zohra El Mestari, Gabriele Lenzini, and Huseyin Demirci. Pre- serving data privacy in machine learning systems. *Computers & Security*, 137:103605, 2024.

[7] Hongsheng Hu, Zoran Salcic, Lichao Sun, Gillian Dobbie, Philip S. Yu, and Xuyun Zhang. Membership inference attacks on machine learning: A survey. *ACM Comput. Surv.*, 54(11s), sep 2022.

ISSN 2281-5147

[8] Touseef Iqbal and Shaima Qureshi. The survey: Text generation models in deep learning. *Journal of King Saud University-Computer and Information Sciences*, 34(6):2515–2528, 2022.

[9] Amon Rapp, Lorenzo Curti, and Arianna Boldi. The human side of human-chatbot interaction: A systematic literature review of ten years of research on text-based chatbots. *International Journal of Human-Computer Studies*, 151:102630, 2021.

[10] Cynthia Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligenæ*, 1(5):206–215, 2019.

[11] Liuyang Wang and Yangxin Yu. Research on text summarization generator method based on input text linguistic features and copy mechanism. In *2021 16th International Conference on Intelligent Systems and Knowledge Engineering (ISKE)*, pages 586–590. IEEE, 2021.

[12] Xi Wu, Matt Fredrikson, Somesh Jha, and Jeffrey F. Naughton. A method- ology for formalizing model-inversion attacks. *2016 IEEE 29th Computer Security Foundations Symposium (CSF)*, pages 355–370, 2016.

ISSN 2281-5147

ISSN 2281-5147

CAN BUSINESS-TO-GOVERNMENT DATA SHARING SERVE THE PUBLIC GOOD?

Bárbara da Rosa Lazarotto*

Abstract

Data is considered to be the world's biggest business, leading some to affirm that it can be considered a commodity. Access to data has been essential to promote competition and innovation between different stakeholders, including the public sector. The European Union has enacted a series of Regulations that overlap and interconnect with the main objective of enhancing the sharing of data from all parties. In this context, this research aims to explore them and analyze if they indeed assist business-to-government data sharing.

Table of Contents

* Barbara is an Early-Stage Researcher at Vrije Universiteit Brussel for the LeADS project with a Master's in Legal Studies from the University of Minho. After obtaining her Bachelor in Laws, Barbara worked as a Persecutor Assistant where she handled fundamental rights cases. In her Master, Barbara she had the opportunity to come across the intersection between fundamental rights and technology. At LeADS, Barbara researches the topic of "Public-Private data sharing from "dataveillance" to "data relevance". Due to her background working and researching fundamental rights and technology, she expects to expand her knowledge on how to promote data sharing in a data economy society while protecting the privacy rights of individuals and boosting innovation.
barbara.da.rosa.lazarotto@vub.be
This work is supported by the European Union's funded project Legality Attentive Data Scientists (LeADS) under Grant Agreement no. 956562.

ISSN 2281-5147

Keywords

Data sharing– Business-to-government (B2G) – Competition and innovation – European Union regulations – Data as a commodity

1. Introduction

In recent years, data has become one of the most valuable assets in the world, driving innovation and competition. For example, tech giants like Alphabet and Facebook have user data that accounts for $1.4 trillion of their combined market value (European Parliament, 2021). When multiple entities can use data at the same time (known as non-rivalry), its value increases without losing quality, but this often benefits big firms more. This creates an imbalance in the market as these firms hold onto large amounts of data and share very little with others, such as the public sector.

One big thing about data sharing is when companies and governments exchange info, known as business-to-government (B2G) data sharing. It's not new, but now it's more about sharing cool new datasets from business operations, not just dealing with rules

ISSN 2281-5147

and taxes. This kind of sharing could make markets fairer and help new ideas and competition. The European Union's 2019 report on "Competition Policy for the Digital Era" talked about how sharing data with smaller firms could help them compete with the big data companies (EU Commission, 2019).

In this context, sharing data between companies and governments has typically been about following rules and paying taxes. But now, it's about sharing new types of data gathered through business activities. This could level the playing field in the market, encouraging more competition and new ideas by giving data to other businesses and consumers. According to the 2019 European Union report on "Competition Policy for the Digital Era," sharing data with smaller firms could make the market more competitive (EU Commission, 2019). And it's not just about making money and being competitive. The OECD says that using and sharing data could boost the economy by 0.1% to 2.5% of GDP, showing how it could change public services and decision-making, from improving healthcare and finance to making transportation, education, and infrastructure more efficient (OECD, 2019).

It's important to find a balance between protecting data and citizens' privacy and recognizing that private companies collect tons of data that often can't be used publicly, even though it's really relevant. The European Union has put in place rules to help with sharing data across different areas. These rules are still evolving, and we're still figuring out how they'll affect personal data protection. This article aims to dig into these complexities and see how the new rules will impact how data is shared and protected in the future.

2. The Evolution and Impact of Data: From Historical Burden to Modern Asset

The word "data" comes from the Latin word "datum," which was first used in mathematics in the 17th century. It wasn't until the 18th century that "data" started to be commonly used to talk about the results of investigations, setting the stage for its modern meaning: a mix of number info that forms the basis for arguments and analyses (Rosenberg, 2013).

Throughout history, people have been collecting, storing, and analyzing data to keep track of things. For example, way back in the day, folks in the Nile Basin kept records about farming to figure out how the river's tides worked and predict future crop

ISSN 2281-5147

amounts. But for a long time, governments and companies thought dealing with data was a hassle and didn't see much economic benefit. That all changed when computers and the internet came around in the late 20th century. This made it possible to create interactive and responsive information setups that blurred the lines between physical and digital.

The way technology has evolved, data isn't just a hassle anymore – it's become a valuable resource with many uses: as information, an economic asset, and a public good. This change could be as groundbreaking as the agricultural and industrial revolutions (Floridi, 2014). By looking at different sets of data, we can uncover detailed information about individual behaviors and predict future events like climate changes and disease outbreaks, which can improve and personalize services. With the increase in digital information, private companies have started to gather and process large amounts of data, gaining insights into consumer preferences and making big profits. This has changed the way we see data – from a hassle to a valuable asset (De Gregorio & Ranchordas, 2020).

Back in the day, people didn't think data was worth much. Storing data was seen as a hassle with little payoff. But with the rise of digital tech, data has become a valuable asset, like oil in driving economies. Just like oil, raw data needs processing to create valuable insights. For example, tech companies gather and analyze loads of user data (like preferences for leather shoes) to make profits, leading to more targeted advertising and personalized services. Another comparison likens data to capital, representing the resources that companies and technologies require to operate. Data capital is highly adaptable and can be processed by machines to power business models and technological progress. For instance, a strawberry farm might use data on weather and machinery to improve harvesting, showing the value of data in different industries.

The process of gathering, processing, and getting useful insights from data has led to big tech companies having a ton of data all to themselves. According to a report from the European Parliament, user data made up $1.4 trillion of the combined $1.9 trillion market value of Alphabet and Facebook (European Parliament, 2021). Different industries value different kinds of data based on what they need it for. For example, data about machinery, transportation, and environmental factors can be super valuable for a company that grows strawberries, helping them figure out how

ISSN 2281-5147

machines can make fruit collection more efficient. The more diverse and extensive the data collection, the more insights and applications can come out of it. Even though one person's data might not be worth much on its own, data from millions of people can be super valuable, especially for marketing (Sadowski, 2019).

Data monopolies rely on a data-driven network effect, meaning that the more data a company has, the better its products and services become, like behavior-based advertisements, thanks to improved recommendation systems. This creates a situation where businesses and users depend on a particular company's data collection, leading to lock-in effects that affect data distribution for the public good. Think of data monopolies as medieval fortresses hoarding data behind thick walls. Only those inside benefit, while smaller companies and public institutions struggle to access this valuable resource. Data functions as a form of capital—many technologies and organizations couldn't operate or generate value without it. Data capital, with its digital record, machine-processability, and high mobility, can be converted into economic capital and is crucial for data collection, storage, and processing infrastructure, including smart devices, online platforms, data analytics, and server farms. Driven by the logic of capital accumulation, data collection continually increases in scope and scale (Sadowski, 2019).

In this context, sharing data is a powerful way to spread the benefits of data across society. Once data is created, it can be shared and used by different people for various purposes in both the public and private sectors (Apráez, 2021). A 2019 OECD report points out several benefits of data sharing, such as making how companies work more transparent, increasing accountability, empowering users, creating new business opportunities, and improving efficiency (OECD, 2019).

This second point highlights the idea of data as a common good, which the EU has been promoting recently. This perspective sees data as a valuable resource that can benefit society, so it should be shared for further use to create new and innovative solutions for common challenges, like dealing with the COVID-19 pandemic. However, viewing data as both an economic resource and a common good that should be shared may create challenges, especially in today's data-dependent economy. This requires careful management to prevent potential misuse of shared data and calls for strong data protection measures.

ISSN 2281-5147

3. What is business-to-government data sharing?

Data sharing" is a term that's a bit fuzzy in its definition. The European Commission defines it as "any form of data flow or access between governments, companies, and individuals." However, scholars have their own takes on it. Some emphasize sharing of big, high-quality datasets across different industries, while excluding business-to-government (B2G) data sharing or consumer data portability (Richter, 2019). Others focus on continuous access to specific data categories, differentiating it from one-off data transfers (Feasey and de Streel, 2020). On the other hand, some take a broader approach, similar to the Commission's, encompassing any data transfer between organizations or individuals.

It's important to know that "data sharing" is different from "data access" and "data portability." Article 15 of the GDPR gives people the right to access their personal data, which empowers them. Article 20 of the GDPR, which introduced the right to data portability, was meant to boost competition by letting users switch services easily, but it's actually given more power to data subjects. In short, this study defines "data sharing" as the exchange of data to create a fairer and more efficient data environment, leading to significant economic and societal benefits. Expanding on this, we can break down data sharing into different types, such as between businesses (B2B), between governments and businesses (G2B), and from businesses to governments (B2G).

The idea behind data sharing is that it's a win-win for both the economy and society. Data is seen as something that doesn't get used up when you use it, so the more people who use and share data, the more valuable it becomes. Sharing data can break down big data monopolies, boost innovation, and help the economy grow by letting more data move through different businesses and people. According to the European Parliament, data sharing can create up to 20-50 times more value in the economy, potentially making up 0.1% to 2.5% of the European Union's GDP, depending on the type of data involved. Sharing data can also help businesses directly by expanding their market reach, giving them insights into their performance, and improving their supply chains. But it's not just about money. Sharing data can also have a big impact on the public good. For example, the World Bank showed how mobile phone location data helped trace COVID-19, predict Ebola outbreaks, and track dengue fever. Plus,

ISSN 2281-5147

sharing data can help expose fraud, corruption, and criminal activity through advanced data analysis.

Data sharing can be split into two main groups: voluntary and compulsory sharing. Voluntary sharing happens when entities willingly exchange data without being forced. This includes entities sharing data directly with each other through mutual agreements, as well as platforms that help with data sharing using standardized protocols and agreements (Rukanova, 2023). It also includes cases where entities donate data willingly for public or scientific use and collaborative agreements to achieve common goals. On the other hand, compulsory sharing occurs when entities are required to share data due to regulatory or legal obligations. This includes laws or regulations that mandate data sharing to ensure transparency or competition, as well as agreements enforced by regulatory bodies requiring entities to share data with specified parties. Third parties are often tasked with managing and facilitating data sharing between entities, and legal or regulatory frameworks compel entities to share data to correct market imbalances or ensure public safety, such as the Data Act.

4. Understanding the Drivers and Obstacles of Data Sharing

The existing research has looked at what makes data sharing work or not work. Most of the attention has been on scientific and government situations, but lately people are starting to look more at businesses. But lots of studies forget about outside stuff like politics, society, and money, which leads to messy research across different areas. So, we want to check out the things that stop data sharing, splitting them into three groups: organizational, technological, and environmental. Even though we look at these things one by one, they often mix together in real life.

Organizational factors are like the internal aspects that shapes how a company or organization works. This includes things like how big the organization is, its history, the relationships between people there, and the overall culture. Self-interest, the way people see the company from the inside, and what they know about the organization also really matter. Data is like the money and the connection inside an organization. So, deciding whether to share data is all about the habits and ways of thinking in the organization. If sharing data isn't a big part of how the organization works, then trying to make it happen can be tough. Also, if different groups inside the same organization,

ISSN 2281-5147

like bosses and regular workers, have different ways of doing things, it can make sharing data even more complicated.

One big barrier is the lack of trust between organizations. When there's no trust, they might not want to share data because they're worried it could be used for other stuff without their permission. Also, competition in the market can stop them from working together, making them think it's every man for himself and they should keep all their data to themselves.

When it comes to data sharing, technological factors like hardware and software play a big role. It's super important for organizations to have compatible systems, because different tech setups can cause a lot of issues. If organizations collect and process data in different ways, it can lead to all sorts of problems like incompatible formats, standards, and databases. Some organizations even use their own unique systems to make it hard for others to share data. It's like everyone speaking different languages and not being able to understand each other. Just like interpreters are needed to bridge language gaps, we need technological solutions (like data standards) to overcome data sharing barriers. Plus, keeping data safe during sharing is a big challenge, and it can cost a lot to make sure systems and data formats are all aligned. This cost can sometimes outweigh the financial benefits of sharing.

Environmental factors include legal, socioeconomic, and political influences. These factors are connected to organizational and technological aspects, impacting data sharing in different ways. Socioeconomic factors include societal structures, trust among citizens, and community interest. For example, in a cultural environment with more trust between citizens, data sharing will be more common. Political factors also impact data sharing, with political preferences influencing the willingness of governments and stakeholders to share data. Legislation is important, but balancing data-sharing regulations is tough. Over-regulation can create rigid conditions, while under-regulation can leave gaps that hinder data sharing. Economic factors, like market failures and a business model that makes companies depend on resources offered by other companies, can also hinder data sharing. Data monopolies and information asymmetry create a competitive environment where "data-rich" organizations are hesitant to share with "data-poor" ones. The high costs of ensuring technical compatibility and the risks associated with sharing data further discourage organizations from sharing.

ISSN 2281-5147

5. The European Data Strategy and the sharing of data for the public good

When it comes to businesses sharing data with the government (B2G), the EU has been working on rules to make it easier for public and private parties to exchange data. For example, they've got this thing called the EU Free Flow of Non-Personal Data Regulation, which is meant to get rid of unfair national barriers and stop non-personal data from getting locked in.

In 2020, the European Commission introduced the "European Data Strategy" to tackle the problem of companies holding back data due to trade secrets and proprietary measures. The aim is to create a single market for data, making it easier for data to flow across the EU among businesses and consumers. The plan includes legislative acts to break data monopolies and ensure that data can be used for the public good while complying with European values and regulations.

The Data Governance Act (DGA), which came into effect in June 2022, is a big part of the data strategy. It sets up a framework for private companies to use public sector data. The main goal is to build trust in data transactions, covering public and private non-personal data and personal data that's shared voluntarily. The DGA regulates data intermediation service providers, public sector bodies, and data altruism organizations, encouraging the wide reuse of public sector data for both commercial and non-commercial purposes. Data intermediation service providers act as neutral third parties that connect data providers (individuals and companies) with data users. Their aim is to make the exchange of data, especially personal data, secure and trustworthy. Data altruism organizations are like charity donation centers but for data. Instead of donating clothes or food, people and companies can donate their data to help solve societal challenges like public health or urban planning. These tools aim to build trust in voluntary data sharing, bridging the gap between the public and private sectors while protecting individual rights as outlined by the General Data Protection Regulation (GDPR).

The Open Data Directive, which came into effect in July 2019, is all about promoting digital innovation by making it easier to reuse public sector data. It mandates that important datasets should be available in machine-readable formats, so that both businesses and non-commercial organizations can use the data for creating new and

ISSN 2281-5147

cool solutions. However, there have been some issues with getting it up and running. Progress has been slow in member states, and like the Data Governance Act, it mainly focuses on making public data available to the private sector without requiring them to share their own data in return. This one-sided approach limits the potential benefits of data sharing for the public good.

The Data Act, which started in January 2024, aims to make sure that data is shared fairly among different companies. It encourages sharing data between small and big companies, including personal and non-personal info. For example, Article 5 lets users share data with other companies if they ask, but big tech companies can't get this data. The shared data has to be given under fair conditions. Article 4 also says that users can access and use data from their own products for free. The Act also makes it necessary for companies to share data with the public sector in special situations, which is a big change from the old rules that mainly focused on public-to-business and B2B data sharing.

Another recent regulation targets collecting and sharing data related to short-term accommodation rental services, such as Airbnb. This proposal is all about setting up a way for hosts and platforms to share data more easily. The idea is to make it simpler for local governments to put their rules into action, cut down on red tape, and help with city planning by getting all the data to line up. The rules are meant to deal with the issues that short-term rentals are causing in European cities, like higher rents and the impact on tourism. It also aims to sort out the uncertainties in the rules and the lack of data that local governments are dealing with (Scassa, 2017; Ranchordás, 2018). The proposal says that platforms have to gather and share info like hosts' names, ID numbers, addresses, and contact details. Platforms also have to let hosts say where they're renting out, so it's easier to make maps that show where short-term rentals aren't allowed. With these rules, the EU wants to find a balance between promoting data sharing and protecting competition and people's privacy, and build a strong data economy that helps society.

Despite the progress in the laws, the European Commission still needs to do more to make sure that everybody can easily share and reuse data. Right now, they're mainly focused on letting private companies use public sector data, but we need to think about making private companies share some of their data too, especially data related to connected objects. It's important for the European lawmakers to really think about

ISSN 2281-5147

the power balance between private and public sectors. We need strong rules to stop private companies from keeping all the data to themselves and to give the public sector access to that data. This way, we can make sure that data sharing is fair and safe for everyone, and that it benefits society as a whole.

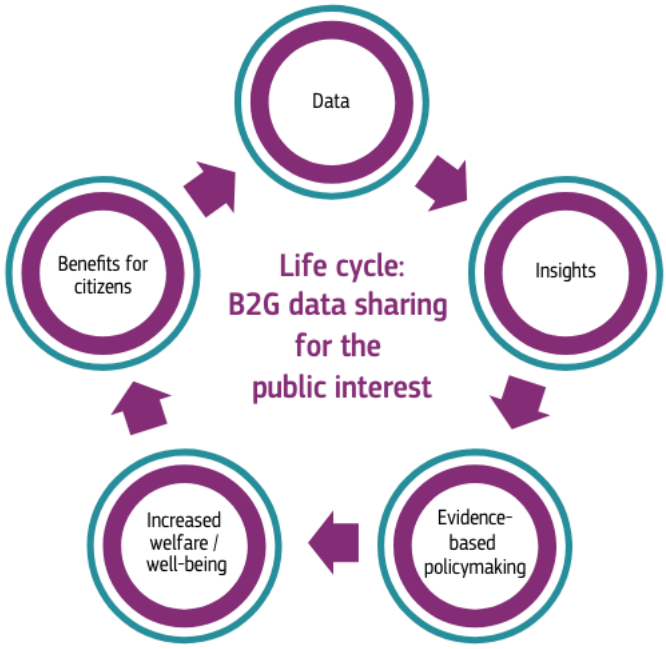


Figure 1. B2G Data Sharing For The Public Interest source: Alemanno A. Towards a European strategy on business-to-government data sharing for the public interest. Final report prepared by the High-Level Expert Group on Business-to-Government Data Sharing European Commission. 2020 Oct 16.

ISSN 2281-5147

6. Case Study: Business-to-Government Data Sharing in Smart Cities

Data is relevant for smart cities because it helps with decision-making, makes public administration more efficient, and tackles urban challenges effectively. It powers the cool technologies that make urban areas more efficient, sustainable, and responsive. With this in mind, this section will look into how sharing data from businesses to government can help build smart cities.

