DATA IN LEGAL PHILOSOPHY AND PRACTICE: CONTEMPORARY CONCEPTUAL AND REGULATORY DEVELOPMENTS

TOMAS BERKMANAS

Vytautas Magnus University, Lithuania

SAULĖ MILČIUVIENĖ

Vytautas Magnus University, Lithuania

SUMMARY. This article analyzes the recent evolution of the concept and regulation of data as an integral attribute of the digital age. At the conceptual level, two extreme approaches are in focus. First, the so-called *dataism* approach, characterized by a speculative but also often radical and shocking holism. The other approach is dominated by an equally radical, inert, and problematic skepticism. This paper proposes an alternative philosophical position towards data (as a phenomenon), analogous to sociological jurisprudence or legal realism in the philosophy of law. Various relevant aspects are analyzed at the regulatory level, ranging from attempts to define 'data' in legislation to possible tectonic breaks in the paradigm of copyright and *sui generis* rights. In this context, it is highlighted that the protection of the property rights of data holders, which has been the subject of a narrow academic debate, even though data have a growing economic value and are increasingly important for economic growth and prosperity. This paper calls for a more balanced approach to data while also balancing the interests of data holders, data users, and society. This may also require some significant changes in the regulation of copyright and *sui generis* rights, in which data are only subject to copyright and *sui generis* protection in very limited cases. (Translated with www.DeepL.com/Translator, free version).

KEYWORDS: data, dataism, copyright, digitalization.

INTRODUCTION

Recently, data has emerged as an influential force that profoundly shapes various facets of our society, ranging from business and governance to healthcare and education. The undeniable potential of data-driven tools to revolutionize industries and enhance human life has catapulted data into the forefront of intense debates and speculations among scholars across diverse fields of art and science. Consequently, pivotal questions demand urgent attention. Firstly, how should the phenomenon of data be conceptualized without losing rational adequacy and common sense? How should it be regulated to ensure its responsible and ethical use?

To formulate a sustainable conceptual and regulatory framework, it becomes imperative to adopt a well-balanced paradigmatic approach towards data and its integration into our lives. This article is an attempt to contribute to the

contemporary scholarly discussion on data, beginning from profound philosophical considerations and extending to their tangible reflection in the legal realm.

General/philosophical considerations on data usually fluctuate between two extremes. While some envision a data-centric future where everything, including human beings, is transformed into data, others express skepticism regarding the reduction of everything (including – and especially – humanity) to mere algorithms and data management. By looking closer into these considerations and making certain parallels in the history of ideas, we may find that another more scholarly adequate and socially beneficial approach is possible, one that provides a more commonsensical theoretical foundation for practical approaches that determine the landscape of EU legal regulation concerning data today.

Thus, further analysis in the article proceeds with an in-depth examination of how EU legislation defines data, encompassing its expansive scope and the diverse categories that fall under its purview. From research data, which underpins scientific exploration, to dynamic data characterized by real time updates and personal data bearing sensitive information about individuals, we aim to unravel the multifaceted nature of data and its far-reaching legal implications.

In the sphere of intellectual property, copyright law plays a pivotal role in protecting original works of authorship as soon as an author fixes the work in a tangible form of expression. However, when it comes to data, the situation becomes significantly more intricate. Generally, copyright may not protect raw data; nevertheless, it does come into play when data is organized in the structured form of a database. As we delve into this domain, we shall uncover the unique rights conferred by the Database Directive, which includes copyright protection for databases displaying originality and the *sui generis* right for databases entailing substantial investment.

To achieve this objective, the article is structured around five main focal points:

1. Presenting and critically assessing the philosophical considerations surrounding data.
2. Proposing a less speculative general approach to data which is also more coherent in relation to further development of science around it.
3. Analysing the existing definitions of data within the legal framework of the European Union.
4. Scrutinizing the interplay between copyright law and data, illuminating the legal intricacies surrounding data protection.
5. Discussing potential future developments in data regulation, contemplating the challenges and opportunities.

By undertaking this comprehensive exploration, we endeavor to contribute to a more profound understanding of data's impact and its governance, fostering an

environment where data-driven advancements can flourish responsibly and ethically in the service of the common good.

