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