Cities are getting more and more important, with about 55% of the world's population expected to live in urban areas by 2050. As cities grow, governments are realizing that using data can help them do a better job and come up with new and targeted policies. Data from private companies like phone companies, online platforms, transportation services, and energy providers can be really useful for solving urban problems using technology. The term "smart city" is used a lot, but it means different things depending on who you ask. For this case study, we're saying smart cities are a mix of physical, digital, and human systems that are all about finding new ways to deal with the challenges of growing urban areas.

The European Data Strategy Regulations haven't shown their effectiveness yet, especially with the recent enforcement of the Data Act. The Data Governance Act introduced ways for the private sector to share data voluntarily, which could help smart city initiatives. However, there's a risk that tech companies could get access to even more data without reciprocating, creating imbalances between the public and private sectors. The Open Data Directive requires public high-value datasets to be available in machine-readable formats, allowing both businesses and non-commercial entities to use the data for smart city solutions. But the directive has been slow to implement across member states and focuses mainly on making public data accessible to the private sector without requiring them to share data in return. This one-sided approach limits the potential benefits of data sharing for the public good in smart cities.

The Data Act is the latest law that might help smart cities. It allows government bodies to ask private companies for data in specific situations, which is the first time this kind of sharing has been required by law. It also says that data must be shared during public emergencies, which could help smart city projects by giving them more

ISSN 2281-5147

data to manage things like climate crises. The rules for businesses sharing data with the government are pretty strict and limited. They only say that data has to be shared in emergencies or when it's really important for dealing with emergencies. This means the Data Act isn't trying to make a big system for sharing business data with the government, instead it's just for special situations. At first, the law said that government bodies could ask businesses for data for any good reason, but they took that out of the final version. This change might be because the law was made after the peak of the COVID-19 pandemic, when there was a lot of sharing data between businesses and the government in the EU. But the law still lets the government get data in other ways, like buying it, getting it for free, or working with businesses, which are things that smart cities often do (Lazarotto, 2022).

7. Conclusion

In conclusion, it is crucial for European lawmakers to really look into and deal with the power dynamics between private and public sectors. We need some solid rules to stop private companies from hoarding data and to make them share it with others, including the public sector. This will create a fair and balanced setup for sharing data, which will be good for everyone. The European Union's efforts to share data are making smart cities more efficient and responsive to public needs, but it's too early to say if it will really benefit sharing data between public and private sectors in smart cities. We need to keep working on fixing power imbalances and making sure data benefits everyone. With the right laws and collaboration between public and private sectors, we can make the vision of smart cities powered by shared data a reality.

8. Selected Readings

Apráez, B. E. (2021). Reconsidering the public-private data dichotomy in the European Union's data sharing policies. European Journal of Law and Technology, 12(1).

Crémer, J., Yves-Alexandre, M., & Schweitzer, H. (2019). Competition policy for the digital era—Publications Office of the EU (europa. eu), Publication Office of the European Union.

ISSN 2281-5147

da Rosa Lazarotto, B. (2022). The Data Act: empty promises for business-to-government data sharing? A critical analysis of the Proposal on the Data Act and its implications for the redistribution of data. Privacy in Germany, (5/2022), 1-6. https://doi.org/10.37307/j.2196-9817.2022.05.04

De Gregorio, G., & Ranchordás, S. (2020). Breaking down information silos with big data: a legal analysis of data sharing. In Legal Challenges of Big Data (pp. 204-231). Edward Elgar Publishing.

Feasey, Richard, and Alexandre de Streel. Data sharing for digital markets contestability: towards a governance framework. Centre on Regulation in Europe asbl (CERRE), 2020.

Floridi, L. (2014). The fourth revolution: How the infosphere is reshaping human reality. OUP Oxford.

Gawer, A. R., & Srnicek, N. (2021). Online platforms: Economic and societal effects. Panel for the Future of Science and Technology (STOA) European Parliament.

Madison, M. J. (2020). Tools for data governance. U. of Pittsburgh Legal Studies Research Paper, (2020-23), 29-43.

Ranchordás, S. (2018). On Sharing and Quasi-Sharing: The Tension between Sharing-Economy Practices, Public Policy, and Regulation. The Sharing Economy: Exploring the Challenges and Opportunities of Collaborative Consumption, Santa Barbara, CA, Praeger, 263-289.

Richter, H., & Slowinski, P. R. (2019). The data sharing economy: on the emergence of new intermediaries. IIC-International Review of Intellectual Property and Competition Law, 50, 4-29.

Rosenberg, D. (2013). Data before the fact.

Rukanova, B. et al., 'Public Value Creation through Voluntary Business to Government Information Sharing Enabled by Digital Infrastructure Innovations: A Framework for Analysis', Government Information Quarterly 40, no. 2 (1 April 2023): 101786–101786, https://doi.org/10.1016/j.giq.2022.101786.

Sadowski, J. (2019). When data is capital: Datafication, accumulation, and extraction. Big data & society, 6(1), 2053951718820549.

ISSN 2281-5147

Scassa, T. (2017). Sharing data in the platform economy: A public interest argument for access to platform data. UBCL Rev., 50, 1017.

ISSN 2281-5147

ISSN 2281-5147

COLLECTIVE CONSENT, RISKS, AND BENEFITS OF DNA

Xengie Doan*

Abstract

Health data is sensitive and sharing it could have many risks for per- sonal or shared genetic data. So how can impacted individuals consent together? Collective consent has been used in person, but no digital collective consent exists yet. Challenges span legal-ethical issues and technical properties such as transparency and usability. To address these challenges, this work uses genetic data sharing as a use case to better understand what tools and methods can enhance a user-friendly, transparent, and legal-ethically aware collective consent.

Table of Contents

* Xengie is a nonbinary scientist with Master's in Bioinformatics from the University of Oregon. After obtaining their Master's, they worked as a bioinformatician at the Stowers Institute for Medical Research assembling genomes, analyzing repetitive DNA sequences in cancer genomes, and creating semi-automated analysis pipelines. They then worked as a bioinformatics engineer at Sage Bionetworks helping build infrastructure, community sourced metadata models, and user-friendly dashboards and tools as part of a data coordination center for the Human Tumor Atlas Network. Now, Xengie is a PhD student with Dr. Gabriele Lenzini in the IRiSC lab at the SnT, University of Luxembourg working on transparent, private, and user-centered consent eHealth data sharing in the EU with LeADS.

This work is supported by the European Union's funded project Legality Attentive Data Scientists (LeADS) under Grant Agreement no. 956562.

xengie.doan@uni.lu

ISSN 2281-5147

Keywords

Consent – Genetic Data – eHealth – GDPR– Business Process Improvement

1. Understanding Genetic Data Sharing and Consent for Everyone: A Story

Have you ever used a DNA test before? Maybe someone you know has. This is a common experience, as millions of people worldwide have bought and used a direct-to-consumer genetic test kit. We begin with a fictional story to illustrate the challenges in consent and sharing genetic data. The state is set in the quaint town of Genetica, where there was a family known for their unity and love, the Helixes. One day, the youngest member, Ada, driven by curiosity, decided to take a direct-to-consumer DNA test. From the website, it looks easy to do. She pays for a test and is excited to uncover the mysteries of their ancestry and share the exciting findings with her family. She follows steps outlined in the image below, which is taken from FamilyTreeDNA.com.

ISSN 2281-5147

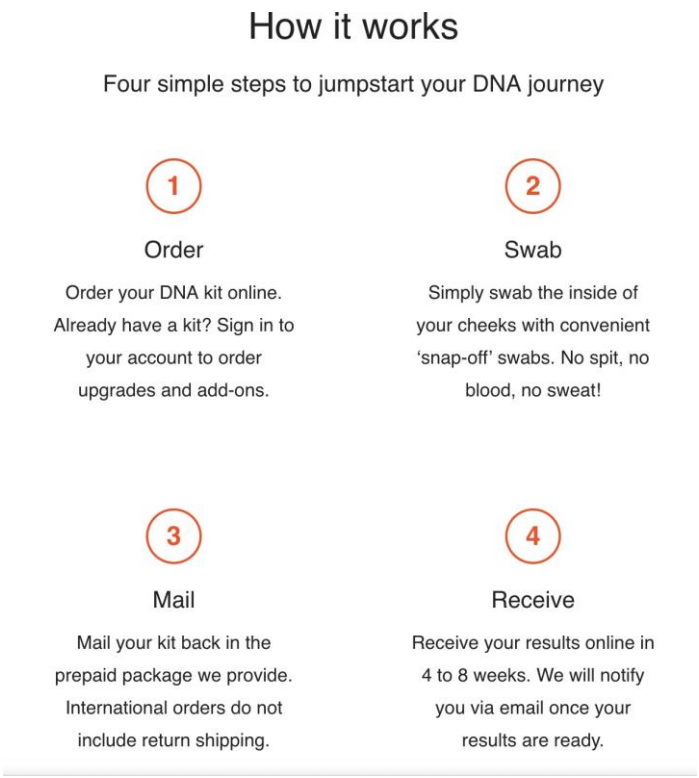


Figure 1: Steps for a DNA kit from FamilyTreeDNA.com.

As the results came in, Ada was thrilled. They were descendants of an interesting region of the world and had a possibility of diabetes! Eagerly, she shared the news at the next family gathering, expecting joy and wonder. Instead, she was met with a mixture of reactions.

Her uncle, a teacher, frowned deeply. "Did you consider the privacy risks, Ada?" he asked. "These companies collect DNA and ancestry information and don't store it securely. Haven't you heard of the recent data breach? They gather vast amounts of genetic data that could be used in ways we can't even imagine yet.

ISSN 2281-5147

Ada's cousin, a lawyer, chimed in, "He's right. Law enforcement has used such data to track suspects without anyone's notice, consent, or knowledge ... ever heard of the Golden State Killer? Insurance companies might access it to increase premiums, and schools could use it to bar someone's admissions. The implications are far-reaching."

The family matriarch, Grandma Helix, who had seen much in her years, spoke softly, "Not all of us are equally at risk, my dear. What if someone will be discriminated against based on this information? Maybe our diabetes will be a sign of poor health and used against us. What if your future kids do not want their data shared? Your decision affects us all, yet we had no say in it."

Ada's heart sank as she realized the gravity of her action. The DNA data, like threads weaving through their lives, connected each family member, carrying potential risks they hadn't consented to.

As the discussion unfolded, the Helixes understood that while science might not yet predict their ancestry and health risk accurately, the future held possibilities that could turn their genetic information into a double- edged sword. The family decided to establish a formal process for such decisions, ensuring that everyone's voice was heard and recorded. They realized that in the tapestry of genetics, each thread was vital, and every pattern mattered.

From that day on, the Helixes approached their shared genetic legacy with caution and respect, knowing that the choices of one could unravel the fabric of many.

2. So Why Should I Care?

If you were placed in the story, how would you feel?

Researchers have been studying how we can all agree (or not) to share this kind of sensitive personal information. Personal information is anything that might identify someone under European law, and there are specific types that are called "sensitive" because there are added concerns. Genetic and health data is one of them, along with things like religion. Genetic data cannot be changed (like in the movie, *Gattaca*), and it can encode information about someone's family history, health, and more. It can be

ISSN 2281-5147

hard to remove identifying information from DNA data, especially if there are multiple databases people can use to cross-reference. Now the world is collecting more data than ever, and it's not just about one person anymore. It can also be used for big data analytics, like machine learning algorithms or artificial intelligence. Our data can tell stories about our friends, family, and even people we don't know but who are like us in some way. It can be a big help for scientific and medical research, but it can also be leaked in data breaches or used without consent. We need a group decision-making system that's clear, fair, and respects everyone's rights—just like how indigenous groups have been doing it for a while.

As databases become more comprehensive, using a mix of public and privacy information increases the likelihood of re-identifying individuals from supposedly anonymized data. Researchers have cross-referenced datasets to uncover patterns that point to specific individuals, showing how easy it is to use public data to reveal sensitive information. This could lead to scenarios where a person's genetic traits, predispositions to diseases, or even their full names could be exposed without their knowledge. 23andMe also had a data break where millions of people's health data was leaked, which was released on the dark web and poses a privacy risk.

The traditional model of individual consent is not equipped to handle the complexities of genetic data, which inherently involves more than just the individual. This is where the concept of collective consent gains importance. It acknowledges that decisions about genetic data sharing should not be made in isolation. Instead, they should involve all those who could be affected by the data's disclosure. Collective consent allows for a more democratic approach, where the rights and concerns of everyone are considered, and decisions are made with the consensus of the group.

The implementation of collective consent would require a new frame- work that respects individuals while also recognizing the interconnected nature of genetic data. It would also need to have clear communication and understanding among all parties involved, ensuring that the benefits and risks of data sharing are transparent and well-understood. Under the GDPR, the key data protection regulation in the European Union (EU), this type of specific, informed consent is key.

ISSN 2281-5147

Collective consent would expand individual ethics to include collective ethical approaches, resulting in a more equitable approach that empowers individuals and groups to have a voice in decisions that could impact their privacy and well-being. This should be clearly communicated and user- friendly to foster a sense of shared responsibility and mutual respect in the management of sensitive genetic information.

The Sensitivity of Genetic Data: Sharing health data, particularly genetic information, carries significant risks. Even when anonymized, genetic data can potentially be traced back to the individual it came from (Erlich et al., 2018). This could reveal not just personal health risks, but also familial connections and population level health risks. Some benefits could be that health information about diseases helps to improve the family, but there are also risks. Such information could be used by the police to find suspects or adjust insurance premiums (Joly et al., 2013), affecting not only the person who agreed to share their data but also their relatives and others with similar genetic markers. In addition, there can be informal sharing of genetic results, like those from consumer DNA tests. But what if family members disagree with this decision? They have no formal say in the matter, even though they could be impacted. This may be especially complex when in the future, those generations wish they could track down how DNA data was shared and delete it. If there is no formalized tracking of the consent and data, it would become impossible to manage personal" data after a few generations.

Collective Consent: This is where collective consent comes in, which can help to protect people's privacy and create a formal system for shared notice, consent, decision-making. Collective consent is a way for groups to agree on sharing genetic data. This idea comes from indigenous com- munities, who have asserted their right to approve biomedical research collectively, rather than individually (Hudson, 2009). This approach re- quires engagement with research partners and respect for the community as stakeholders rather than mere subjects, including considering non-Western cultural beliefs. To adapt it to everyone, collective consent would involve shared decision-making and keeping records of who agrees or dis- agrees with sharing genetic information. This is not just about being fair; it's about respecting each person's autonomy and rights. A system that formalizes decision-making and consent could protect privacy and individual rights. Collective consent works for indigenous

ISSN 2281-5147

groups anymore; and it could be adapted for broader use in genetic data sharing. While the European Data Protection Board, an independent European Union body that is in place to ensure a consistent application and enforcement of data protection law, stated in their guidelines for DNA data recognize genetic data as personal data that may affect more than one person (e.g., may identify all members of a family because they are all genetically related), the current laws are designed for individuals, not groups, making enforcement tricky (*12178/03/EN WP 91 Working Document on Genetic Data*, 2004, Kuru and de Miguel Beriain, 2022).

A Framework for Digital Collective Consent: Collective con- sent has been effective for ensuring autonomy and rights for indigenous groups, where decisions are made by community leaders. However, in the era of extensive data collection, similar approaches may need to be digitalized to operate the new era of digital data sharing. While the goal is to develop a framework for digital collective consent, many challenges must be addressed first, including the transition from individual to collective models and the integration of user needs and risks. While collective consent is recognized in traditional settings (e.g., for physical consent where the participants are in-person), its digital counterpart is still in development. There are complex issues to resolve, such as balancing the right to privacy against the need to inform, defining decision-making processes, and establishing governance structures for the collective. Legal, ethical, technical, and trust-related factors all play a role in shaping digital collective consent. Consent itself is also complicated. Things like digital literacy, reading level, and data management preferences vary among users (Niemiec et al., 2018. Systems must consider informational transparency, storage and access controls, and user interface design. Trust in the institution requesting consent is also crucial and can be influenced by past events, such as data breaches.

A promising solution is dynamic consent, a flexible model that allows individuals to manage their consent and interact with research projects (Haas et al., 2021, Mascalzoni et al., 2022). It could potentially align with collective consent, but research on group decision-making in a digital con- text for genetic data is lacking. Questions remain about the appropriate level of transparency and the most relevant user attributes. This research looks to find a framework for digital collective consent that can address the needs that people have, help businesses improve their services, and share information

ISSN 2281-5147

clearly. This leads to the research questions, which will not address the full problem, but try to answer a small part of it.

3. Research Question

The journey towards a formalized, legal-ethical framework for collective consent in genetic data sharing is filled with challenges. This work will explore the challenges of collective consent, then research user needs and the potential for tools and methods to increase transparency of the systems and address user needs. Legal, ethical, business, and stakeholder considerations must be balanced to create a consent process that is both effective and respectful of individual and collective rights.

To answer this, here's what I did:

- Looked at the Gaps: I worked with legal and ethical experts to check the ongoing challenges in consent for sharing health data in the EU.
- Checked the Fine Print: I worked with another lawyer to also read through the privacy policies of companies that test your DNA to see if they're clear, relevant, and fair.
- Asked People What They Want: Then, we asked regular adults in Germany what they wanted from consent and what format (video, text, newsletter, comic, infographic) they would prefer.
- Tried It Out: We tested methods to improve information transparency and clarity with a company in Norway to see if they could make better privacy policies and consent processes.

4. Results

4.1 Legal-Ethical Gaps

Self Determination: Imagine you're signing up for a new health app on your phone. It asks you to agree to terms and conditions that are really long and full of complicated words. This is about giving your permission, or consent, to let the app use your health

ISSN 2281-5147

information. But it's not just for your doctor anymore; now, this info might be used for other things like ads, which are not necessary for your care. The problem is, that these apps don't always make it clear what you're agreeing to. Sometimes they don't even have a privacy policy that's easy to understand. So, when you just tap "agree," you might not really know what you're getting into. This makes people worry that we're not really in control of our own information, which is super important when it comes to our health.

So, are we really making our own choices if we don't fully understand what we're agreeing to? It's important that we can make our own decisions, which is called *self-determination*. This right is reflected in legal and ethical guidelines and regulations in Europe, and consent is a key part of lawful data processing and biomedical research.

Genetic Data is Shared: In Europe, the data protection board guidelines explore the way genetic data is shared between people. However, its just a guideline, so there are no laws that force people to address collective data privacy. This is because the rules, if they went into action, would be very complicated. The current system is based on individual rights, and there is no precedent or framework for how to manage multiple people's rights together. There are no rules for agreements or disagreements. What if one person doesn't want to share the data, but another person does? Or what if someone doesn't even want to know the results or risks at all? Some think we can sort out these tricky situations by starting with the rules we already have and working from there. Different countries are also trying their own ways to balance everyone's rights. For example, sometimes they decide that knowing about health is more important

than keeping a secret and a doctor can share important health information with relatives.

Specific Consent: The General Data Protection Regulation (GDPR) says that consent should be clear and specific. This means you should know exactly what you're agreeing to and why. But sometimes, it's hard to be specific. Like when scientists collect data for research, they might not know all the ways they'll use it in the future. So, they ask for a "broad consent," which is like saying, "I trust you to use my data for good things later on." The law tries to protect us by saying we need to give specific consent. But if we're too strict about this, we might end up having to say "yes" over

ISSN 2281-5147

and over again for each little thing, which can get really tiring. There's a part of the law (a Recital that tries to expand on the main body) that tries to help by saying it's okay to give broad consent for science, as long as it follows good ethical standards. But this isn't supported in the core regulation, which still talks about being specific. So, it's confusing in practice and people can interpret it in different ways.