1. FROM SKEPSIS TO APOCALYPSE: CONTEMPORARY GENERAL SPECULATIONS ON DATA AND A NON-SPECULATIVE WAY OUT

Contemporary generalizations on how the relationship between data and humans will develop in the future – in other words, the potentialities behind the new realities of the *data-driven* age and science, which is still *human* age and science – fluctuate between two extremes. On the one side, we find almost apocalyptic scenarios of a complete change in worldview from anthropocentric to data-centric that will eventually lead to the elimination of humans from the game, or even their disappearance altogether. Importantly, these scenarios are not considered mere fantasies but real possibilities (Harari 2017: 443, 454, 461). Others are skeptical of the possibility of transforming everything to data, especially that which concerns the life of a human (as a whole of senses, emotions, thoughts). Also on the skeptical side, *data* is considered to be *nothing special,* just another detail of/in our lives among many others, something of considerably lesser importance in certain hierarchies (for example, from traditional intellectual law perspective it is considered to be merely raw material from which a valuable intellectual outcome/product is derived). Both approaches are not only speculative and impractical, but also constitute another *deja vu* in the history of ideas.

The first approach is otherwise called *dataism* or *data religion*, which allegedly "has already conquered most of the scientific establishment" (Harari 2017: 428 *et seq.*). What is important here is that in its final form, this approach allegedly states that not only is everything *transformable* into data but that everything (the world, the universe, etc.) in fact *is* data. This kind of ontological generalization resembles many other so-called holistic claims that are nothing more than rhetorical acts of speech. On the other hand, they are correctly called a *form of religion* as they cannot be proven, only believed in. Some examples of this are the political holism of Carl Schmitt or the economic holism of Karl Marx. The latter, as a matter of rhetorical form, was very well expressed in the old Hollywood movie *Network* in the famous sermon of Mr. Jensen to Mr. Beale, which we will try to rephrase here, with slight changes to adjust to dataism: "It is the system of *data* which determines the totality of life in this universe. That is the natural order of things today. That is atomic and subatomic and galactic structure of things today. <...> There is no America, there is no democracy. There is only *Facebook* and *Google* and *Wikipedia* and *Web of Science*, *EOSC*, *EBSCO* and *Cambridge Core*. Those are the *nation*s of the world

today. We no longer live in a world of nations and ideologies. The world is a college of *data bases* intrinsically determined by the immutable bylaws of *data flow*. The world is a *data flow*" (*cf.* Lumet 1976; emphasis mine/authors).

If we would follow this rephrasing in the movie, Mr. Beale would go on air preaching about *dataistic* cosmology after hearing this sermon, eventually emphasizing that it leads to the loss of our humanity and the transforming of humans into humanoids. Regardless, it appears that nothing has changed in this style of *insightful* and *scholarly* universal generalizations and corresponding apocalypticism which renders humans *small* or even *disappearing*, besides/inside this grandiose totality of either currency, corporations and business or data, its bases, and its flow. From this worldview, we – as all other organisms (including senses and emotional, and intellectual capacities) – transform into mere algorithms; Beethoven's Fifth Symphony and the flu virus are simply two different patterns of dataflow, capitalism and communism – two competing data processing systems, and so on (Harari 2017: 429–430).

As a matter of religion, however, this approach has its own enigmas and taboos; or, to take a more elegant Popperian approach to handling problems, *questions that must be suspended or put aside.*[1] One of such suspended questions is exactly "where do these great algorithms [as data operating tools] come from?" (Harari 2017: 457). At a more general level, behind this query is a good old epistemological question: how exactly do we *know*? How does our intellect (especially as a matter of innovation) work? What *is* intellect? After the linguistic turn in philosophy, the latter question was reformulated to the even more down-to-earth question: what is language? (i.e., substratum through which our knowing/intellect is expressed). Accordingly, if contemporary philosophy has no clue how our fundamentally linguistic intellect works or even what it is, how could we possibly know the same things about AI? If epistemology still struggles to grasp how innovative scholarly ideas come to our mind (Popper 2005: 7 *et seq*.), how could we know whether genuine algorithms somehow spring into existence, either in HI or AI?