Consent is Relevant Even when its not required by the law, ethical consent plays a crucial role in data processing. First, what is the difference? Legal vs. Ethical Consent:

- Legal consent refers to obtaining permission from individuals to process their personal data based on legal requirements (such as the General Data Protection Regulation, GDPR).
- Ethical consent, on the other hand, goes beyond legal obligations. It ensures that data processing respects human dignity, autonomy, and privacy.

Second, ethical consent is relevant as a legal safeguard to help make sure people's right are taken care of. Some researchers (Staunton et al., 2019) argue that ethical requirements such as consent and transparency could serve as safeguards to help inform the data subject of their rights. Third, to uphold self-determination, consent should be asked. Ethics scholars have prioritized this decision making, or autonomy, and they are part of the internationally recognized guidelines regarding people's rights for medical research, like the Belmont Report, to ensure people's rights are safe in biomedicine. It should also be considered in a collective sense,

when multiple people's autonomy is involved.

Without consent, there would be less self-determination and safe- guards for data collection, sharing, and processing activities.

4.2 Genetic Testing Policies

Privacy policies are like public explanation of the internal rulebooks that companies follow when handling your personal data. They are required to write their data processing activities so you can be informed of what is happening, how they store the data, and more. Transparency means being open and clear about what they're doing

ISSN 2281-5147

with your information. While the companies should write everything they are doing and have it been clear what is happening, many privacy policies fall short, causing confusion and misaligned expectations. Maybe you think they are using the best privacy techniques, but your data can be easily hacked. Companies that test DNA data are dealing with sensitive genetic data (like DNA, your family tree, health and wellness information, and more). They have millions of users, but how well do they communicate their practices to users?

We looked at privacy policies from the 6 top companies chosen and analyzed the sections that talk about how they share genetic data. Shockingly, 81% of these explanations were vague, using terms like may, possibly, perhaps, etc. 37% were confusing with more than 2 distinct subjects or purposes for processing the data (some had 10+ unique purposes in one section), making it hard for users to understand. The GDPR requires clear, direct, and transparent information so they might not fully meet the legal requirements. In addition, the way the information is framed is tailored for legal experts, so it makes it hard for a normal person to understand what is happening. They also don't address the collective responsibility of sharing DNA data, which can affect your family. They just say how it will affect their single customer, which makes it seem like it's not a big issue. Some companies only share one risk of anything happening, while others detail about the possible risks. Some could be great and affect your family. So, we suggest making policies more user-centered by framing it in an useful way to general users and more risk aware.

4.3 User Needs and Desires

Consent forms for individuals are complicated. So how can we make them better and try to apply them to collectives?

We looked at what people wanted from consent and the way they want it presented. We took a portion of the consent process and translated it into an infographic, video, text, newsletter, and comic. Then we inter- viewed 24 German adults about their expectations and experiences with consent forms.

We found that people have different goals when reading consent forms, and while everyone wanted easy to understand information, some relied more on trust (e.g., how

ISSN 2281-5147

much they trusted their doctor or the organization) or placed a burden on themselves to record the consent decision.

The infographic was the top voted way to receive consent information because it helped understanding and they could focus on important info. It suited serious scenarios (like health data consent), where comics were seen as childish or unserious for health data consent. Things like structure and readability were also important because they helped make the form engaging and easier to follow.

The participants also wanted a centralized digital platform to manage consent over time, and possibly to revoke consent.

4.4 Business Testing

Following the previous results, we were interested in how to improve privacy policies before they go public. They should reflect the company's own actions for how they collect, use, and share your data. To create better privacy policies, we need a solid approach, or methodology. While there are many methods, none considers the context. In this ongoing work, I tested the usefulness of methods for clarifying privacy policies and consent processes with 13 employees at a company in Norway to see if they thought it would be useful for their jobs.

Results are still being published, but it seems that the employees do appreciate the methods for giving a more structured, visual, and understandable way to process very complex information.

5. Limitations and Future Work

First, these studies cover many different areas with a limited scope. Many of the studies mentioned were pilot projects to check the feasibility and interest in the method, idea, or framework. While useful, they might not cover all aspects or scenarios. More in-depth, larger-scale studies could provide a more comprehensive view. Second, there may be sampling bias. For different studies, the way participants were chosen for the study could introduce bias. This is especially true for the testing in Norway because we were constrained by the availability of people on different teams. In the German

ISSN 2281-5147

study, we looked for an equal number of people of different ages, education levels, and sexes. A more diverse participant pool would enhance the study's validity. We also only looked at one company, and it might not be as reliable. More deliberate sampling and larger and more varied samples would strengthen the findings.

The work was also around a specific domain, so it would not be applicable to others. The work focused on health and DNA data and applying these findings to other fields (like finance or social media) might not work because consent norms can differ across domains. Consent has a strong history in biomedicine and people are used to it in health-related scenarios. From the study with German participants, they took it more seriously than cookie consents.

The studies also had a short-term focus. We looked at privacy policies and laws at one point in time, and the user studies were for immediate perceptions, understanding, or engagement. We did not explore how perceptions might change over time, or measure beyond people's perceptions. Maybe people with have different ideas after working with the concepts for longer, or putting it into practice.

Future work includes repeating and extending the studies. They should be repeated, and if people were interviewed, more people should be interviewed. It should try to get people from many different demographics, like country, age, sex, etc. to be more generalizable, so the results could apply to more people. For the user needs, the study should be repeated in different countries to get a better understanding of how the context (like how comics are not serious enough) might change from country to country. It would be interesting to have a prototype of collective consent using all the suggestions from this work to see how people react to it.

6. Conclusion

From this, we have a better understanding of what users want and can try to translate the findings for collective consent design. This can tie into previous methods such as the analysis of privacy policies, which can be shown before the consent form, and be prototyped all together into a dynamic consent system. Altogether, it is a first step into building a digital collective consent, deeply considering the gaps, user-needs, and possible implementations in small businesses. As shown in Figure 2, the dynamic

ISSN 2281-5147

consent prototype should be built and tested based on a specific use-case for collective consent on the findings from this work, then be used iteratively to build better systems from the ground up.

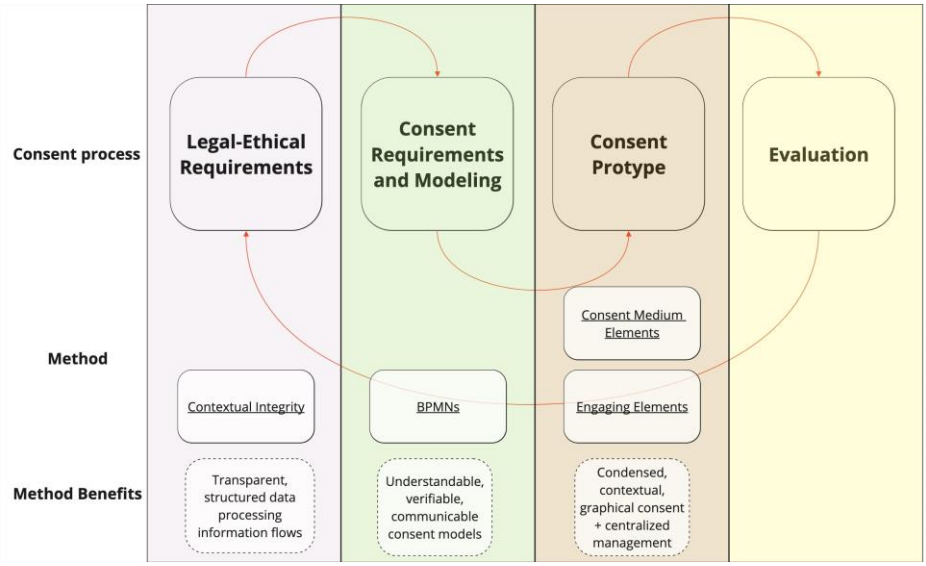


Figure 2: A visual summary of the work as it targets different parts of the consent process, the methods tested, and the top findings

References

12178/03/en wp 91 working document on genetic data. (2004). https :

/ / ec . europa . eu / justice / article - 29 / documentation / opinion - recommendation/files/2004/wp91 en.pdf

Erlich, Y., Shor, T., Pe'er, I., & Carmi, S. (2018). Identity inference of ge- nomic data using long-range familial searches. Science, 362 (6415), 690–694.

Haas, M. A., Teare, H., Prictor, M., Ceregra, G., Vidgen, M. E., Bunker, D., Kaye, J., & Boughtwood, T. (2021). 'ctrl': An online, dynamic consent and participant

ISSN 2281-5147

engagement platform working towards solving the complexities of consent in genomic research. European Journal of Human Genetics, 29 (4), 687–698.

Hudson, M. (2009). Think globally, act locally: Collective consent and the ethics of knowledge production. International Social Science Journal, 60 (195), 125–133

Joly, Y., Ngueng Feze, I., & Simard, J. (2013). Genetic discrimination and life insurance: A systematic review of the evidence. BMC medicine, 11, 1–15.

Kuru, T., & de Miguel Beriain, I. (2022). Your genetic data is my genetic data: Unveiling another enforcement issue of the gdpr. Computer Law & Security Review, 47, 105752.

Mascalzoni, D., Melotti, R., Pattaro, C., Pramstaller, P. P., G¨ogele, M., De Grandi, A., & Biasiotto, R. (2022). Ten years of dynamic con- sent in the chris study: Informed consent as a dynamic process. European Journal of Human Genetics, 30 (12), 1391–1397.

Niemiec, E., Vears, D. F., Borry, P., & Howard, H. C. (2018). Readability of informed consent forms for whole-exome and whole-genome se- quencing. Journal of Community Genetics, 9 (2), 143–151. https:

//doi.org/10.1007/s12687-017-0324-6

Staunton, C., Slokenberga, S., & Mascalzoni, D. (2019). The GDPR and the research exemption: Considerations on the necessary safe- guards for research biobanks [Number: 8 Publisher: Nature Pub- lishing Group]. European Journal of Human Genetics, 27 (8), 1159–1167. https://doi.org/10.1038/s41431-019-0386-

ISSN 2281-5147

ISSN 2281-5147

TO USE OR NOT TO USE? RE-USING HEALTH DATA IN AI DEVELOPMENT

Fatma Sümeyra Doğan*

Abstract

This study examines the re-use of health data in the context of AI development, focusing on regulatory frameworks governing this practice under the European Health Data Space. It explores how transparency and the protection of personal data are balanced with the need for innovation in healthcare. By analysing real-world examples and the application of General Data Protection Regulation principles, particularly transparency, this study assesses whether health data can be re-used for AI-driven healthcare advancements without undermining individuals' data protection rights.

Table of Contents

* PhD Researcher and Marie Skłodowska-Curie Action's Fellow at Jagiellonian University under the Legality Attentive Data Scientists (LeADS) project. Her PhD research focuses on health data governance in the EU in relation to emerging technologies. Email: av.sumeyradogan@gmail.com

This work is supported by the European Union's funded project Legality Attentive Data Scientists (LeADS) under Grant Agreement no. 956562.

ISSN 2281-5147

Keywords

Secondary Use of Health data – Medical AI – European Health Data Space – GDPR – Transparency

1. Introduction

There is no doubt that healthcare systems could benefit from technological developments to address staff shortages and the increasing number of people who need treatment. AI in healthcare can make a big difference by improving access to treatments and providing more personalized care. It can help doctors diagnose illnesses more accurately, discover new medicines, predict diseases, and obtain support medical professionals by analysing complex data, suggesting treatment options. However, using AI also brings up important issues like protecting people's data and ensuring privacy.

Maintaining the protection of our personal data and privacy carries utmost importance especially in healthcare domain. One of the reasons for this is that we cannot change almost all the data related to our health. For example, our genetic

ISSN 2281-5147

information is fixed from birth and cannot be altered. Similarly, our medical history, such as past surgeries, chronic illnesses and other health conditions we've experienced, is permanent. Unlike passwords or other personal identifiers that can be changed if compromised, these aspects of our health are unchangeable. This makes it crucial to protect this data from misuse or unauthorized access because any breach could have long-lasting or permanent consequences, potentially affecting our privacy, insurance, employment and even the quality of care we receive.

However, at the same time, we must use data to train AI technologies. For example, to develop AI systems that can accurately diagnose diseases, we need access to a vast amount of medical records and imaging data. This data helps the AI learn to identify patterns and make accurate predictions. Similarly, for AI to assist in drug discovery, it must analyse extensive datasets about how different substances interact with the body. Thus, if we want to develop, use and benefit from AI technologies, we must allow the use of data to train these new technologies. To achieve these advancements, the concept of re-using data becomes crucial.

With the increasing potential of AI technologies to in healthcare, the use of health data has become a central concern. While AI can enhance diagnostic accuracy and personalize treatment, its development relies heavily on access to vast amounts of health data. This raises significant legal questions, particularly concerning the re-use of such data beyond its original purpose. In light of the European Health Data Space proposal, which aims to enable the secondary use of health data, this study investigates whether this framework adheres to the General Data Protection Regulation's transparency requirements. Specifically, it examines how transparency can be maintained when data is anonymized or pseudonymized and considers the challenges posed by AI's opacity in data processing.

2. Motivations for Heath Data Re-use

Re-using data refers to the practice of using existing data for purposes beyond its original purposes. For instance, patient records from routine medical visits can be re-used to train AI systems to predict potential outbreaks of contagious diseases such as COVID-19. Additionally, big health datasets can be used to assess post-marketing adverse events and thus the safety of pharmaceutical products. By analysing this data,

ISSN 2281-5147

AI can identify risk factors and help implement preventive measures. For example, AI algorithms can detect patterns in patient records, such as the emergence of symptoms in specific demographics or regions, which may indicate the early stages of an outbreak. The AI can then model potential scenarios, allowing healthcare providers to respond proactively by increasing resources in high-risk areas or recommending targeted interventions like vaccination campaigns or public health advisories. This data-driven approach enables more effective prevention and management of health issues before they become widespread. Additionally, data from fitness trackers and wearable devices can be re-used to enhance AI algorithms that promote healthier lifestyles by providing personalized recommendations on exercise and diet. This re-use of data is essential for advancing AI technologies and unlocking their full potential to benefit society. It allows us to maximize the value of existing data while ensuring that new insights and innovations can be achieved without repeatedly collecting the same information.

In this study, the terms 'secondary use' and 're-use' of data are used interchangeably, as both concepts involve utilizing existing data for new purposes to derive additional value and insights.

2.1 Real World Examples of Data Re-use

In order to provide more concrete examples of the re-use of data in the health sector, we will give a few of them in the following. Google's Automated Retinal Disease Assessment harnesses artificial intelligence to aid healthcare practitioners in detecting diabetic retinopathy, a condition where high blood sugar levels damage the blood vessels in the retina, potentially leading to blindness if left untreated. This technology also has the potential for AI algorithms to further assist clinicians in recognizing other medical conditions (*ARDA*, n.d.). Google collaborated with Moorfields Eye Hospital located in the UK to assemble a dataset of eye retina images. Subsequently, Google Health trained an artificial intelligence system capable of predicting the development of a type of eye disease. A study was conducted to evaluate this system against expert clinicians. The findings suggest that Google's AI system can forecast whether an eye may develop the disease within the next six months as accurately as clinicians (*Using AI to Predict Retinal Disease Progression*, 2020). Additionally, Google explored potential clinical uses of this system, showcasing the promise of AI in preventive medical studies. According to Google this technology now has been used widely in India and

ISSN 2281-5147

practitioners reported that it enables them to examine more patients in a day, which is crucial in over-populated areas (*Dostrzeganie Potencjału - Google*, n.d.).

Moreover, Google has improved this technology and developed an innovative method to predict the risk of a heart attack by analysing images of a person's retina. This advancement also utilizes artificial intelligence to scan the eye and identify risk factors for cardiovascular diseases. By examining the retinal blood vessels, the AI system can accurately predict the likelihood of a heart attack, offering a non-invasive and efficient way to assess heart health, although this method is not yet widely used in clinical practice (Poplin et al., 2018). However, its innovative approach could complement traditional methods and potentially become more common as the technology advances and gains broader acceptance. This method, developed in collaboration with various research institutions, highlights the potential of AI in preventive healthcare. By leveraging retinal images, which are relatively easy to obtain, this approach could revolutionize how heart disease risks are assessed, potentially leading to earlier interventions and better health outcomes. The significance of this development lies in its ability to provide a quick, non-invasive diagnostic tool that can be used widely, especially in settings where traditional methods might be impractical.

On a different study, Altsman and his team at Stanford University utilized statistical analysis and data mining techniques to detect patterns in extensive datasets. They developed a "symptomatic footprint" for drugs that could cause diabetes (Yousefi, 2022, p. 4). By partnering with Microsoft Research, they examined user's anonymous Microsoft search engine logs. Through this investigation, they discovered that combining the drugs named Paxil and Pravachol can lead to diabetes. This conclusion was drawn from the observation that patients taking these two drugs together exhibited a notable increase in searches for terms associated with diabetes, such as "fatigue" and "loss of appetite," indicating high blood glucose levels. This data-driven research provided a crucial, life-saving finding that traditional methods might have missed. Altsman asserts that restricting access to data would be detrimental to research, as data is a vital source of inspiration, innovation, and discovery in medicine. He believes that the ability to analyse new data sources offers unprecedented opportunities to identify problematic drugs or drug combinations much earlier than previously possible (Stanford University, 2016).

ISSN 2281-5147

3. Rules and Regulations to Process the Health Data

Data is a key element in training AI technologies, making it crucial to obtain. This raises the question: how can innovators legally access data? To answer this, we must examineregulations, such as those in the European Union, which will be the focus due to the limitations of this study. (Artice 3(1) of the Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation), 2016)

3.1 The General Data Protection Regulation

The General Data Protection Regulation is a law from the European Union designed to protect personal information. It ensures that companies handle data with care and respect. Under the General Data Protection Regulation, individuals have the right to know what data companies have about them, to correct any inaccuracies and to request the deletion of their data if it is no longer needed. Companies must obtain consent before using personal health data and are required to keep it safe from data breach and misuse. They must also be transparent about why they are collecting data and how it will be used. Companies that fail to follow these rules can face significant fines. In essence, the General Data Protection Regulation gives individuals control over their personal data and ensures it is protected.

The General Data Protection Regulation includes special rules for protecting health data because this type of information is highly sensitive. Health data includes medical records, genetic information and any details about an individual's physical or mental health. Because of the sensitive nature of this data, General Data Protection Regulation imposes stricter rules to ensure it is handled with the highest level of care. Under the General Data Protection Regulation, as a general rule, companies and organizations must obtain explicit consent from individuals before collecting or using their health data. This means they must clearly explain why the data is needed and how it will be used and individuals must agree to it. Additionally, health data must be kept secure to prevent unauthorized access or misuse. This includes using advanced security measures like encryption. Encryption is the process of converting data into a coded format that can only be accessed by authorized individuals, ensuring that even

ISSN 2281-5147

if the data is intercepted, it cannot be read by unauthorized parties. The stricter rules are in place to protect individuals' privacy and to prevent any potential harm such as identity theft, discrimination or unauthorized use of personal health information, that could come from the misuse of sensitive health information. By mandating stringent measures for handling health data, the General Data Protection Regulation aims to safeguard personal health information and build trust in how it is managed.

3.2 A New Regulation to Govern the Health Data

The strict protection rules imposed by General Data Protection Regulation on health data have inadvertently inhibited the development of AI technologies by restricting access to the data needed to train and improve these systems. To address this challenge, proposal for the European Health Data Space was introduced to improve healthcare quality and continuity across Europe (European Commission, 2022). Another key reason is to accelerate medical research and innovation. With access to larger and more diverse datasets, researchers can conduct more comprehensive studies, leading to quicker and more significant medical advancements. This collaborative approach, supported by the European Health Data Space, will establish clear rules and frameworks that facilitate the secure sharing of health data across borders and between institutions. This will help in developing new treatments, understanding diseases better and improving public health outcomes. The European Health Data Space also aims to empower individuals by giving them more control over their health data. During the discussion phase at the EU Parliament, an opt-out mechanism was proposed, allowing individuals to choose not to participate in the secondary use of their health data, ensuring that their consent remains central to the process. (*European Health Data Space*, n.d.)By making it easier for people to access and manage their medical records, the European Health Data Space aims to promote modernisation in the healthcare systems of the European Union.

3.3 Secondary Use under the European Health Data Space

The secondary use framework under the European Health Data Space is designed to facilitate the re-use of health data for purposes beyond the initial care of patients, while ensuring robust data protection and privacy. Once the proposal fully adopted by the Member States, health data will be collected from various sources, such as hospitals, clinics and wearable devices. Before this data can be used for secondary

ISSN 2281-5147

purposes, it undergoes a process of anonymization or pseudonymization to enhance confidentiality. Anonymization involves removing all personal identifiers so that individuals cannot be identified, while pseudonymization masks identifiers by replacing them with codes (pseudonyms) that can only be re-linked to the original data under strict conditions.

Strict protocols will be established to determine who can access the data and for what purposes. Researchers must apply for access through a regulated process, which is overseen by independent authorities known as health data access bodies. These bodies are established under the governance frameworks of the European Health Data Space, ensuring their operation is guided by strict legal and ethical standards. Composed of experts in data protection and relevant scientific fields, these bodies are independent from research institutions which allows them to impartially evaluate applications. This board assesses the potential benefits of the research or project against privacy risks, as well as other potential risks such as data misuse, ensuring that data is used responsibly. Approved users access the data through secure environments, often referred to as data safes or data access platforms, which implement advanced security measures such as encryption, secure login and monitoring to prevent unauthorized access or data breaches. The European Health Data Space establishes comprehensive governance frameworks that outline the responsibilities of all parties involved in data handling and usage.

The framework promotes transparency by requiring public disclosure of who is using the data, for what purposes and the outcomes of their research or projects. Accountability measures are in place to address any misuse of data, including penalties and corrective actions, such as revoking access to data or imposing fines. By enabling the safe and legal re-use of health data, the European Health Data Space aims to drive innovation in healthcare, support the development of new treatments and technologies and enhance public health strategies. This approach maximizes the value of existing data while maintaining high standards of data protection and privacy, ensuring that the benefits of data re-use are realized without compromising individual rights.

ISSN 2281-5147

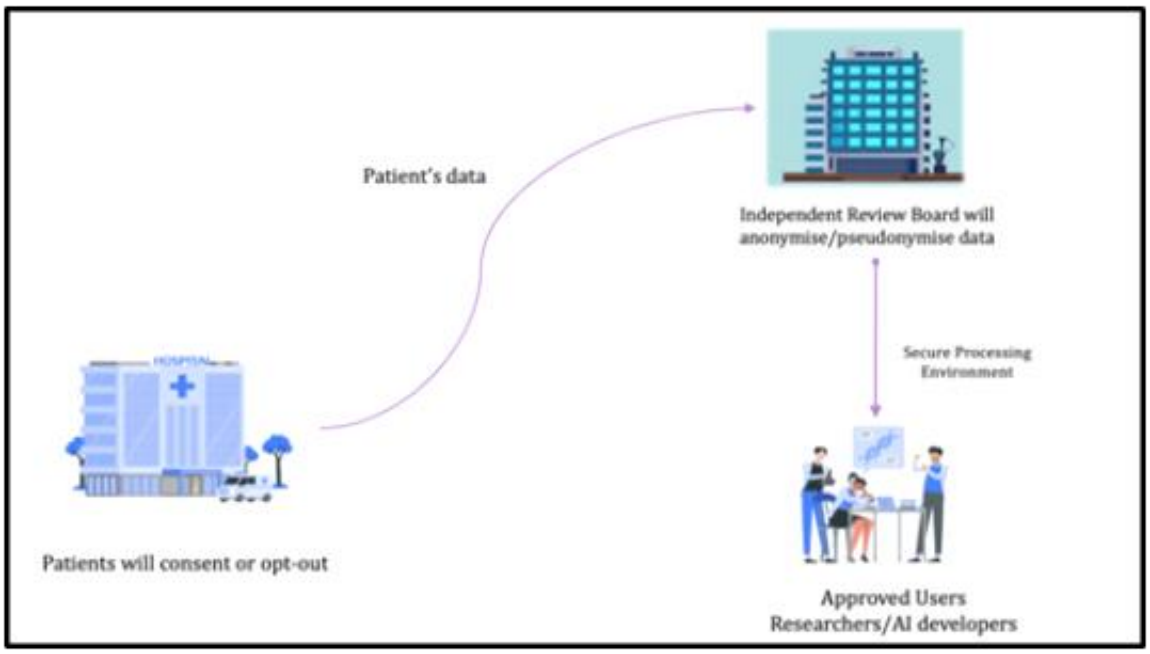


Figure 1- An example for the re-use of health data under the European Health Data Space proposal

4. Data transparency challenges for Health Data re-use

Understanding transparency in data processing is crucial as it forms the backbone of responsible data management practices. Under General Data Protection Regulation, transparency means that data controllers must provide clear, accessible and comprehensive information to individuals about how their personal data is collected, used and shared. This principle is particularly relevant when considering the legal and ethical considerations of the secondary use framework under the European Health Data Space. As the European Health Data Space enables the re-use of health data for research and other secondary purposes, it must comply with stringent transparency requirements mandated by the General Data Protection Regulation. Specifically, Article 14 of the General Data Protection Regulation sets forth detailed obligations for data controllers to inform individuals about the collection and use of their data, when the data is obtained indirectly. This requirement ensures that individuals are fully aware of how their health information is being utilized and protected within the European Health Data Space framework, thus maintaining trust and upholding privacy standards.

ISSN 2281-5147

4.1 Transparency under General Data Protection Regulation

Transparency in data processing is a fundamental principle aimed at ensuring individuals understand how their personal data is collected, used and protected. This principle is essential for building trust between data subjects, who are the individuals whose data is being processed, and data controllers, which are the entities that determine the purposes and means of processing personal data.

Data controllers must communicate clearly and openly about their data processing activities. This involves informing individuals about what data is being collected, the purposes for which it is collected, how it will be used and who will have access to it. Transparency also requires explaining the legal basis for data processing, which can include consent - especially critical for health data- from the data subject, the necessity for the performance of a contract, compliance with a legal obligation, protection of vital interests, public interest or legitimate interests pursued by the data controller.

Additionally, individuals must be informed about their rights regarding their personal data. These rights include access to their data, correction of inaccuracies, deletion of data (known as the right to be forgotten), restriction of processing and data portability, which is the right to receive their personal data in a structured and commonly used format and to transfer that data to another data controller. Transparency also encompasses information about how long personal data will be retained and the security measures in place to protect it. This ensures that individuals are aware of the lifespan of their data and the steps taken to safeguard it against breaches and unauthorized access.

When personal data will be shared with third parties, data controllers must disclose this information, specifying who the third parties are, the purpose of sharing and how the data will be protected in the process. Providing contact information for the data protection officer or another responsible entity is crucial, allowing individuals to seek further information or lodge complaints about data processing activities.

If data processing involves automated decision-making, including profiling - where personal data is used to evaluate certain aspects of an individual, such as their behaviour, preferences or health-, they must be informed about this aspect. They should understand the logic involved, the significance and the potential consequences of such processing. By ensuring these aspects of transparency, data controllers help

ISSN 2281-5147

individuals make informed decisions about their personal data and exercise their rights effectively. Thus, the transparency principle fosters trust, accountability and compliance with General Data Protection Regulation.

4.2 Transparency vs. Secondary Use of European Health Data Space

The balance between transparency and secondary use in data processing is a crucial and often challenging aspect of today's data management practices. As mentioned, transparency requires that data controllers clearly inform individuals about how their data is collected, used and shared. Secondary use, on the other hand, involves using data for purposes beyond the original context in which it was collected, such as for research or analytics, which can be distinct from the initial data collection purpose.

This trade-off arises because achieving high levels of transparency often necessitates detailed disclosures about data use, which can sometimes conflict with the need to manage secondary use. This challenge becomes even more pronounced when data users have profit-driven motives. For instance, consider a company that collects health data from wearable devices for research purposes. Initially, the company informs users that their data will be used to improve health monitoring technology and to conduct general health studies. This initial disclosure is straightforward and focuses on the primary purpose of data collection. However, the company's long-term plan involves using this data to develop targeted marketing strategies for health-related products and services, such as personalized dietary supplements or fitness programs. To maximize profits, the company might not fully disclose these secondary intentions to users at the time of data collection. By keeping these plans less transparent, the company can more easily obtain consent from users who may otherwise be hesitant if they knew their data would be used for targeted marketing. In this context, the opt-out mechanism provided by the European Health Data Space offers individuals the ability to refuse secondary uses of their data.

The overarching question this study seeks to address is: For the sake of innovation, should we give up on protecting our personal data and privacy? The European Health Data Space regulation proposal will provide access to health-related data through its secondary use of data framework. It also includes several safeguards, such as anonymization and pseudonymization. Nonetheless, it remains unclear how the transparency principles of the General Data Protection Regulation will be adhered to

ISSN 2281-5147

within this new framework. This study aims to investigate whether the secondary use schema of the European Health Data Space proposal aligns with General Data Protection Regulation's transparency rules.

As discussed, this goal does not have a straightforward answer. Transparency is a multifaceted concept and ensuring it when data is processed in anonymized or pseudonymized forms presents particular difficulties. For example, it is challenging to maintain transparency about data use when the data itself has been altered to remove personal identifiers. Additionally, AI technologies are known for their opacity in how they process and analyze data, making it difficult to fully understand and communicate how data is being used. Developers benefiting from secondary use of health data sould be more transparent about their purposes for using the data.

4.3 Guiding principles for transparency

To address these transparency challenges, we can draw insights from the U.S. Food and Drug Administration, an authoritative body that regulates medical devices and their software, has established guiding principles for transparency in machine learning-enabled medical devices. These principles highlight the need for clear communication about how devices are used, including their intended purpose, development, performance and the underlying logic of their algorithms. Drawing from the Food and Drug Administration's recommendations, it is clear that effective transparency involves not only providing relevant information about a device's functionality and performance but also ensuring that such information is accessible, timely and comprehensible to users.('Transparency for Machine Learning-Enabled Medical Devices', 2024)

Incorporating these principles into the European Health Data Space framework could address some of the transparency challenges. For example, clear and ongoing communication about how health data is used and how AI systems make decisions could help bridge the gap between data anonymization and user understanding. Policymakers and developers should consider adopting strategies similar to those outlined by the Food and Drug Administration, such as enhancing user interfaces to present information more clearly, providing timely updates, and using human-centered design principles to make data use more transparent.

ISSN 2281-5147

Ultimately, this study seeks to determine whether a balance can be struck between leveraging health data for technological advancements and maintaining stringent transparency and data protection standards as mandated by the General Data Protection Regulation. By aligning with best practices in transparency, as suggested by institutions like the FDA, we can better safeguard individual rights while unlocking the full potential of health data for technological progress.

5. Conclusion

In summary, the European Health Data Space represents a significant step forward in the responsible use and protection of health data. We have explored the balance between transparency and anonymity, highlighting the importance of clear communication and robust anonymization techniques to protect individual privacy while enabling valuable medical research and technological advancements. Key legal and ethical considerations, such as adherence to the General Data Protection Regulation and the transparency requirements of it, ensure that data processing within the European Health Data Space framework is conducted legally.

Looking to the future, the potential of the European Health Data Space to revolutionize healthcare is immense. By facilitating the secure and legal re-use of health data, the European Health Data Space can drive innovations in personalized medicine, improve public health strategies and support the development of new treatments and technologies. However, challenges remain, particularly in maintaining the delicate balance between transparency and privacy and ensuring that data re-use does not compromise individual rights.

As the landscape of health data use continues to evolve, it is crucial for individuals to stay informed about their data rights and the measures in place to protect their privacy. By understanding the principles of transparency and the importance of data protection, we can promote a more informed and engaged public that supports the re-use of health data. This collective awareness will help ensure that the benefits of the European Health Data Space are realized while safeguarding the privacy and data protection rights.

ISSN 2281-5147

6. Bibliography

ARDA: Using Artificial Intelligence in Ophthalmology - Google Health. (n.d.). Retrieved 11 May 2024, from https://health.google/caregivers/arda/

Dostrzeganie potencjału—Google. (n.d.). Retrieved 11 May 2024, from https://about.google/intl/ALL_pl/stories/seeingpotential/

European Commission. (2022, March 5). *Communication from The Commission to The European Parliament and The Council A European Health Data Space: Harnessing the power of health data for people, patients and innovation*. https://doi.org/10.1163/2210-7975_HRD-4679-0058

European Health Data Space: Council and Parliament strike deal. (n.d.). Retrieved 17 April 2024, from https://www.consilium.europa.eu/en/press/press-releases/2024/03/15/european-health-data-space-council-and-parliament-strike-provisional-deal/

Poplin, R., Varadarajan, A. V., Blumer, K., Liu, Y., McConnell, M. V., Corrado, G. S., Peng, L., & Webster, D. R. (2018). Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nature Biomedical Engineering*, *2*(3), 158–164. https://doi.org/10.1038/s41551-018-0195-0

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the Protection of Natural Persons with Regard to the Processing of Personal Data and on the Free Movement of Such Data, and Repealing Directive 95/46/EC (General Data Protection Regulation), EP, CONSIL (2016). http://data.europa.eu/eli/reg/2016/679/oj/eng

Stanford University. (2016, March 30). *Harnessing big data to better understand what happens when we mix drugs*. Welcome to Bio-X. https://biox.stanford.edu/highlight/harnessing-big-data-better-understand-what-happens-when-we-mix-drugs

Transparency for Machine Learning-Enabled Medical Devices: Guiding Principles. (2024). *FDA*. https://www.fda.gov/medical-devices/software-medical-device-samd/transparency-machine-learning-enabled-medical-devices-guiding-principles

ISSN 2281-5147

Using AI to predict retinal disease progression. (2020, May 18). Google DeepMind. https://deepmind.google/discover/blog/using-ai-to-predict-retinal-disease-progression/

Yousefi, Y. (2022). Data Sharing as a Debiasing Measure for AI Systems in Healthcare: New Legal Basis. *Proceedings of the 15th International Conference on Theory and Practice of Electronic Governance*, 50–58. https://doi.org/10.1145/3560107.3560116

ISSN 2281-5147

ISSN 2281-5147

HOW TO COLLABORATIVELY USE STATISTICAL MODELS IN A SECURE WAY

Maciej Krzysztof Zuziak*

Abstract

The following articles compile research on the brink of privacy, federated learning and data governance to provide a reader with a basic understanding of the nuanced world of decentralised learning systems. It starts from simple notions of personal data and its connection to artifical intelligence. Afterwards, it goes into the realm of statistical learning to explain the basic technocratic lingo in a (hopefully) engaging way. With those topics covered, it proceeds to deliver on the basic notion of Data Collaborative and Decentralised Data Governance - an arcane term that the reader will be familiar with at the end of this lecture. Finally - it poses some open-ended remarks on the future of data analysis done in a way that benefits our communities. While the delivery of the article is relatively simple and straightforward, it also provides the curious reader with links and pointers that would allow them to go deeper into a well of data governance and large AI infrastructure.

Table of Contents

* Maciej joined LeADS in November of 2021 to work on the topic of privacy-enhanced Machine Learning and Personal Data Management at the Institute of Information Science and Technologies "Alessandro Faedo" at the National Research Council of Italy. He deals with the topics of Decentralised Machine Learning and Alternative Data Governance. During his three-years research at the Institute, he published a number of articles on both of those topics on a reputable venues such as IEEE Big Data or ACM FaaCT.
maciejkrzysztof.zuziak@isti.cnr.it, maciej.k.zuziak@protonmail.com
This work is supported by the European Union's funded project Legality Attentive Data Scientists (LeADS) under Grant Agreement no. 956562.

ISSN 2281-5147

Keywords

Federated Learning – Machine Learning – Data Governance – Alternative Data Governance – Data Collaboratives

1. Introduction to a world of big-data

The advent of highly capable intelligence systems is evident to everybody. From talk shows to popular news outlets, terms like *Artificial Intelligence*, *Machine Learning,* and *Data-Driven Economy* are being constantly called out and discussed, with little to no explanation of what is actually hiding behind them. Sometimes, you can hear that those systems *consume* (or perhaps – *require*) a tremendous amount of data to be *trained* – but what does this mean in practice? If you are one of those who encountered those terms in the wild – you may ask yourself a question: *but why should it actually concern me?* Suppose you have pondered on that matter a little bit more. In that case, you might actually rephrase the question a little into the following: *but what is the relationship between my personal information and the behaviour of those systems?*

ISSN 2281-5147

If you had indeed asked yourself one of those (or similar) questions, this text is addressed to dispel your doubts or instil new ones. If you haven't asked those questions yet, but the opening paragraph has sparked your interest, then there will be no better time to start seeking answers to those issues and this text may be a good beginning of that journey. I must disclose that this text is written clearly and succinctly and is addressed to a reader who wants to satisfy their curiosity rather than to an expert who has already spent a fair number of hours (perhaps weeks or maybe years) researching presented topics as it will simplify many steps and definitions to arrive at simple explanations of rather complex things. In Feynman's book titled QED: The Strange Theory of Light and Matter[1], the author states that explaining complex concepts in plain words is an art itself, and – I would like to add after him - as in art, one can fail miserably at it. However, I did my best to make this will digestible for someone who did not touch the matter presented in it previously or perhaps burnt their hand while trying to touch it, as it requires knowledge from a fair number of disciplines to comprehend it fully (and I do not pose myself as one of those who comprehended it. I somewhat believe that the most honest response would be to say that we all struggle to understand it, and the history of that struggle is what we present the world with).

2. Personal Data and Statistical Inference

It may be best to start the journey from the concept of *personal data*. For the sake of simplicity, let's assume that *personal data* is any information that relates to you as an identifiable living individual[2]. The GPS routes saved in your telephone, the photos with nametags on your iPhone, the health records your local doctor keeps…all this can constitute *personal data*. Of course, the reality may be more complex, but for the sake of this text (and for the sake of all other debates about *personal data* that you stumble across from time to time), it is safe to assume that this simplified definition is all you need (and indeed, this will not be that far from the truth).

[1] (Richard P. Feynman & Anthony Zee ([introduction], 2024)

[2] The fairly simple explanation of the term personal data provided on the European Commission website states that *Personal data is any information that relates to an identified or identifiable living individual. Different pieces of information, which collected together can lead to the identification of a particular person, also constitute personal data*. While the concept of personal data is more nuanced in practice, this definition is a fundamental building block for all more subtle interpretation regarding personal data (European Commission, 2024).