Another suspended question could be simplified even further: What exactly *is* data?, i.e., as an ontological matter/object? It is fair to say that most human beings would not agree that the universe consists solely of data and its flow. Most people would not conceive of the drinking of wine as a literal consumption of data. On the other hand, in the light of the digitalization of almost everything, we could agree with the statement that much of what we understand as certain aspects of the world/universe could be somehow *transformed* into data that has digital form. But what is the so-called atom or unit of this digital data, and what exactly *is* it at

---

[1] In this way Popper *solves* apparently fundamental epistemological problem of "how it happens that a new idea occurs to a man" (see Popper 2005: 7 *et seq*).

the atomic (or subatomic) level? According to data scholars, it is a *bit*. But then, what is a bit? We say that it is a certain digit, either 1 or 0, but from the ontological point of view, this is not exactly true. In fact, a bit is a very ontological phenomenon; it is only metaphorically digital. A bit does not consist of zeroes and ones but is *physically represented* by, for example, two distinct levels of voltage or current or two directions of magnetization. A bit is a very ontological tool, as with the physical signifiers of language, used to express physical and metaphysical aspects of the world that are ontologically different from this tool. Bits *themselves* are not, and will never turn into, the sounds of a symphony or colors of a photograph; for that we need speakers or monitors. It is naive to argue that the movie *Matrix* is factually based – that the world *is* (or might possibly be) a totality of flowing green zeroes and ones. On the other hand, if we reject this naivete, this dataistic dizziness, then what exactly remains as/of data from the ontological standpoint? If it is *something more* than levels of voltage or directions of magnetization – something metaphysical perhaps – then are we not drawn back to the old problems of the enigmas of logos and/or language.

With all these problems of data-centric / dataistic approach in mind, another skeptical approach may appear as more reasonable and scientifically adequate. In fact, in comparison to the first approach, it lacks grandiose universalizations and corresponding ontological claims. Instead, it is dominated by relatively simplistic doubts about the possibilities (at the very least) entertained by the first approach: "it's doubtful whether life can really be reduced to dataflows;" "it is equally doubtful whether [human] life boils down to mere decision-making;" "is there perhaps something in the universe that cannot be reduced to data;" and so on (Harari 2017: 458–459).

Nevertheless, the second approach is not without its problems. The first issue is a practical one: by entertaining this *calm down* mentality, we may become ignorant to the processes related to data and AI, as though nothing important and relevant is happening – *ignorant* to the extent that it becomes problematic, especially if it affects political and regulatory processes. From this perspective, the corresponding critique of contemporary political structures and political establishment from the camp of the first approach makes sense (albeit accompanied by colourful rhetorical forms) (Harari 2017: 436 *et seq*.).

Another problem is still more general and philosophical in nature. It appears that the aforementioned skepsis is very much founded on another holistic paradigm with all the corresponding claims, i.e., anthropocentrism. We are still reluctant to reduce humanity and its most fundamental characteristic – human intellect / HI (historically derivative of anthropocentrism is/was logocentrism) to data, or to make ourselves equal to or even inferior to AI. From this perspective, the

second approach might be considered as representing an anthropocentric inertia in the face of the emerging paradigm of datacentrism. However, it appears that in fact both centrisms, particularly because of their holistic claims, have very similar ontological problems. In analyzing the first camp, we posed the questions: *what is algorithm?* or *what is data?* The second camp has its own corresponding questions and problems – not only *what is logos* or *what is language,* but also *what is a human* and *what is a person?* (see, for example, Agamben 2004; Agamben 2000: 63 *et seq*.).

This situation with two general approaches to data and AI is similar to the one in legal philosophy – an oscillation between *jusnaturalism* and *juspositivism*. Speaking very generally, the former is focused on a certain holistic given outside human – as nature, deity, something sacral – that determines law and makes it independent of us, humans, thus independently determining human fate. The latter is focused on a human as the sole and only existential determinant of law and, from the other side, law as something that only serves our human purposes. By following this parallel and especially if we somehow ignore the underlying philosophical problems discussed above, we may endlessly oscillate between two extremes while generally reflecting on data and AI with all the rhetorical but purely speculative arguments – for example, that "eventually the Internet-of-All-Things [whatever that may actually be] may become sacred in its own right" (Harari 2017: 454) or that AI may be hacking the operating system of human civilization (Harari 2023).

But a detour to legal philosophy here was not accidental. In fact, history of legal philosophy (as an example) may offer a certain way out from this oscillation and all the anxieties that it leads to. In the course of the 20th century, a certain transformation has happened in legal philosophy, marked by the fundamental change in methodological approach to the analysis of law as a phenomenon. This could be generally categorized as the change of the modus of primary/initial question – instead of never-ending and, apparently, hopeless speculation on *what [the law is],* philosophers instead decided to turn their gaze to *how [the law functions][2].* Otherwise known as *the law in action* approach, this approach gave birth to sociological jurisprudence and legal realism as separate schools of legal philosophy that have considerably influenced and changed the landscape of legal philosophy since then (see, generally, Freeman 2008: 835 *et seq*.; Wacks 2012: 145 *et seq*.). We think that in order not to immerse ourselves in the speculation and anxiety discussed above, which are more focused on attracting popular attention but not to rational coherency and scrutiny, it would be more reasonable and practical to make the same shift of gaze (or modus of a question) in relation to data and AI. What should interest us more is not so much *what (data, algorithm, AI) is it,* with all the holistic

---

[2]   On the other hand, of course, this decision is very similar to the Popperian-style suspending of certain questions.

generalizations that follow any attempt to answer this question, but more *how it functions/works* and *what practical problems we should solve to make it function more efficiently and to contribute more to the common good?* In the end, these questions are posed to politicians (who act as regulators), but from this perspective we believe that it is much better for them to be administrators only; they should not entertain grandiose visions, but instead solve the related practical problems with coherency, diligence, and carefulness (Harari 2017: 438–439). We will now identify a few areas of practical problem(s) and then, in the next chapter, discuss practical proposals and solutions in relation to some of them.

One example of a practical problem is the impact of the abundance of data that varies widely in its credibility to phenomena such as censorship and providing accurate information to society. "In the past, censorship worked by blocking the flow of information. In the twenty-first century censorship works by flooding people with irrelevant information" (Harari 2017: 462). Accordingly, this new reality requires new approaches to providing society with credible and reliable information.

Another example is the already well-known tension between privacy and openness of data and information. From the dataistic perspective, we could be very enthusiastic about the potential benefits we could achieve if we completely lifted the veil of our privacy and made ourselves completely transparent to algorithms (Harari 2017: 448). However, especially in the light of the EU law, starting from *General Data Protection Regulation*, we could not escape the practical reality of the need for a more balanced approach instead of the aforementioned radical one.

Finally, from the dataistic perspective we could speculate that "dataism inverts the traditional pyramid of learning. Hitherto, data was seen as only the first step in a long chain of intellectual activity. Humans were supposed to distil data into information, information into knowledge, and knowledge into wisdom. However, Dataists believe that humans can no longer cope with the immense flows of data, hence they cannot distil data into information, let alone into knowledge or wisdom. The work of processing data should be therefore entrusted to electronic algorithms, whose capacity far exceeds that of the human brain" (Harari 2017: 429). But upon careful reading of this extract, it is easy to see that what happens here has nothing to do with the pyramid turned upside down. Instead, we have a problem of the first stage – the problem of large amounts of data. And it appears that we have the tool to solve this problem: algorithms/AI. Of course, what exactly AI will produce – information, knowledge, or wisdom (or perhaps all or none of these) – remains to be answered. From a practical perspective, what may be of interest is that this new reality may require new regulatory approaches in the field of the law of intellectual property, for example. Traditionally, this law was focused on the higher stages of the pyramid – knowledge and wisdom – while data and information (with certain

exceptions related to the intellectual input in relation to them) was left outside of its regulatory scope. However, in combination with algorithms/AI as a new holistic phenomenon, data may require novel regulatory approaches that may affect *inter alia* law of intellectual property that is still quite traditional.

In the next chapter, we will discuss certain practical approaches and solutions, especially related to the last aspect of data/AI in action. But first, we will explore the regulatory attempts to explain what data is.