ISSN 2281-5147

Now, let's focus on the concept of *statistical model*. According to Wikipedia – which closely follows D.R. Cox's book on Principles of Statistical Inference (Cox, 2006) – *a statistical model is a mathematical model that embodies a set of statistical assumptions concerning the generation of sample data (and similar data from a larger population). A statistical model represents, often in a considerably idealised form, the data-generating process*. However, this definition may bear little to no meaning to a person not acquainted with statistics (and the one who is acquainted with statistics probably would not require this explanation in the first place), so let's take a step back. Most of us would probably agree that the number of ways a specific phenomenon can be measured is limited. Political polls generally survey only a limited number of participants since surveying all of those entitled to cast a vote would be nearly impossible in the first place. Explaining consumers' trends and turnovers operates on (limited) historical data and is constrained to events that can be (effectively) recorded. Drawing conclusions about the height and size of a specific type of penguin is based on the measurements of penguins in a limited sample since catching and measuring all of them would be considered highly impractical – and so on. It is obvious that while making conclusions about a population at large, the best we can often do is to take a minimal glance over the window of our own reality. On an intrinsic and highly intuitive level – the statistical models are the windows through which we can deduct information about the reality surrounding us. The models themselves are wooden frames, while the model's parameters are the frame's dimensions. Once we choose a frame that has caught our attention (and one which we believe will give us a good outlook on the world around us), we must tune the width and height of that frame – so that it captures exactly what we want it to. *Training a statistical model* means no more than adjusting those parameters with the help of prior information we have obtained.

3. Training Your First Statistical Model

Let's take a look at an illustrative example. Let us assume you want to identify patients with a high risk of a particular genetic disease. Due to a high number of patients and a complicated diagnosis process, you want to automate this process to foster the preventive measures that can prevent the disease's development if applied early. Then, you would need to develop a classification model, i.e., classifier – a hypothesis

ISSN 2281-5147

function that will return 'true' if a specific genetic marker[3] is associated with a higher risk of disease occurrence and 'false' otherwise. Assume that the *hypothesis function* is in the form of a black box. It is hardly a metaphor, just imagine a black steel box arriving at your desk. The black box has two modes: *train* and *infer*. When the black box is in *train* mode, it tries to distinguish between the markers potentially associated with a higher chance of occurrence and those risk-free. The first step would be to switch our black box to train mode and teach it the distinction. As with children, the simplified learning process relies on correlating the features of the object with a corresponding label. While human beings' generalising capabilities give us a remarkable ability to learn a pattern by observing just a few data instances, machines tend to require more examples to familiarise themselves with a specific pattern. On the other hand, their main strength is strictly connected to the ability to detect much more complex patterns that we could potentially detect, making them a perfect tool for identification based on genetic markers.[4]

[3] A genetic marker is a gene with known location on a chromosome that can be used to infer properties about the individual or species. A beautiful comparison of a genetic marker to a *landmark* is presented on the site of National Human Genome Research Institute. In this narration, a genetic marker can be compared to a marker (characteristic location) that can help you navigate through a city that you are not yet familiar with (National Human Genome Research Institute (NIH), 2024)

[4] The human learning process is – of course – more nuanced than that and relies on many different skills developed throughout our upbringing. Interestingly enough, the machines also benefit from learning some fundamental knowledge about the domain. A fairly accessible article by M.G. Levy exploring this topic can be found under the link: https://www.quantamagazine.org/machines-learn-better-if-we-teach-them-the-basics-20230201/ (Levy, 2023).

ISSN 2281-5147

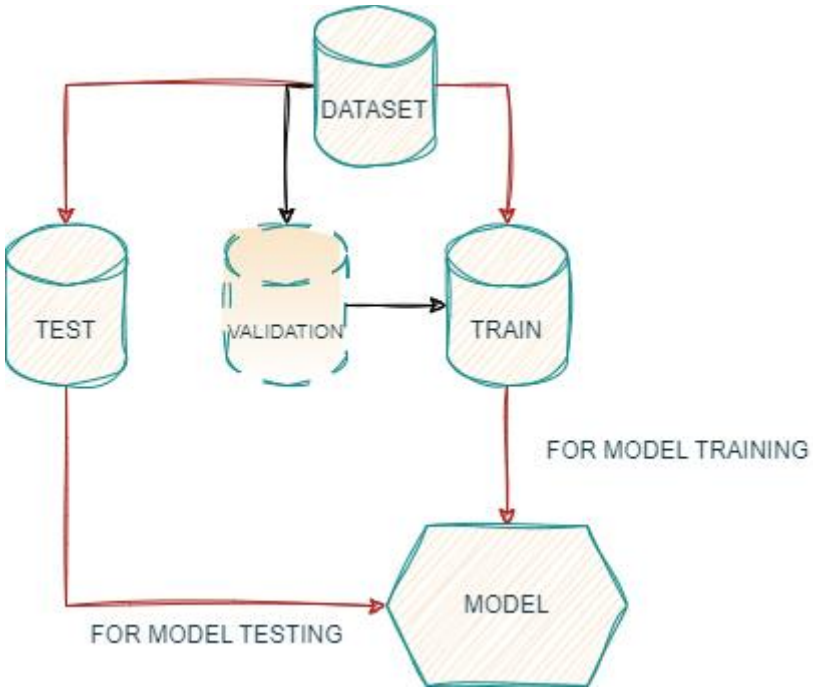


Fig.1: Before the training, we split the dataset into two sets: one for training purposes and one for testing purposes. Sometimes we also distinguish a third set for validation purposes (checking the progress of the training without using the external test part that is reserved only for a final evaluation).

However, to begin the training, we will require a number of pre-classified genetic markers with corresponding labels ('risk-associated and 'risk-free'). The sample of pre-classified markers will constitute our *training set*. Another sample of pre-classified objects is necessary to evaluate the general performance of our model – this will be called our *test set*. The distinction between training and testing datasets is crucial, as we do not want to evaluate the performance of our model based on the same things that we trained it on. Using education as an example, teachers and professors seldom provide students with answers before the test.

ISSN 2281-5147

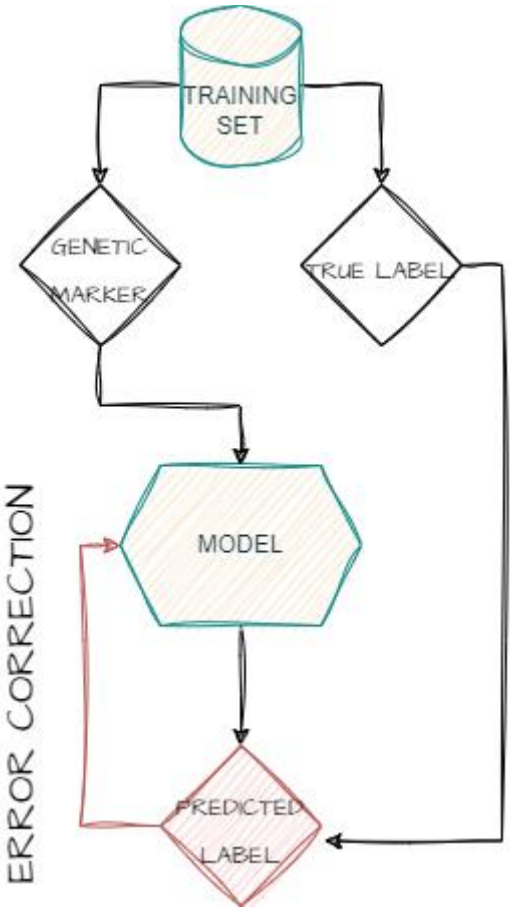


Fig.2: Training phase. The model associates genetic markers with corresponding label. The difference between the predicted and true label is used to correct the model.

The training will consist of putting a pre-classified marker within a black box, scanning, and then picking up another data sample. Depending on the task complexity, it may require a few hundred to a few hundred thousand different data points to be deployed. After the training, when you switch the black-boxin to *infer* mode, it will allow you to make an automatic classification of the data sample. You do that by placing a marker inside the device and then pressing the big red button on the top of the black box. It either returns 'true' or 'false' as mentioned above, where *true* value would imply that the patient is at risk of occurrence of a particular genetic disease. Now, let us imagine that the black box has memory cards that are used to distinguish between markers. Those memory cards preserve the information about what types of markers are associated with a high risk of occurrence. The knowledge is based on markers that they have seen so far. In the machine learning lingo, the memory cards are called *parameters*, while the black box is often called *architecture*. When we use the word *model*, it often implies a specific black-box architecture with a pre-determined set of *parameters*. The

ISSN 2281-5147

markers we use to *train* the black box are part of our *sample*, while the black box will be used to make predictions about the *population*.

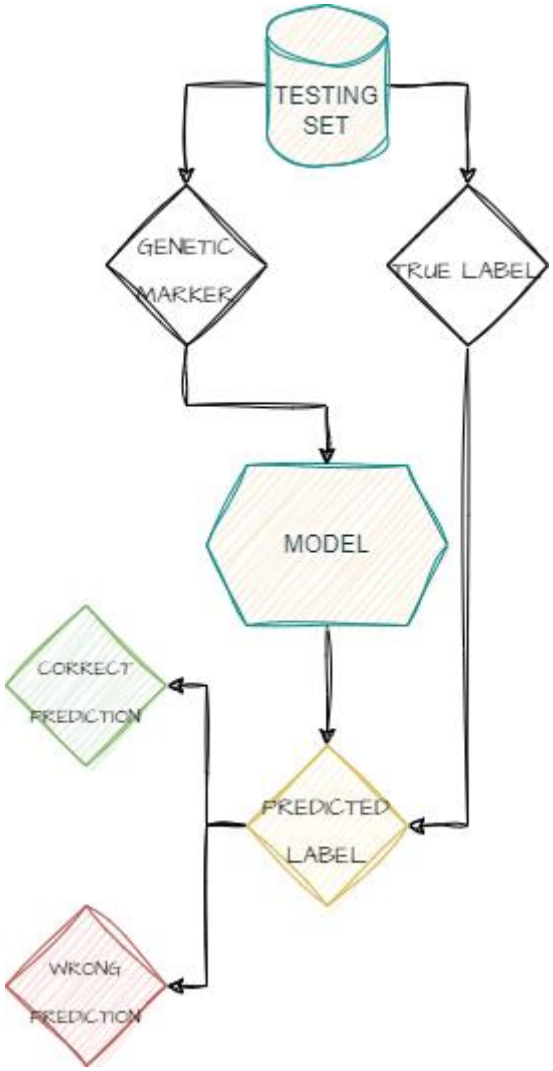


Fig.3: Testing phase of the model. This time, we count the correct and wrong predictions to quantify how well our model performs on unseen data. In contrast to the training phase, this time, we do not use the difference between the true and predicted labels to correct our model.

4. Between Statistical Inference and Personal Data

As can be understood from the last paragraphs, training the model requires some amount of pre-classified data that can be used as a basis for future pattern recognition. The source of this data depends on the particular task we aim to solve. In many

ISSN 2281-5147

instances, the data will be registered during carefully designed experiments. In other instances, the samples will be provided by us directly, as is in the case of digital marketing, where our commercial activities are the basis of sentiment analysis and consumer-type classification. I want to focus on the second case, as it involved the previously introduced concept of personal data.

Genetic markers may – or may not – constitute personal data. Suppose the genetic marker can be uniquely attributed to an individual person. In that case, it is considered a special category of personal data that is guarded by a higher regime of legal protection under the General Data Protection Regulation. As with many cases of human-related data, the final answer to whether it constitutes personal data will be highly context-specific and based mainly on the notion of identifiability (whether a specific sample can or cannot be attributed to a particular person).

There is a catch – acquiring (any kind, not only personal) data is expensive. It requires a lot of infrastructure. Firstly, to collect it, then to store it – not even mentioning any methods of processing it to obtain any meaningful information of our interest. Large technological companies have a natural advantage here. They provide and utilise a large-scale architecture that is mainly based on interconnected social networks – while they still need to abide by the consent-based regulatory framework that we've described earlier, the staggering scale of their enterprises allows them easily to obtain the consent of their users, as those users are often engaging with their services daily (think about online marketplace, video-on-demand services, social networks providers and other entities that monetise heavy user-flow of their platforms). Of course, public infrastructure and individual citizens seldom can deploy at such a scale. It naturally explains the benefit of the size, which can make or break in terms of becoming a champion of the digital race.

5. And The Weak Suffer What They Must?

The pioneers of the digital race, the same that utilise the large-scale infrastructure mentioned earlier, do that in their own name and for their own benefit. It should hardly come as a surprise since those entities are enterprises that main aim is to generate revenue and their investment in said infrastructure was calculated as a long-term investment. How they benefit from it may depend on the particular case

ISSN 2281-5147

scenario. Netflix – subscription video-on-demand service – uses data analysis to train recommendation models that can deliver their clients a suitable set of new titles to watch every time they visit the main homepage (which, in turn – can allow them to retain their customers by engaging them in yet another series). Amazon – a multinational technology company – uses a stream of data from various services they offer to create multi-modal systems predicting behaviours and trends of their customer base. Since the time the media has expressed enormous interest in the Large Language Models (LLM) – models that are capable of (among other) language generation – a number of companies have tried to jump on the bandwagon with their products. Because those models are trained on vast amounts of text and require enormous monetary expenditure – the race was briskly overtaken by industrial champions, with little space left for smaller entities.

You may ask – what about the individual users? Can they also benefit from the blessings of big data advent? Well, that is the question that I try to answer with my research. Undeniably, an individual has a very limited ability to acquire the data, infrastructure and workforce necessary to achieve such a task. There exist open source pre-trained models that are available on the market. Pre-train – in that context – means that most of the hard work has already been carried out, and you only need to adjust the model to your needs. This sounds tempting, but like always – there is a catch. Firstly, those models may not suit our particular needs. In drastic simplification, a model (and an architecture) that was reserved for an image classification task (let's say – distinguishing between cars and pedestrians) would not be a viable choice for a language translation – and vice versa. Secondly, the 'pre-trained' does not mean 'fully-trained'. For the model to suit your desired purpose (whether it will be digital marketing, medical research or language generation) – you will still need to have access to some (possibly personal) data that will be used in the course of pre-training the model. Hence, we have come a full circle.

6. Benefits Through the Collaboration

When individual efforts do not yield a significant result, it is a natural behaviour to look upon the concept of collaboration. In fact, some scholars have noticed that data-driven endeavours do not have a monopolistic nature by default. In fact, they may be

ISSN 2281-5147

more suited towards collaborative effort than they seem at first glance. This observation has led to the development of several concepts, such as *Data Trusts, Data Collaboratives or Data Cooperatives* (Mozilla Insights et al., 2020). While they all come with some nuanced differences, I do not want to dwell on that topic too much. However, I would like to refer an interested reader to the studies on the difference between collaboration-based concepts performed by Mozilla Insights together with Jonathan von Geuns and Ana Bradusescu since their findings are fully open to the public[5]. What matters from the perspective of this essay is the basic understanding of the *Data Collaborative* concept as defined by our research.

In the simplest terms, the Data Collaborative is a common undertaking of independent actors to train a shared model (or a number of shared models) throughout the lifecycle of the collaboration. The cornerstone of each Data Collaborative is its ability to train one statistical model by a number of participants collectively. For example, let's assume that we have five hospitals in Tuscany. All of them possess highly specific data about a number of patients with pulmonological diseases of a certain type. Let's further assume that those diseases can be identified based on X-ray imaging. However, their occurrence is somewhat region-specific, and the symptoms visible on the X-ray image will vary from region to region. Since the disease is highly problematic, those five hospitals want to train a model for its early detection to aid doctors in analysing the X-ray imagining. Seldom will it be the case that those hospitals are able to use some pre-trained model since such a study could not have been conducted yet or the results of such a study (hence, access to a model) are not publicly available. Neither often is it a case that those hospitals can train one model individually. Hence, they create a structure, a *Data Collaborative*.

The *Data Collaborative* is a very wide definition of every structure that fulfils some pre-defined criteria and allows for shared model training. In our article, *Data Collaboratives with the Use of Decentralised Learning,* we have developed four fundamental principles that could characterise such a structure (Zuziak et al., 2023).[6] *Firstly, the data collaboratives should provide an accessible infrastructure for performing various analytical tasks without the necessity to transfer raw data beyond the participants' devices [**Decentralised Data Storage**]*. It means that the hospitals will not make their x-ray images public due to

[5] Link to the studies form September 2020: https://assets.mofoprod.net/network/documents/ShiftingPower.pdf.
[6] Link to the article from June 2023: https://dl.acm.org/doi/10.1145/3593013.3594029

ISSN 2281-5147

privacy reasons, nor will they create a shared data storage that will hold all of their data in one place. It means that the very data that is necessary to train the model will not be transferred beyond the (local) datacentres belonging to the hospitals. You could now be asking: *How is that even possible? Isn't the data absolutely necessary to train the model?* Indeed, it is a perfectly reasonable question that I will try to answer soon. *Once the model is trained (or once an analytical task is accomplished), it should be governed by all the members of the collaborative (in proportion to their marginal contribution). Shared governance is a key guarantee that all the members will benefit from joint participation in the analytical tasks* by collectively making decisions about the future of the model. *Collective-choice arrangements could be realised by allowing participants to collaborate to create their own rules and governance conditions. [**Shared Model Governance**].* Since the model is trained by a consortium of hospitals, it will be governed by the joint board of their representatives. It guarantees that no member of the collective can, without prior consent from others, remove, modify or make the model publicly available. Since all of the hospitals are equally participating in the training, using their own data and incurring infrastructure costs, they must have some common way of making a decision regarding the model governance. In terms of hospitals, it can manifest in a set of rules regarding how the model can be made public and under what circumstances (for example, after unanimous consent of all the members of the collaborative). *The structure of a data collaborative should be mostly implementation-agnostic. This is because any structure or implementation that satisfies the baseline definition and the four essential principles can be treated as a data collaborative - irrespective of the implementation details [**Universality**].* This is mostly due to the fact that such a partnership can be implemented with multiple technologies that are available. While – in the next section – this article overviews only one of them (namely, Federated Learning). By no means can the use of other methods preclude someone from using the term *Data Collaborative*. Since the pool of available technologies is evolving all the time, this ensures that the concept will stay universal long after its first presentation. *Finally, data collaboratives can be established and executed in many different ways, combining the available technology with local needs. However, each data collaborative should be able to perform at least one analytical operation in a distributed environment [**Minimal Utility – Collaborative Computation].*** Since the original goal of the Collaborative was to train one statistical model, it is a natural consequence that it should serve its goal. While the hospitals may opt-in for training multiple models, one common undertaking is an essential criterium for the existence of the collaborative.

ISSN 2281-5147

7. Data Collaborative in Brief

In very brief terms, the ***Data Collaborative*** is a vehicle for training one or more statistical models in collaboration with other members – where they share resources and data to come up with a shared result. But how can the model be learnt ***collaboratively*** without firstly gathering the raw data in one (central) place? Well – and here perhaps you will need to take it at face value – the model is learnt through sharing the intermediate values that are learnt locally. More precisely, each participant is learning their own local model based on their own data. In a subsequent step, a global model is created using a mixture of local models. The technique that we used for the sake of this research is called ***Federated Learning,*** and it was proposed as the more privacy-centered manner of statistical learning some years ago. Referring to an example of hospitals that were made before, each of the hospitals will train its own model for classifying the X-ray images based on the locally available data. The next step will involve merging those models into one global model that can accumulate the knowledge of all its local counterparts. The procedure can continue until a satisfactory result is reached.