2. NEW CHALLENGES IN DATA LEGAL REGULATION

In this chapter, we will examine how certain philosophical questions raised in the previous chapter are reflected in EU legal regulation. As discussed earlier, the philosophical debate addresses the nature of data. Therefore, in this chapter, we will delve into how data is defined in EU legislation. Moreover, the focus is not solely on defining specific objects, but on identifying problems and finding effective ways to enhance the functioning of society for the common good. Addressing this, we will explore the extent of the copyright and the *sui generis* right protection for data and how data sharing issues are more efficiently regulated.

2.1. HOW DATA IS DEFINED IN EU LEGISLATION

As implied previously, the definition of data lacks universality. It can vary substantially across different domains and require contextualization that is especially valid at a practical (regulatory) level. Prior to exploring the interrelation between copyright and data, we will investigate the definition of data as outlined in EU legislation, as well as various categories of data that are distinguished therein.

The general definition of data is provided in the Data Governance Act (Regulation (EU) 2022/868, 30/05/2022) and in the proposal for the Data Act (EU Commission Proposal for a Regulation COM/2022/68, 23/02/2022). Therein, it is stated that "'data' means any digital representation of acts, facts or information and any compilation of such acts, facts or information, including in the form of sound, visual or audiovisual recording" (EU Regulation 2022/868, 30/05/2022) (EU Commission Proposal for a Regulation COM/2022/68, 23/02/2022). This implies that data can be described as a digital depiction of actions, events, or information, regardless of its format, covering a spectrum that incorporates, but is not confined to, recordings in sound, visual, or audiovisual forms. It comprises a compilation of such elements, which collectively contribute to a comprehensive information repository.

Additionally, EU legal acts contain definitions of individual categories of data, which are subject to different legal regulations. Some of these categories are (1) research data, (2) dynamic data, (3) personal data, and (4) metering and consumption data.

Research data act as a foundational pillar in supporting the answers to research inquiries and enriching the scientific domain with empirical evidence. The definition of research data is provided in the directive on Open data (Directive (EU) 2019/1024, 20/06/2019). It states that "'research data' means documents in a digital form, other than scientific publications, which are collected or produced in the course of scientific research activities and are used as evidence in the research process, or are commonly accepted in the research community as necessary to validate research findings and results" (Directive (EU) 2019/1024, 20/06/2019). In other words, in the realm of academic inquiry, the term *research data* pertains to digitally recorded materials distinct from scientific publications. These materials are generated or collected during scientific research activities and serve as pivotal evidence in the research process. Moreover, they are commonly acknowledged within the research community as indispensable in validating research findings and results.

In recent times, the emergence of dynamic data has garnered attention due to its frequent or real-time updates. Given their inherent flexibility and responsiveness, dynamic data holds immense potential in enabling adaptive and up-to-date analyses across various applications. The definition of dynamic data is also provided in the directive on Open data (Directive (EU) 2019/1024, 20/06/2019). It states that "'dynamic data' means documents in a digital form, subject to frequent or real-time updates, in particular, because of their volatility or rapid obsolescence; data generated by sensors are typically considered to be dynamic data" (Directive (EU) 2019/1024, 20/06/2019). The dynamic nature of such data arises from factors like volatility or rapid obsolescence. Notably, data generated by sensors exemplify dynamic data, as they continuously provide fresh inputs, reflecting real-world changes.

As personal data carries sensitive information, its collection, storage, and usage must adhere to stringent privacy and security protocols to protect individuals' rights and safeguard against misuse. The definition of *personal data* is provided in General Data Protection Regulation. (Regulation (EU) 2016/679, 27/04/2016) It states that "'personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person" (Regulation (EU) 2016/679, 27/04/2016). The notion of *personal data* involves information concerning a

recognized or recognizable individual, known as the *data subject*. This identification can occur either directly or indirectly, employing diverse markers.

Metering and consumption data play a pivotal role in informing energy policies, facilitating sustainable practices, and ensuring efficient resource allocation in the energy sector. The definition of metering and consumption data is given in implementing regulation on interoperability requirements (Commission Implementing Regulation (EU) 2023/1162, 6 /06/2023). It states that "'metering and consumption data' means meter readings of electricity consumption from the grid, or electricity fed into the grid, or consumption from on-site generation facilities which are connected to the grid and includes validated historical data and non-validated near-real time data" (Commission Implementing Regulation (EU) 2023/1162, 6 /06/2023). *Metering and consumption data*, are essential for monitoring energy consumption patterns, optimizing energy distribution, and evaluating the efficiency of power generation systems.