There are a few more technical issues that we experiment with in that setting. Firstly, there is of course, an issue of fault detection. Since the number of collaborators can be semi-trustworthy, it is crucial to detect potential intruders and free-riders. Although it can be difficult to imagine it in terms of hospitals (that are public institutions of a high reputation), let us assume that instead of hospitals, the Data Collaborative is formed by a number of industry partners that want to train one model for consumer classification. In such a case, it may be reasonable to expect that some of them may be either interested in obtaining a model for free (without really contributing their own knowledge) or jeopardising the global model (by including false information). In one of our papers called *Amplified Contribution Analysis for Federated Learning* in 2024, we presented an intuitive way of detecting participants that could potentially harm the global model (Zuziak & Rinzivillo, 2024)[7]. Another open issue would be that of *personalisation*. Since the models can be learnt on different data-generating distributions, there may not be a suitable mixture of models that suits all

[7] Link to the paper form April 2024: https://link.springer.com/chapter/10.1007/978-3-031-58553-1_6

ISSN 2281-5147

the participants of the learning. In this case, we would like to make some splits between the members of the collaborative to allow them to learn individualised models. If the hospitals are located in places of vastly different population characteristics, one diagnostic model could not necessarily be effective locally, as it could fail to focus on lcoally relevant traits of the population. In this case, we would like to automatically detect such variations in needs and allow the participants to *personalise* their models. Lastly, there are some issues with privacy that may not be so obvious to spot. In the introduction to statistical models, I have said that memory cards that preserve information about what types of genomes are associated with the higher chance of genetic disease occurrence are called *parameters*. Well, those parameters can also store personal information about objects they were trained on. For example, an attacker may be able to infer whether the genome of a certain person was included in the model training and its corresponding label. This would imply that they will obtain knowledge about whether a certain individual is associated with a higher chance of genetic disease occurrence.

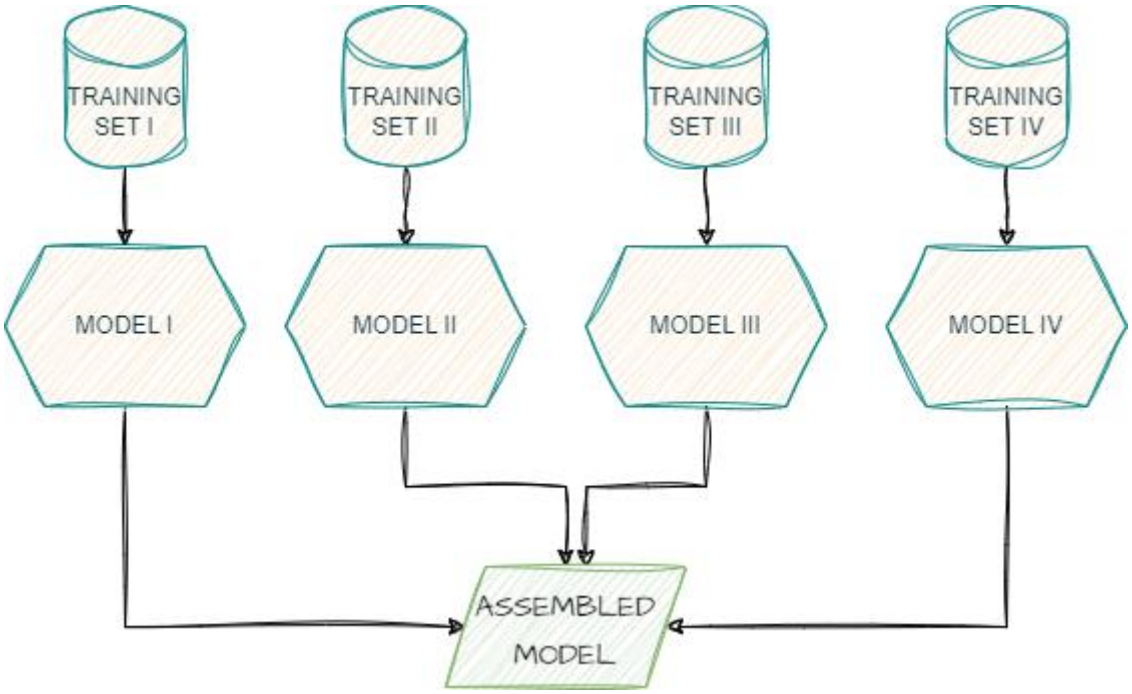


Fig.4: Model assembly. By combining a multiple model into one, a number of different organizations can create can shared model with far great capabilities, without the need to transfer the data directly into hands of one of the organizations.

ISSN 2281-5147

8. The Farewell Note

There are many concepts and nuances that were not included in this text – as they go beyond the brief description that could be delivered in one digestible piece of writing. However – I have – as an author - promised to deliver a set of pointers to a curious reader who would like to expand their knowledge upon this lecture. Throughout this text, I have included numerous links to articles that served as building blocks for this research – and I can firmly reassure every curious soul – that those links provide their own sources, upon which you can further expand your understanding of the topics delivered here.

In the last paragraphs, I have tried to explain the notion of *statistical model* in the most accessible manner possible. I have also related the concept of a model to one of personal data and why – in some cases – the availability of training samples may be highly dependent on the size of the business infrastructure (and – in turn – how this benefits larger players). This all ties back to the concept of *Data Collaborative* – a formal association of individual members that joints efforts in training one common statistical model. In the afterwords, I want to express my belief that we all value our personal data. Hence, we should not turn away from the difficult discussion on how we all may benefit from it when it comes to statistical inference. As the examples were simplified in order to explain the described concepts, they are by no means restrictive to the number of ways that the *Data Collaborative* can be used. From real-time traffic jam inference to research on genomes, common efforts in statistical analysis may impact our lives in a number of ways. It only depends on us and our communities how we will use it to benefit us and those around us.

9. Selected Readings

Cox, D. R. (2006). Principles of Statistical Inference. Cambridge University Press. https://doi.org/10.1017/CBO9780511813559

European Commission. (2024, September 1). What is personal data? [Govermental Website]. European Commission Website. https://commission.europa.eu/law/law-topic/data-protection/reform/what-personal-data_en

ISSN 2281-5147

Levy, M. G. (2023, February 1). Machines Learn Better if We Teach Them the Basics. Quanta Magazine. https://www.quantamagazine.org/machines-learn-better-if-we-teach-them-the-basics-20230201/

Mozilla Insights, Jonathan van Geuns, & Ana Brandusescu. (2020). Shifting Power Through Data Governance (p. 22) [White Paper]. Mozilla Foundation. https://assets.mofoprod.net/network/documents/ShiftingPower.pdf

National Human Genome Research Institute (NIH). (2024, September 1). Genetic Marker (Glossary Entry) [Govermental Website]. National Human Genome Research Institute (NIH). https://www.genome.gov/genetics-glossary/Genetic-Marker

Richard P. Feynman & Anthony Zee (introduction_. (2024). QED: The Strange Theory of Light and Matter. Princeton University Press. https://press.princeton.edu/books/paperback/9780691164090/qed?srsltid=AfmBOorY57KBeFZCtddUvUs0yZLLSUlSGOtGFAlXQnIZKNrQM9MNIFhM

Zuziak, M. K., Hinrichs, O., Abdrassulova, A., & Rinzivillo, S. (2023). Data Collaboratives with the Use of Decentralised Learning. Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency, 615–625. https://doi.org/10.1145/3593013.3594029

Zuziak, M. K., & Rinzivillo, S. (2024). Amplified Contribution Analysis for Federated Learning. In I. Miliou, N. Piatkowski, & P. Papapetrou (Eds.), Advances in Intelligent Data Analysis XXII (pp. 68–79). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-58553-1_6

ISSN 2281-5147

ISSN 2281-5147

POLICING THE AI ADJUDICATOR: THE STUDY OF ALGORITHMIC ACCOUNTABILITY FOR AUTOMATED DECISION MAKING SYSTEMS IN THE PUBLIC SECTOR.

Mitisha Gaur*

Abstract

AI is ubiquitous in public and private sectors used for optimizing tasks through complex data analysis. While the technology is promising, its use in high-risk domains raises concerns about trust, fairness, and accountability. This chapter analyzes AI backed automated decision-making systems being used by public authorities and advocates for a strict governance framework based on meaningful transparency, risk management and algorithmic accountability practices focused on safeguarding fundamental rights and upholding the rule of law by adhering to the principles of natural justice.

*Mitisha is a Marie Skłodowska-Curie Action's fellow working as an Early Stage Researcher (Law) with the Legality Attentive Data Scientists (LeADS) Project funded under the EU's Horizon 2020 Framework. The primary focus of her research is the use and regulation of high-risk artificial intelligence systems deployed in adjudication environments such as those in courts, regulatory bodies, government departments etc.; during her research Mitisha is focused on studying transparency and legal viability vis-à-vis AI systems and the social impact these systems have when deployed in adjudication environments. She is based at the Lider Lab, Scuola Superiore Sant'Anna, Pisa (Italy).
Mitisha.Gaur@santannapisa.it
This work is supported by the European Union's funded project Legality Attentive Data Scientists (LeADS) under Grant Agreement no. 956562.

ISSN 2281-5147

Table of Contents

Keywords

Algorithmic Accountability – Meaningful Transparency – Automated Decision Making– Public Authorities – Artificial Intelligence

1. Introduction

Artificial intelligence has turned into a seemingly ubiquitous presence with its use spanning over multiple sectors such as energy, finance, education, healthcare, navigation and public administration. The central appeal of using AI based technology (AI Systems) lies in the purported public sentiment around its superintelligence. The primary driver of this superintelligence is associated with the ability of AI Systems to

ISSN 2281-5147

identify patterns within a dataset and generate insights by using analytical techniques rooted in statistical analyses, recognition of recurring patterns, mathematical computations etc. These techniques guide optimisation efforts for various activities across sectors. For example: the use of AI based prediction analytics can help in the optimisation of energy load across power grids by harnessing user data to decide where electricity is to be supplied in order to be compatible with the user requirements that vary across homes, industries and commercial buildings. Similarly, the use of these prediction based analytics fuelled by AI Systems has also permeated more dynamic and sensitive fields such as the financial sector where AI systems are used by banks for the assessment of credit applications, within public administrations to disburse government subsidies to persons eligible under welfare schemes and also by law enforcement authorities in order to decipher criminal activities in areas with high criminal activities. The governance framework applicable to an AI System is determined by the level of risk which may be associated with the AI System. The classifications for the levels of AI risk adopted by the European Union's (EU) AI Act which is the primary legislation governing AI systems across the EU, are divided in four broad categories, namely (1) Unacceptable Risks: AI systems marked for unacceptable risk are prohibited from being used and include AI Systems acting as social scoring systems used by financial institutions to evaluate candidates for their creditworthiness based on behavioural data regarding spending habits, credit history etc., AI Systems which aim to manipulate children or other vulnerable groups such as emotional manipulation through the use of virtual assistants, the use of AI Systems for real-time remote biometric processing such as emotion recognition of individuals in work spaces etc.; (2) High-Risk: The tasks performed by AI systems in circumstances which may have a significant and (potentially) harmful impact on the quality of life as well as the freedoms and liberties enjoyed by human beings are classified as high-risk tasks. Consequently, the AI Systems used to perform, augment or assist in the performance of any such high-risk tasks are termed as High-risk AI Systems. These include the use of AI systems for law enforcement functions such as those focused on evaluating the viability of evidence in the course of investigation or those used to evaluate the risk of a person becoming the victim of criminal offences etc., the performance of public administration functions such as to evaluate the eligibility of applicants for public benefits such as welfare benefits, healthcare assistance and associated services; (3) Limited Risk: These includes chatbots used in

ISSN 2281-5147

customer service and AI Systems with capabilities to create deepfakes,; and lastly (4) Minimal Risk: These include AI Systems that are used to perform low-risk functions such as AI systems acting as spam filters, writing and text editing tools etc.

The regulatory matrix under the AI Act varies across the 4 risk levels namely- (1) the AI systems exhibiting unacceptable risk are prohibited from being used; (2) the ones exhibiting high-risk are bound by a comprehensive set of legal obligations which include periodic and event-based compliances that are associated with both the technical and organisational requirements focused on use of high-risk AI systems such as risk assessment and mitigation, issuance of instructions of use, fundamental right impact assessment, conformity assessment, technical documentation, record-keeping etc.; (3) AI systems with low risk are bound by minimal reporting requirements and finally, (4) the AI Act exempts the use of AI systems with minimal risk from its purview, however with the increase in the use of generative AI tools, this may change.

The two-fold regulatory obligations, namely: technical and organisational, that are imposed on relevant stakeholders engaged with high-risk AI Systems, which include providers of AI Systems i.e. entities that develop and subsequently license a high-risk AI System and a deployer who may be a natural or a legal person such as an organisation, company, public authority, that uses an AI System to perform functions.

This chapter is focused on the use of AI Systems by government departments such as taxation authorities, family and child welfare departments etc as well as by judicial authorities such as courts, tribunals etc (collectively referred to as Public Authorities). The use of AI systems by Public Authorities has a direct impact on the health, safety and fundamental rights of the decision-subjects. The acknowledgement of the risk associated with the use of AI systems in this domain is also reflected in the AI Act's classification of an AI system used by Public Authorities to assist in performing sensitive tasks such as the dispensing of public welfare, assist judges in researching and interpreting facts etc. as a high-risk AI system.

ISSN 2281-5147

2. Automated Decision Making Systems and Public Authorities: Preserving the Adjudicatory Fabric.

There are multiple applications of AI Systems within Public Authorities, however for the purposes of this chapter, the central focus lies on the use of AI Systems in their capacity as automated decision-making systems (ADMS). These ADMS may be machine learning based statistical tools which provide quantifiable indications to the user such as rate of successful resolution of a legal dispute (whether in favour of the petitioner or the defendant) based on a given set of facts or provide a risk based scoring associated with the applications they process such as the application to request public welfare funds.

These types of inputs by the ADMS have a material impact on the manner in which the user of the ADMS views the applications presented to them. Another popular ADMS tool is the newer generative AI Systems (GenAI) such as the famous large language models ChatGPT and Gemini, that are backed by natural language processing technology and may be designed to provide answers to the questions a user may pose to it. The mimicking of human behaviour by GenAI may lead to increased trust between the deployer and the AI System, however, numerous investigations have observed flaws within the GenAI system which has been observed to produced fictitious answers to queries posed to it. This phenomenon has been termed as *hallucinations*. A prominent example is when ChatGPT constructed a fictional caselaw to support its answer to a question placed before it.

Against this backdrop, the efforts to govern the development and use of High-risk ADMS by Public Authorities, a crucial factor to consider is the methodology of use associated with such an ADMS.

The decision making process across Public Authorities is divided into four (4) key stages: (1) acquisition of information based on which a decision has to be made; (2) the analysis of the information; (3) selection of decision based on the analysis of information and lastly; (4) the implementation of the selected decision.

ISSN 2281-5147

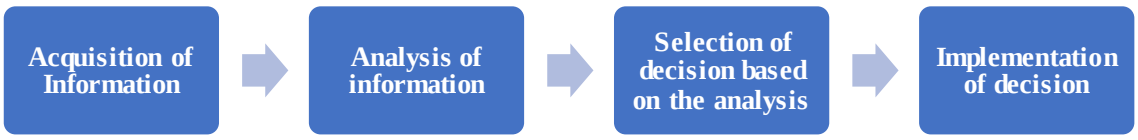


Figure 1: Stages Of Decision Making Within A Public Authority

The manner in which ADMS is used by Public Authorities is materially affected by the stage of decision making within which such ADMS is deployed, as the instructions of use, associated risk as well as transparency requirements will differ. Prior to delving into the technical and organisational constraints attached to the ADMS, it is crucial to understand the context within which the ADMS is deployed by the Public beyond the simplistic reduction of "to assist in decision making". This assistance can be understood as a plethora of tasks and ranges in the level of automation associated with it. This can be at the performance of simple tasks such as the streamlining of information focused on expediting tasks (low level of automation), the assistance in writing a judgement (moderate level of automation) or the choosing of a decision based on historical data on behalf of the Public Authority (high level of automation).

The endeavour of decision making by a Public Authority is guided by the balancing of many crucial duties and associated responsibilities shouldered by such Public Authorities. These include the duty to uphold and protect the rights of citizens, the responsibility to exercise the rule of law, and the duty to adhere to the principle of natural justice. These principles of natural justice are the very fulcrum of robust judicial systems (such as courts and tribunals) and quasi-judicial systems (such as

ISSN 2281-5147

government department and boards providing licenses and administrative rulings based on legal statutes) across the world. These principles of natural justice are as follows- (1) The adjudicating authority must not be biased whether in favour of or against the persons seeking legal recourse; (2) Pronouncing of a reasoned order by the adjudication authority; (3) Absence of unjustifiable delay in adjudication; (4) Ability of a person to make legal representation in front of the adjudication authority and; (5) Adequate notice to be provided to a person to prepare for the legal proceedings initiated against them. Consequently, the material impact awarded by the principles of natural justice upon the decision making processes by Public Authorities is two-fold: (1) allows Public Authorities to build precedent and; (2) the adherence to the principles of natural justice allows for examination of the judgements of Public Authorities by supervising authorities such as superior courts with appropriate jurisdiction on the subject matter.

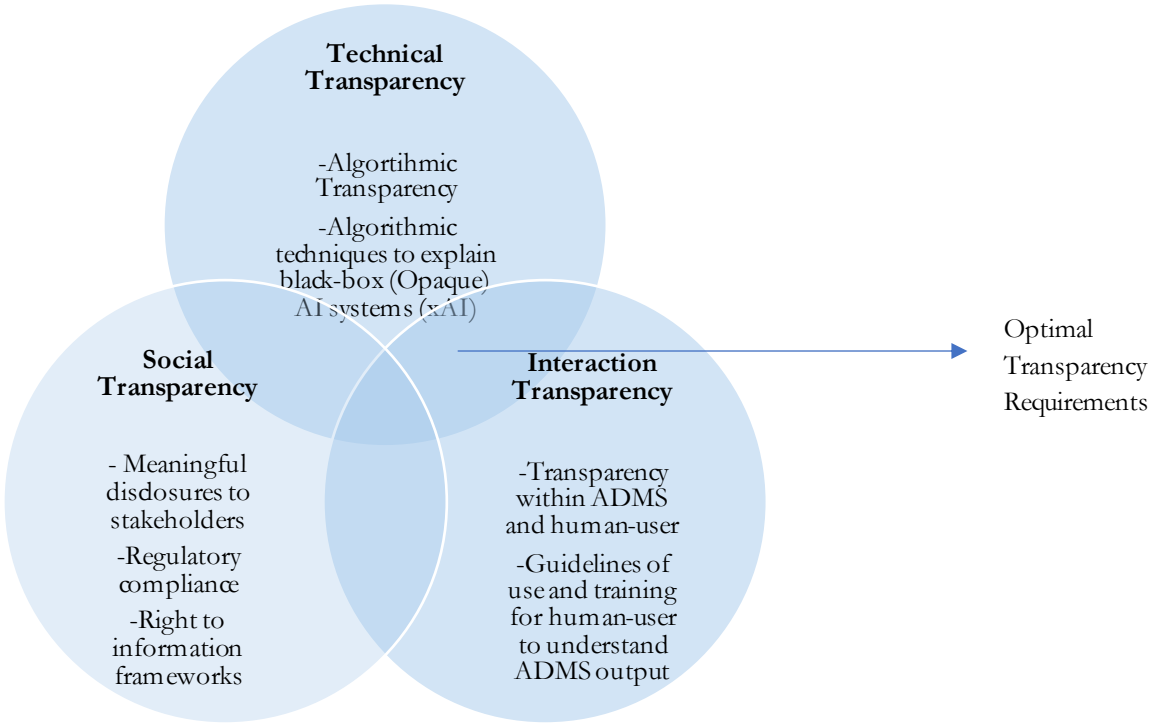
Against this backdrop, this chapter focuses on three crucial issues associated with the use of ADMS by Public Authorities, namely (1) How to develop an ADMS which can be safely deployed within a Public Authority to assist in carrying out judicial and quasi-judicial functions?; (2) How to ensure that the ADMS is deployed safely within a Public Authority and is being used in the correct context? and lastly; (3) How to protect the persons subjected to these decisions from adverse effects of the ADMS use by Public Authorities?