To conclude, data is mainly defined in EU legislation as a digital representation of acts, facts, and information. Data refers to factual information collected, recorded, or represented in a structured form, with various types serving specific purposes. Research data supports scientific inquiries and provides empirical evidence, excluding scientific publications. Dynamic data gains attention due to its real-time updates and flexibility, exemplified by sensor-generated inputs. Personal data includes sensitive information about identifiable individuals, requiring stringent privacy and security measures. Metering and consumption data are pivotal for informing energy policies and optimizing resource allocation in the energy sector. In relation to our discussion in chapter 1, we saw how a practical conceptual approach to data, as reflected in concrete regulatory articulations, is different and distant from the corresponding speculative generalizations.

2.2. DATA: REGULATORY CHALLENGES AND SOLUTIONS

Generally, copyright law does not protect data. This principle is based on the idea/expression dichotomy, which distinguishes between protected forms (such as specific expressions) and unprotected contents (ideas or data). While data itself is not eligible for copyright protection, the situation becomes more complex when dealing with data compilations.

The Berne Convention (Berne Convention, 09/09/1886) protects collections of literary or artistic works, but it does not explicitly mention data collections, nor does it grant copyright to facts or news of the day. However, modern conventions such as the Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPs Agreement, 15/04/1994) and the WIPO Copyright Treaty (WIPO

Copyright Treaty, 20/12/1996) provide broader protection to compilations of data or other materials, especially if the selection or arrangement of their contents represents intellectual creations.

Within the EU intellectual property framework, data itself does not receive any protection. However, when data is organized in the structured form of a database, it becomes eligible for certain rights. The Database Directive (Directive 96/9/EC, 11/03/1996) introduced two distinct rights for the protection of databases. The first right, granted under Article 3(1), pertains to copyright protection for databases that constitute the author's own intellectual creation based on the selection or arrangement of their contents. The second right, governed by Article 7 of the Database Directive, is known as the *sui generis* database right. This right has been a topic of debate since its introduction and aims to protect databases that have undergone substantial qualitative and/or quantitative investment in obtaining, verifying, or presenting their contents. The intention behind this right is to incentivize European companies to invest in creating databases and foster growth in the data economy.

The ruling of the Court of Justice of the European Union (CJEU) in *Football Dataco and Others v. Sportradar and Others* is related to the scope of database protection. The case *Football Dataco and Others v. Sportradar and Others* (joined C-403/08 and C-429/08, 4/10/2011) revolves around the issue of copyright protection for sports data compilations. Football Dataco and others were seeking copyright protection for their database, which contained football match data, claiming that the database compilation represented an original intellectual creation. The CJEU clarified that for a data compilation to be eligible for copyright protection, it must meet the standard of originality. Originality, in this context, means that the selection or arrangement of the data must be the result of creative choices, indicating the author's own intellectual creation. The CJEU ruled that merely investing significant effort, skill, and labor in compiling the data is not sufficient to establish the originality and claim copyright protection. As a result, the Football Dataco database did not meet the threshold of originality and, therefore, could not be granted copyright protection under EU law.

The Football Dataco decision has implications beyond its specific case that extend to denying copyright protection for many contemporary compilations of machine-generated data. These databases are often automatically created by algorithms that process predefined data types for specific sectoral applications, with minimal human intervention, based on preset configurations set by the data collecting entity (Hugenholtz 2018). The generation of such databases is driven by technical or functional considerations rather than creative choices, making them ineligible for copyright protection as they lack the necessary originality. This aligns with the general principle that copyright protection requires human authorship. In the case

of dynamic databases exchanging data, human intervention is typically limited to establishing the initial framework, and the ongoing processes of data collection, generation, updating, and structuring are largely automated and continuous, happening in real-time. As a result, copyright protection is not applicable to these types of databases due to their automated and non-creative nature (Hugenholtz 2018).