3. The Transparency Triad: Informing ADMS within a Public Authority

The common thread across these three challenges (as discussed in Section 2) is that by virtue of the expectation of transparency from Public Authorities, the decisions of Public Authorities as well as any associated information which aides and assist such decision making is subject to explanation under the mechanisms of the access to information framework, through which an individual can seek specific information from Public Authorities. The combined reading of the legal requirements, duties and responsibilities as well as the explanation requirements associated with Public Authorities, transparency associated with decision making emerges as a central theme. Therefore, it is evident that ADMS deployed within Public Authorities must also comply with necessary transparency requirements. The transparency mandate

ISSN 2281-5147

associated with the use of ADMS within an Public Authority is tri-fold and comprises of (1) Technical transparency: This form of transparency is associated with the inner workings of the ADMS and the ability of the ADMS to provide a meaningful explanation about the output it produces; (2) Interaction Transparency: This form of transparency is associated with the ability of the human-user of an ADMS to adequately understand the inner workings of the ADMS and make an informed decision as to whether or not the output produced by the ADMS must be relied upon; and finally (3) Social Transparency: This pertains to the sharing of information (such as the underlying technology, the trustworthiness and safety) vis-à-vis the ADMS by the Public Authority with relevant stakeholders such as citizens, persons subjected to the decision in which an ADMS was involved, regulatory bodies etc. The triad of these three types of transparency related requirements creates the optimal transparency requirements for a Public Authority deploying ADMS.



ISSN 2281-5147

Figure 2: Tri-fold Transparency Mandate For Public Authorities Deploying ADMS

4. The Ever-Shifting Landscape: Context and Public Authority in ADMS

The deployment of an ADMS within a Public Authority raises crucial questions, particularly concerning the role of context of use and how it informs decision making. This also informs the manner in which the ADMS may be used in a safe and trustworthy manner, while protecting the interests of the developers, deployers as well as decision-subjects.

Contextual clarity during the development and the deployment of ADMS in a Public authority is crucial. This context is far ranging from the stage of decision making within which an ADMS is deployed to the decision-subjects and whether they are minors or members of a vulnerable class, the level of technical expertise showcased by the human-user relying upon the computations of the ADMS such as the level of AI literacy which dictates their ability to adequately comprehend and rely on the decision-outcome of an ADMS. Another crucial constraint is that context within an ADMS is ever changing and the technical infrastructure of the ADMS must evolve accordingly to accommodate it. For example: Changes in legal regulation or social norms may directly impact the context within which an ADMS must be deployed or relied upon.

Another crucial layer of contextual clarity within an ADMS is a defined purpose for which the ADMS is being used by an organisation. Is it designed to automate routine tasks like eligibility checks or delve into complex areas like parole decisions? The level of automation and the associated degree of human oversight may vary significantly depending on this background. For instance, an ADMS flagging the possibility of fraudulent tax returns might have a lower human oversight threshold compared to a system assessing child custody disputes.

The impact of an ADMS decision directly depends on who it affects, therefore a crucial consideration is that any decisions which directly or indirectly impact minors or vulnerable populations such as migrants and refugees, people with disabilities, racial and ethnic minorities etc. necessitate a more nuanced understanding of context.

ISSN 2281-5147

Public authorities rely on qualified users to interpret and implement ADMS outputs. The level of technical expertise these individuals possess forms another crucial layer of context, as noted previously. Therefore, comprehensive training becomes paramount to ensure that the users of the ADMS can understand the limitations of the ADMS such as hallucinations, presence of bias in data which leads to algorithmic discrimination, lack of transparency, improper data quality and diversity (which may again circle back to the problem of bias within the dataset used to develop AI systems) and can critically analyse its recommendations before implementing the same. This is also pertinent to combat cases of automation-bias within human-users where users are observed to overly rely on the decision outcome produced by an ADMS. In some cases, additional data or context not captured by the system might be crucial for the final decision, therefore it is crucial that in the absence of the same the user of the ADMS possesses adequate levels of AI literacy to spot the challenges associated with the ADMS and take necessary steps. Therefore, it has been noted that owing to these possible shortcomings, high-risk AI systems should not be deployed without meaningful human oversight.

The final, and perhaps most critical, aspect of context is its dynamic nature. Public policies, demographics, and technology evolve constantly, this is relevant more so when the subject matter is the use of AI Systems within Public Authorities. An ADMS designed for efficient distribution of unemployment benefits during an economic downturn might need adjustments during a period of low unemployment. Regular reviews and updates are essential to ensure the ADMS adapts its algorithms and data sets to reflect the ever-changing environment such that the ADMS may produce results relevant to the current societal norms and use based requirements.

Public authorities face a complex challenge in deploying ADMS effectively. Striking a balance between automation and human oversight, ensuring fairness for all decision-subjects, and continuously adapting the system to a dynamic environment requires a nuanced understanding of context. Therefore, as discussed previously, transparency and an adequate level of AI literacy are key here. Public authorities need to be transparent about how they are using ADMS and put in place mechanisms for individuals to challenge automated decisions.

ISSN 2281-5147

5. Public Authorities and Algorithmic Accountability: The Final Piece Of The ADMS Puzzle

The final crucial piece of the puzzle which focuses on the developing and deploying an ADMS within a Public Authority, is algorithmic accountability. Algorithmic accountability is the practice of holding the deployers of algorithms responsible for its effects. This inclusion of responsibility through algorithmic accountability has a direct impact on the manner in which the ADMS is not only developed and deployed within a Public Authority to augment decision making but also how it is perceived socially.

The ADMS is a technical component or a tool which is deployed within a social and organisational environment, therefore this interaction between the technical components as well as the social and organizations components creates an interdependent ecosystem referred to as a sociotechnical system (STS). The theory of STS is essentially an organisational development approach that focuses on the complexities associated with workflow within an organisation based on the interaction between social (such as persons, levels of education and technical skills), organizational (such as processes, timelines and task flows) and technical elements (such as hardware and software components) within an organisation (collectively referred to as the STS Stakeholders). The characteristics of an STS, which vary greatly from one organisation to another, impacts the interaction that a technical component such as an ADMS has in the face of contextual information which are driven by the social and organisational factors within an STS. To simplify it further, imagine a big system within the Public Authority, like a machine with many parts. To make all the parts work well together, they need clear rules. These rules cover how different stakeholders interact with the system and how the system itself works. Therefore, for optimal functioning these rules should be transparent, meaning not only must they be easy to understand but also that each stakeholder must know the rules applicable to itself as well as its counterparts and the information pertaining to whether such rules have been followed or not should be readily available. Also, there needs to be a dedicated person in charge, making sure everything runs smoothly (human oversight). Things get even more complex if this system works with other similar ones, like connecting different machines. In these cases, it's crucial to have clear rules and a clear chain of command to avoid confusion or problems.

ISSN 2281-5147

The concept of algorithmic accountability is closely associated with the adherence to the rules that govern an STS. Therefore, in order to truly ensure algorithmic accountability, it is crucial to divide the accountability frameworks based on subject matter. In keeping with this, the algorithmic accountability framework is divided into 4 main parts, namely (1) Technical Accountability; (2) Organisational Accountability; (3) Social Accountability and; (4) Regulatory Accountability.

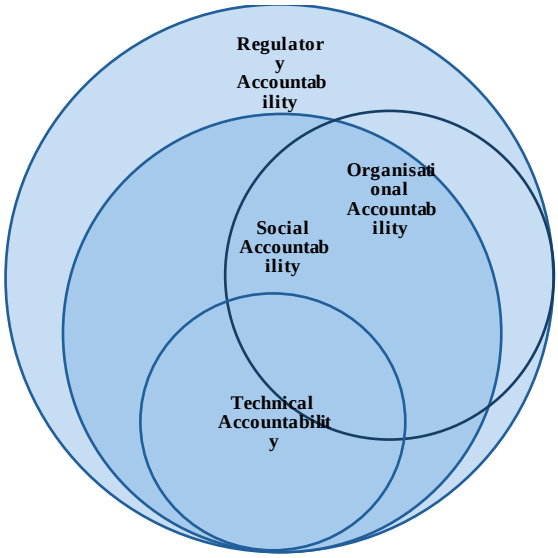


Figure 3: Interplay Between The Components Of Algorithmic Accountability

To ensure a fair and ethical application of ADMS within a Public Authority, a multi-pronged approach to accountability is crucial. Imagine a complex STS tasked with making critical choices such as the ADMS used within a Public Authority to investigate possible fraudulent activity vis-à-vis tax benefits and is required to function with transparency and fairness. This can be achieved through a layered framework encompassing each: technical, organizational, social, and regulatory accountability measures.

Technical accountability focuses on the inner workings of the ADMS, like ensuring the system is built well and uses reliable data. Meaningful and multi-faceted transparency, as explained in the previous sections, is key. These include deliberations such as "Can the users understand how decisions are made?" The presence of clear and meaningful explanations, even for those without technical expertise, are essential. Further, rigorous testing and audits are crucial to identify and eliminate bias before deployment of the ADMS within the Public Authority. This prevents the system from

ISSN 2281-5147

perpetuating social inequalities through skewed datasets based on which the ADMS may have been trained. Consequently, data governance becomes vital as well, focusing on pertinent questions such as "Where does the data come from?" and "How is the data used?" A robust technical accountability framework ensures that the data is accurate, complete, and collected ethically, with individuals having control over their personal information, is used by the ADMS. Finally, security and explainability are important. Strong cybersecurity measures protect the system from tampering, and the ability to meaningfully explain decisions helps identify errors and promotes fairness.

Organizational accountability focuses on responsibility within the STS itself, such as the assignment of clear roles within a complex environment. Defining roles and responsibilities ensures everyone working with the STS understands their part and more importantly, can be held accountable, in case of an adverse event such as ADMS led bias propagation. While ADMS helps to automate decision making tasks, human oversight, as noted previously, remains crucial. Human involvement through oversight allows for questioning decisions, considering the presence of contextual factors that the algorithm might not take into account, and maintaining alignment with ethical and legal principles associated with the use of high risk AI by Public Authorities such as ADMS. The ADMS user training provided by Public Authorities in conjunction with the developers of the ADMS, empowers those human-users interacting with the ADMS to understand its limitations and capabilities, as well as providing thorough instructions of use to the ADMS users which include risk escalation mechanisms, adverse output mechanisms etc. This training should be focused on empowering the human-user to analyse the ADMS outputs critically, identifying potential biases or errors. Risk management in these scenarios also remains crucial, as organizations implementing ADMS need a well-defined plan to identify and mitigate potential risks.

Further, social accountability empowers the public to hold institutions using ADMS accountable. Here, the focus also lies on meaningful transparency and encouraging public engagement and discussion. Public awareness ensures they understand how ADMS are used and its potential impacts on decision subject as well as the wider impact population that the societal fabric is comprised of. The core tenet of social accountability is transparency in communication, which focuses on the right to explanation vis-à-vis persons and their right to know how ADMS are used by Public

ISSN 2281-5147

Authorities which will consequently affect their lives. Additionally, it is to be noted that social accountability fuelled by public participation through public meetings such as townhouses, public consultations and conducting public polls allows for meaningful public involvement in the development and deployment stages of an ADMS. This has the ability to help in identifying potentially material issues before they arise and ensuring the system is designed ethically, in compliance with legal regulations and with social good in mind. Independent audits (by algorithmic watchdog organisations and citizen's rights groups) and reviews provide valuable insights and identify areas for improvement. Finally, grievance redressal mechanisms are essential for fairness and building public trust. Persons who are subjected to ADMS by the Public Authorities should have clear and accessible ways to challenge unfair or discriminatory decisions made by ADMS, which must be resolved within a stipulated timeframe.

Lastly, regulatory accountability sets the ground rules for the operation of an ADMS by a Public Authority to aide its efforts to perform material public functions, through establishing rules and regulations. Regulatory frameworks (such as the EU's AI Act) may be focused on defining clear expectations for fairness, transparency, and accountability, which may be imposed on the developers and deployers of the ADMS by means of regulatory compliance. Another useful solution may be the establishment of independent regulatory bodies, focused on overseeing the use of ADMS by Public Authorities, monitoring adherence to compliance and investigating potential breaches. The practice of impact assessments (which includes both algorithmic impact assessment (focused on the technical robustness of the ADMS) as well as the fundamental rights impact assessment (focused on the impact of the ADMS on the fundamental rights of the decision subjects) is a welcome and crucial tool, focused on the evaluation of potential risks ex-ante deployment, allowing for mitigation strategies to be put in place. These forms of ex-ante requirements are preferred over the enforcement of ex-post sanctions, in the case of high-risk AI such as the use of ADMS by Public Authorities since the degree of harm caused by a biased or faulty ADMS may cause material harm to the decision-subject which may not be mitigated through sanctions or monetary compensation. This brings us to sanctions and enforcement mechanisms such as operational injunctions enforced on ADMS until the algorithmic shortcomings are tackled, that are used to ensure accountability and deter misuse of the ADMS by Public Authorities.

ISSN 2281-5147

6. Conclusions

These four pillars of accountability are interconnected and work best when implemented in harmony. Technical measures ensure the fairness and transparency of the ADMS itself, while organizational measures establish clear roles and responsibilities for those deploying and using the system. Social accountability empowers the public through awareness, participation, and grievance redressal, and regulatory accountability sets the ground rules through regulation, independent oversight, and regulatory enforcement. By weaving these pillars together, we can ensure that ADMS are used responsibly, ethically, and with the public good at the forefront. This multifaceted approach allows us to build trust in the complex world of automated decision-making, ensuring it serves society effectively and fairly.

The primary roadblock in the investigations pertaining to the use of ADMS by Public Authorities is the seemingly opaque algorithms which are often found to be in use and consequently rupture the requirement of algorithmic transparency and the ability of users to perceive or explain (to the decision subjects) the inner workings of the ADMS on which the Public Authority relies. This opacity, in turn, while creates distrust in the minds of the decision subjects vis-à-vis the ADMS also provides a leeway to Public Administrations to dodge questions regarding inner mechanisms of the algorithms in use. Further, there has been an observed lack of internal training which leads to either the misuse, over-reliance or the unreserved mistrust regarding the use of the ADMS. Additionally, the observed pattern of the development and deployment of an ADMS for use by a Public Authority is that the same institution dons the hat for both the deployer as well as the developer, therefore there is no internal accountability or balance of powers within these two roles, which may lead to the unchecked perpetration of bias and lack of social as well as regulatory accountability in case of a misadventure at the hands of the Public Authority relying on an ADMS.

The benefits of using a high-risk AI System such as an ADMS cannot be considered in insolation with the responsibilities associated with the use of such tools. This becomes even more crucial in light of the fact that the deployer and user of the ADMS is a Public Authority, an organisation that wields immense power and its actions or

ISSN 2281-5147

inactions have a significant impact on the lives of people. Therefore, a holistic approach is required that spans across all STS Stakeholders (organisational, social and technical) while developing, deploying and using such an AI Systems, rooted in upholding meaningful transparency, promoting meaningful human oversight and keeping the two in check through use of algorithmic accountability measures.

7. Selected Readings

(1) Laux, J., Wachter, S., & Mittelstadt, B. (2024). Trustworthy artificial intelligence and the European Union AI act: On the conflation of trustworthiness and acceptability of risk. Regulation & Governance, 18(1), 3-32.

(2) Wang, A., Kapoor, S., Barocas, S., & Narayanan, A. (2024). Against Predictive Optimization: On the Legitimacy of Decision-making Algorithms That Optimize Predictive Accuracy. ACM Journal on Responsible Computing, 1(1), 1–45.

(3) Roehl, U. B. U. (2023). Automated decision-making and good administration: Views from inside the government machinery. Government Information Quarterly, 40(4), 101864.

(4) Mökander, J., & Axente, M. (2023). Ethics-based auditing of automated decision-making systems: Intervention points and policy implications. AI & SOCIETY, 38(1), 153–171.

(5) Chopra, A. K., & Singh, M. P. (2018). Sociotechnical Systems and Ethics in the Large. Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, 48–53.

(6) Ruckenstein, M., Lomborg, S., & Hansen, S. S. (2020, November). Re-humanising automated decision-making. In Workshop report from the ADM: Nordic Perspectives research network.

(7) Malgieri, G., & Pasquale, F. A. (2022). From Transparency to Justification: Toward Ex Ante Accountability for AI. SSRN Electronic Journal.

ISSN 2281-5147

(8) Diakopoulos, N. (2020). Accountability, transparency, and algorithms. The Oxford handbook of ethics of AI, 17(4), 197.

(9) Olsen, H. P., Hildebrandt, T. T., Wiesener, C., Larsen, M. S., & Flügge, A. W. A. (2024). The Right to Transparency in Public Governance: Freedom of Information and the Use of Artificial Intelligence by Public Agencies. *Digital Government: Research and Practice*, *5*(1), 1–15.

(10) Kaminski, M. E. (2023). Regulating the Risks of AI. Forthcoming, Boston University Law Review, 103, 22-21.

(11) Novelli, C., Casolari, F., Rotolo, A., Taddeo, M., & Floridi, L. (2024). AI Risk Assessment: A Scenario-Based, Proportional Methodology for the AI Act. Digital Society, 3, 13.

ISSN 2281-5147

ISSN 2281-5147

THE PERILS OF VALUE-ALIGNMENT

Robert Lee. Poe*

Abstract

This essay argues that global AI governance risks institutionalizing violations of fundamental rights. It critiques the ethical foundation of AI governance, observing that moral objectives are being prioritized over legal obligations, leading to conflicts with the rule of law. The essay calls for a re-evaluation of AI governance strategies, urging a realistic approach that respects citizens, legal precedent, and the nuanced realities of social engineering, aiming to provide an account of some of the dangers in governing artificial intelligence—with an emphasis on Justice.

Table of Contents

* Researchet at Scuola Superiore Sant'Anna and Ph.D Candidate (National Doctorate, University of Pisa), robertlee.poe@santannapisa.it
This work is supported by the European Union's funded project Legality Attentive Data Scientists (LeADS) under Grant Agreement no. 956562

ISSN 2281-5147

Keywords

Non-discrimination – Fair Machine Learning – Distribute Justice – Legal Design– AI Ethics

1. Introduction

An alarm bell rung may be doubted by those who hear it. Neither party in that relationship is without the possibility to error in judgement due to ignorance of circumstance in the calculation of risk. We may not have all the relevant facts, and even if we did, we may not understand why that fact-set is the relevant set and not another. Written alarms seek to draw attention to a risk and should justify the attention being drawn. This alarm has been written because the field responsible for preventing algorithmic discrimination has developed the tools and methodology being used to discriminate; and AI governance, in a quest to make the world a better place, has led to the standardization of automated distributive decisions that engage in real-life, systematic violations of the fundamental right to non-discrimination (See e.g. ISO/IEC TR 24027, 2021[1] and NIST Special Publication 1270, 2022).[2]

Recent work took note of a study (Raghavan et al., 2020) which concluded that automated hiring software for pre-screening and interviewing candidates "debias" in accordance with independence-based group fairness metrics, and we argued that such practices would likely be in violation of Article 21 of the Charter of Fundamental Rights (hereafter Charter), because the Court of Justice of the European Union (Court) has found preferential treatment in the hiring context to be limited to tie-breaking scenarios (Poe & El Mestari, 2024). This is not a fringe legal conclusion, neither in the academic study of algorithmic discrimination nor in the wider EU non-discrimination law discourse. Meanwhile, human-resource software companies like Workday sell automated hiring and promotion systems throughout Europe, and they readily state in their adverts to their pursuit of a global diversity, **equity**, and inclusion policy (Global Blueprint for Belonging and Diversity, 2024).[3] Workday also readily

[1] https://www.iso.org/standard/77607.html
[2] https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.1270.pdf
[3] https://www.hci.org/system/files/2024-05/Global Blueprint.pdf

ISSN 2281-5147

divulges their use of the fair machine learning techniques that will likely be found to engage in unlawful discrimination (Global Impact Report on VIBE, 2024).[4]

I believe this phenomenon is the result of a number of failures, but this brief essay identifies two: a failure (1) to understand what distributive decisions are and how the process of distribution relates to the fundamental right of non-discrimination, and (2) to understand the difference between moral and legal facts and reasoning. The essay will begin with an introduction describing the risk. The first proper section of the essay will address the question of whether system design is to conform with the law or whether the law is to conform with system design when hard moral and legal questions are both relevant to design, and the second section will give a nuanced primer on distributive decisions.

First, a necessary definition: A distributive decision is a function performed by a *designer* (e.g. an authority, guardian, provider, employer) of a distributive decision-making process, allocating *distribuendum* (e.g. goods, services, burdens, offerings) among *distributees* (e.g. applicants, recipients, patients, customers, employees, citizens). This essay will rely heavily on the above definition and the terminology within it. The description is precise, but the idea is simple: a teacher distributing grades to students, an employer distributing job offers to candidates, a bank distributing loans to applicants, a state distributing benefits to citizens—all of these are distributive decisions. In other words, the study of distributive decisions then is the study of those real-life, distributive decisions that take place all throughout human activity within jurisdictions, only some of which are subject to legal scrutiny in accordance with the principle of equal treatment.

At the very least, the automated distributive decisions subject to legal scrutiny are those that have legally significant effects. But traditionally, the distributive decisions subjected to legal scrutiny are those whose authority has ". . . a direct impact on others' lives. We may be civil servants, shopkeepers, employers, landlords or doctors who decide over how public powers are used, or how private goods and services are offered. In these non-personal contexts, non-discrimination law intervenes in the choices we make . . ." (*Handbook on European Non-Discrimination Law*, 2018, p. 42) Now, of course, these are general comments about a complicated topic i.e., the scope

[4] https://www.workday.com/content/dam/web/en-us/documents/other/workday-global-impact-report.pdf

ISSN 2281-5147

of EU non-discrimination law, which takes seriously a variety of balanced interests, but the core idea is a reasonable summary based on the enumerations of Article 21 of the Charter in light of the Equality Directives and the Court's jurisprudence (e.g. in employment relations).

Fair Machine Learning (FML) is (not entirely but predominately on the policy side) the doctrinal, technical application of a philosophy known as Distributive Social Justice. *Distributive Justice* asks the question: how should the boons and burdens of a society be shared amongst the members of a society? Distributive *Social* Justice as *an* available answer to that question. Data scientists search within samples of personal and non-personal data for evidence of numerical inequalities between groups of persons in the outcomes of past distributive decisions, in an effort to determine whether those distributive decisions were Just. This approach came naturally to the field because social scientists necessarily study recorded **outcomes** of decision processes. The study of discrimination in outcomes has a long and celebrated history, and so it was naturally relied upon for the creation of the tools for addressing algorithmic discrimination in legal-technical design.

Distributive decisions are increasingly becoming automated, and the data scientist is one among others (e.g. the philosopher, legal scholar, activist, and politician) that are defining how such decisions should be made, often institutionally (limited oversight of decision-making). For the data scientist, a decision is Just if it is not biased. But bias, for the FML community, no longer means a deviation from the *true value* of a parameter or variable but instead a deviation from *group equality* or similarity in decision outcomes. This is the difference between truth as an objective and Distributive Social Justice as an objective in learning from data, described statistically—nothing fancy. Once bias is defined as group equality or similarity in decision outcomes, the data scientist deems an outcome that is unequal or dissimilar between groups of persons, Discriminatory and therefore Unjust, if the grouping represents the sub-groups of a protected characteristic (e.g. sex, gender, race, ethnicity, and so on). And when standardizing the design of automated distributive decisions, the data scientist requires distributive decisions to be Non-Discriminatory and therefore Just. That conception of Justice and Discrimination is different than what can be found in the law (I will argue for good reason), but it satisfies the moral intuitions of many (risk creation)—at least at first glance (risk reduction).

ISSN 2281-5147

This Distributive Social Justice logic is not only being applied to automated distributive decisions, but in general data quality checks. The conception is being baked into the concept of *quality* data. Information that conveys a group disparity is removed or given a threshold of acceptability. This logic has been advocated for: during the training of models, during the mining of data (automated collection process) itself (that will later train other models), during the human-verified data quality check, and during an investigation of the outcomes of automated distributive decisions. The error is subtle yet *disastrous*. Nuance is to be had and is present throughout the body of the essay and other works of the author and colleagues, but the logic of FML described above is a synthesis of the non-legal, yet highly influential logic that risks being standardized. One might feel strongly about social justice, but there is a difference between debating about the lawful scope of positive action social policy and debating about not having a scope—without scope there can be no proportionality.

2. Legal Design or Design made Legal?

First, legal *sense* needs to be distinguished from *nonsense*. In legal philosophy this is done theoretically through reliance on the separation thesis which holds that there exists a difference between what the law is and what the law ought to be. The separation thesis is well-founded given that the answer to most legal questions is straightforward due to sufficient statutory clarity, legislative history, and the past application of the law with similar fact patterns relevant to the facts in a specific case. There is neither time nor space to develop a full account of legal methodology to show the extent of its rigor here, so it will have to suffice to say that practicing lawyers can be held accountable for making frivolous legal arguments on behalf of their clients—that is, at least, in practice the formal process that distinguishes reasonable legal arguments from the unreasonable ones in accordance with professional licensing standards.

In the practice of translating fundamental rights jurisprudence into technical specifications for automated distributive decisions, the law of a given moment and jurisdiction should have priority over political or moral passion when providing legal expertise that will later guide design and the live use of such systems in a jurisdiction.

ISSN 2281-5147

A cornerstone of AI Governance generally, and Trustworthy AI specifically, is the premise that unlawfulness in AI systems inherently undermines their ethical standing—a point made repeatedly in the Guidelines for Trustworthy AI (also cited in the preamble of the AI act)—while also accommodating and encouraging the lawful, moral aims of individuals, businesses, and institutions operating within those boundaries. While many engage in and applaud going "beyond the law" to achieve a desired moral aim, the law itself must not be contradicted. And, before asking "in what direction," it should be understood that the cases which further apply, and thus define, fundamental rights are among the most controversial.

While moral conflict may sometimes be resolved through an appeal to reason, morals themselves are not the design of reason—whether natural or artificial. Morals are inherited from generation to generation, constructing our intuition through the observation of conduct and cultural learning—whether familial, local, or societal—appealing to our sense of proportionality for agreement with others, whose moral axioms are similar enough to allow for the identification of conflicts between axioms in application to a shared and relevant fact pattern. "Thought experiments" are useful for showing others that hard moral questions exist regardless of our awareness of them. But hard moral questions exist because the experimentees experience a contradiction of held axioms and force a synthesis, a moral reconstruction—prioritization via re-balancing, removal, or replacement of morals. However, there is no substitute for nuance. The moral wealth of humanity (with beautiful and tragic complexity and contradiction) cannot be expressed in a standard, nor should we delude ourselves that it might. While one's own moral compass might appear linear, non-discrimination law is emphatically not. Instead, it is proportional. The interpretation and application of the Charter of the Fundamental Rights of the European Union begins and ends at the Court of Justice of the European Union. If moral standardization is the goal, at least rely on legal standards to form the axiomatic basis.

If a moral conclusion mirrors a legal one, then at most it is a supplementary justification for lawful behavior. If a moral conclusion deviates from a legal conclusion, then it was brought about by a moral, not legal, interpretation. The difference between a hard question of morality and a hard question of law is, itself, the rule of law—through which equal treatment grew. To whatever extent the results

ISSN 2281-5147

of an ethical impact assessment have an obligatory nature, this point should be taken seriously. And, if AI ethics experts legitimize their value selection by deriving them from the fundamental rights of a given jurisdiction, then it follows that the definition should be left to fundamental rights jurisprudence. Thus, the question is not should the Court set the standard but, instead, has the standard already been set?

The question of value-alignment is not the question of whether and how an artificial intelligence can be made Good. The question of value-alignment is, in practice and in policy, whether and how current-generation AI system design (which is advancing faster than our responsible decision-making about it) can be aligned with legal obligations—both present and emerging—and their designers held accountable for violations. The objective outlined by the European Commission has been twofold: create a legal framework encouraging the adoption and innovation of AI systems in both the private and public sector, while ensuring that "high risk" purposes are brought into alignment with standards for health, safety, and the protection of fundamental rights. The enthusiastic yet cautious approach of the EU is a response to several immediately cognizable intranational and international concerns, ranging from socio-political stability to economic and defense competitiveness.

The most complex of value-alignment questions require a deep understanding, not solely of the related legal or technical subject-matter but, most importantly, the foundations which interconnect them. The problem is not that solutions are unavailable because of the novelty of artificial intelligence and the application of it. The problem is that the solutions are in the past and the researchers are in the future—more precisely in a Utopia. The crucial difference between the theoretical construction of a Utopia and the actual attempted construction of a utopia is the systematic violation of fundamental rights, always. Before the point where scholars are able to align a super-intelligent being with the Good—a dangerous and confused task to begin with—it should first be understood how to align an automated decision-making system (of any kind) with legal obligations and not the other way around. Otherwise, we are quite literally talking about a data revolution in the legal-political sense of the term.

This manuscript is about events, distributive decisions that occur in jurisdictions whose actions may have legal consequences, and if violations of equal treatment occur there may be redress for victims; but, if it is not obvious, the argument is that the

ISSN 2281-5147

definition of fundamental rights modeled technically must be modeled based on a faithful and current interpretation of the law—that means jurisdiction (cannot be global if about fundamental rights).

3. Distributive Decisions and Consequences

A few clarifications about the meaning of the term equal treatment are in order. First, *equal treatment* is a separate concept from *equal outcomes*. Outcomes may be called equal in a *factual* sense, indicating that two or more sums are identical, where those sums represent the sub-groups of a protected characteristic; in a *reactionary* sense, expressing disapproval or disgust when two or more sums are not identical, where those sums indicate differences between the sub-groups of a protected characteristic; in a *moral* and *political* sense, where theories of justice based on egalitarianism and social justice seek to resolve those differences through social policy; or in a *legal* sense, where the term is used synonymously with a term of art known as "substantive equality" which takes note of a legal fact—in this instance the Court of Justice of the European Union's jurisprudence on the application of direct and indirect discrimination doctrine and the scope of positive action social policy—and attempts to either predict how the Court might apply those legal doctrines in future cases or advising how the Court *ought* to interpret the law in those future cases in line with the goal of "substantive equality."

Treatment may be called equal in a *factual* sense, when a distributive decision does not prefer specific distributees or groups of distributees over others in relation to the creation and application of a standard. In the case of automated distributive decisions, evidence of difference in treatment in the factual sense is implemented by-design and evidence are discoverable with sufficient access to the system (more on this later). Treatment may also be called equal in a *reactionary* sense, expressing disapproval or disgust when a distributive decision does not measure each distributee or group of distributees in relation to the same standard; in a *moral* or *political* sense, where theories of justice based on merit and procedural justice seek to resolve those differences through social policy; or in a *legal* sense, where the term is used synonymously with a legal term of art known as "formal equality" which takes note of a legal fact—in this instance, again, the Court's jurisprudence on the application of direct and indirect

ISSN 2281-5147

discrimination doctrine and the scope of positive action social policies—and attempts to either predict how the Court might apply those legal doctrines in future cases or advising how the Court *ought* to interpret the law in those future cases in line with the goal of "formal equality."

Going forward I will use the term "sameness of treatment" and its derivatives for the expression of the *factual* concept of equal treatment, and I will use the term "sameness of outcome" and its derivatives to express the *factual* concept of equal outcomes. I will use the term "equal treatment" and its derivatives to express the principle of equal treatment. The principle of equal treatment can be found under Article 2 (1) of Directive 2000/78 establishing a general framework for equal treatment in employment and occupation: the principle of equal treatment "shall mean that there shall be no direct or indirect discrimination whatsoever on any of the grounds referred to in Article 1." Thus, equal treatment is satisfied where a distributive decision neither (1) prefers a sub-group *based* on a protected characteristic, or an indissociable proxy for a protected characteristic, over another in relation to the creation or application of a standard, *unless* such differential treatment can be justified by concerns of health, safety, or the protection of the rights and freedoms of others, where the means of achieving the aim are appropriate and necessary (direct discrimination); nor (2) applies a standard that disproportionately affects a sub-group of a protected characteristic *based* on a dissociable characteristic, *unless* the use of the dissociable characteristic is justified by a legitimate aim and the means of achieving that aim are appropriate and necessary (indirect discrimination).

In short, neither differences in treatment nor differences in outcome are necessarily discriminatory. I place the definitions above knowing full-well the meaning may remain elusive in the hope it becomes clearer through the explanation of distributive decisions, sameness of treatment, sameness of outcomes, equal treatment, and the principle of proportionality.

Again, a distributive decision is a function performed by a *designer* (e.g. an authority, guardian, provider) of a distributive decision-making process, allocating *distribuendum* (e.g. goods, services, burdens, offerings) among *distributees* (e.g. applicants, recipients, patients, customers, citizens). Not all distributive decisions are subject to equal treatment. But it must be remembered that distributive decisions happen in legal jurisdictions and thus are the subject of legal consideration when falling into the scope

ISSN 2281-5147

of law. A comprehensive account of legal considerations would begin with examining the "why" behind the purpose inherent in a specific distributive decision, revealing whether it falls within the scope of the law. It should then address the "who," identifying the relevant roles involved, followed by the "where," which, in part, determines the jurisdiction. Next, it should consider the "what," outlining the rights to hold and transfer the distribuendum, and conclude with the "how," detailing the means of pursuing the purpose, including infrastructural considerations for personal and non-personal data protection when distributive decisions are automated and digital or processed the old-fashioned way. Information about the design of a distributive decision can be merely stated or presumed, ex-ante conforming and evident, or, where no access to the decision-process itself is available, witnessed in the outcomes of a distributive decision. In this essay, I am concerned with the equal treatment aspect of the "how" of distributive decision-making; but in the paragraph below, I am solely concerned with the sameness of treatment aspect of the "how" of distributive decision-making.

So, how is a distributive decision made? It is made in two distinct phases: *categorical ordering* and *patterning*. **Categorical ordering** is essential to the **creation** of a standard and **patterning** is essential for the **application** of a standard. A standard is by nature a categorical ordering—distinct from patterning because it is a prerequisite of patterning—which parses the relevant from the non-relevant information for the measurement of distributees. A standard is a categorical ordering because out of all observable categories of information about distributees, through which they may be compared, it must be determined which information is useful and proper for a given measurement (feature space) (i.e. what qualities does the successful candidate have for a given job posting).

Once a standard is designed via categorical ordering, it must be designed via patterning: each distributee may be described in relation to that standard, ranked, and the distribuendum may be distributed accordingly. This process will result in inequalities in the outcomes received by distributees or groups of distributees where they are, in fact, different in relation to the standard. If a designer of a distributive decision finds different outcomes undesirable (the non-legal conception of Good or Bad Discrimination), the designer might, from the outset, create a standard which prefers certain distributees or groups of distributees over others, a re-ordering guided

ISSN 2281-5147

by a patterning dilemma: preferential purpose. For such a designer, it is necessary to not let the decision be based on a well-defined categorical ordering of correct information, because such a standard would result in different outcomes.

There are two limiting cases where sameness of treatment and sameness of outcomes are *compatible*. First, it is possible that a well-defined categorical ordering of correct information treats distributees or groups of distributees the same and, at the same time, results in the same outcomes. Such an instance is only possible where distributees or groups of distributees are, in fact, the same in measurement with the standard. Second, it is possible to create a standard of such minimal content so at to not recognize differences between distributees or groups of distributees, such that all distributees or groups of distributees are both treated the same in relation to the standard while ensuring the same outcomes. The crucial qualitative difference between the two limiting cases that ***must*** be understood is that in the first a designer reaches sameness of outcomes via factual equality (a process bounded by the categorized descriptions of distributees competing under a specific standard but cooperating in societies allowed for by such standards, i.e. the preservation of spontaneous order in centralized, distributive decisions) and, in the second, the designer reaches equal outcomes via mere presentation (a process bounded by designer choice in preference).

In between these two limiting cases, sameness of treatment and sameness of outcome are *incompatible*. A designer may set a threshold at any point between these two extremes, demarcating how much difference in outcomes between distributees or groups of distributees is acceptable in their distributive decision, and by doing so, necessarily demarcates how much difference in treatment between distributees or groups of distributes is acceptable. Properly understood, the quantitative trade-off of the threshold is the sameness of treatment in one hand of a designer and the sameness of outcomes in the other.

Example A: imagine an automated loan approval system, where the algorithm is designed to approve loans based on credit score (categorical ordering: e.g. on-time payment history, credit-debt ratio, open and closed lines of credit, and so on, to predict the likelihood of loan default), it may result in different approval rates among various sub-groups of protected characteristics, not because it is *based* (if protected characteristic or indissociable proxy was not a category) on those protected

ISSN 2281-5147

characteristics but because there may be a correlation between the protected characteristic and creditworthiness in the target population. By setting a threshold (necessary use of protected characteristic or an indissociable proxy) for acceptable variance in approval rates between sub-groups of a protected characteristic (sameness of outcomes, difference in treatment), the bank increases the likelihood that applicants who would have been deemed uncreditworthy for the loan specifics are given loans—putting the borrower at risk of loan default by-default (based on the protected characteristic) and the lender at risk of a loss. Remember, while a bank is likely to shift that burden to others via fraud and/or taxpayer bailouts or socialization, the borrower will be left with little recourse.

Example B: imagine an automated hiring system, where the algorithm is designed to evaluate candidates based on qualifications such as education, work experience, and skill assessments (categorical ordering: e.g., degree level, years of relevant experience, results of standardized tests, and so on, to predict job performance). It may result in different hiring rates among various sub-groups of protected characteristics, not because it is based (if protected characteristic or indissociable proxy was not a category) on those protected characteristics but because there may be a correlation between the protected characteristic and certain qualifications in the target population. By setting a threshold (necessary use of protected characteristic or an indissociable proxy) for acceptable variance in hiring rates between sub-groups of a protected characteristic (sameness of outcomes, difference in treatment), the company increases the likelihood that candidates who would have been deemed less qualified for the job specifics are hired—putting the company at risk of decreased performance and the candidate at risk of struggling in the role. Remember, while a company may try to address performance gaps through training or reassigning roles, the hired candidate might face challenges in job satisfaction and career growth.

When a distributive decision process is automated, evidence of difference in treatment exists. In short, it exists in the space between a representative sample and generalizable hypothesis assumptions, and so by measuring the relationships between distributees or groups of distributees in the data sample and comparing them with the relationships in the outcomes of an automated distributive decision, violations of sameness of treatment are quantitatively detectable—relevant to direct discrimination in the same way differences in outcome are relevant to indirect discrimination.

ISSN 2281-5147

4. Conclusion

As we move forward in the development and standardization of AI systems, it is crucial not to oversimplify the complex issues at hand. The questions surrounding AI governance and the alignment of technology with legal standards are far from straightforward. This essay offers just a glimpse into the intricate challenges we face—challenges that require deep, ongoing analysis and public discourse. While I have highlighted some critical points, this discussion is only a piece of a much larger puzzle.

The development of AI systems cannot be separated from the legal frameworks that protect fundamental rights. As we standardize technologies that have the potential to profoundly impact society, we must ensure that these standards are informed by a nuanced understanding of both legal obligations and the realities of social engineering. The issues at stake are not just technical or academic; they have real consequences for the lives of individuals and the functioning of societies. Therefore, it is imperative that these debates extend beyond closed circles of experts and become part of a broader public discourse.

ISSN 2281-5147