Other important cases of the CJEU are *Fixtures Marketing Ltd v. Oy Veikkaus Ab* (C-46/02, 9/11/2004), *Fixtures Marketing Ltd v. Svenska Spel Ab* (C-338/02, 9/11/2004), *British Horseracing Board Ltd v. William Hill* (C-203/02, 9/11/2004), and *Fixtures Marketing Ltd v. OPAP* (C-444/02, 9/11/2004). In these decisions, the CJEU emphasized that the *sui generis* right does not extend to databases that are mere byproducts of an organization's primary activities. Consequently, *sui generis* rights may not broadly apply to databases created by public administrations or private companies that are solely necessary for their internal functioning, products, or services. This includes records maintained by public administrations and automatically captured or machine-generated data, like data from IoT devices, as they are considered primarily byproducts. Therefore, licensing based on copyright or database rights may not be suitable for such data sets (Graux 2021).

The shift away from relying on copyright and *sui generis* rights for data is crucial. When data is not protected under these rights and is instead offered as a service through subscription agreements, it grants the data provider complete contractual freedom. Data providers can determine the terms of use without the need to consider public interests or adhere to exceptions provided by copyright or database rights. In essence, moving towards data as a service allows data holders to undermine the delicate balance established by current intellectual property rights legislation (Graux 2021: 8).

In the case of *Ryanair v PR Aviation* (C-30/14, 9/11/2004), the CJEU ruled on the issue of *screen scraping*, a practice where online travel agencies collect flight data from Ryanair's website and display it on their own platforms. Ryanair claimed that this method infringed its database rights. The CJEU concluded that while the databases used by Ryanair to provide flight information to customers could be protected under the Database Directive, the act of screen scraping itself did not infringe those rights. The CJEU stated that screen scraping did not involve the extraction or re-utilization of a substantial part of the database, which is required for a breach of the rights. Therefore, screen scraping of flight data from Ryanair's website was considered lawful under the Database Directive.

However, the CJEU's decision in *Ryanair v PR Aviation* confirmed that Ryanair could impose contractual restrictions on third-party use of its database. Since the database was not eligible for copyright or *sui generis* right protection, the Database Directive's exceptions were not relevant. This meant that contractual freedom

prevailed, allowing Ryanair to set its own rules and regulations regarding data sharing, and granting it substantial control over how the database is used by others (Graux 2021: 8–9).

The shift towards data-as-a-service has undeniable benefits, but data users must be aware of their increased reliance on contractual terms set by data providers. When intellectual property rights are not applicable or not referenced in subscription agreements, the absence of such rights removes the balancing effect that legislations and courts aim to achieve. Data-as-a-service, facilitated by Application Programming Interfaces, web services, and online marketplaces, has become the dominant paradigm for data sharing, enhancing flexibility and usability. However, this trend leads to a growing dependence on contractual provisions that may not offer the same protections as intellectual property rights. While contractual freedom allows for tailored agreements, it also erodes assurances like exceptions for lawful use and transferability. To navigate this landscape, parties subscribing to dynamic data services should take some precautions. They should verify if the contract terms are independently drafted or follow balanced best practice templates. Additionally, they should assess whether the defined usage rights adequately cover their intended use, considering that exceptions allowed under intellectual property rights may not apply. Furthermore, they should ensure stability in the agreement, especially regarding data retention and access if the service becomes unavailable. While some concerns may be temporary, as laws evolve to accommodate dynamic data services, data subscribers should exercise caution until a more systematic and balanced legal framework is established (Graux 2021: 8–9).

To cope with the underlying issues, the European Union adopted Data Governance Act (EU Regulation 2022/868, 30/05/2022), which creates the processes and structures to facilitate data, and under legislation procedure is the Data Act (EU Commission Proposal for a Regulation COM/2022/68, 23/02/2022), which clarifies who can create value from data and under which conditions.

CONCLUSIONS

Contemporary generalizations on what is data and how it (to include AI) will impact humans and humanity fluctuate between two extremes. One approach, known as dataism, presents us with apocalyptic scenarios of the complete change of the worldview from anthropocentric to data-centric that will eventually lead to the elimination of humans. Others look skeptically at the possibility of transforming everything to data. Both approaches are not only speculative and impractical, but they also constitute a certain *deja vu* in the history of ideas.

The ontological generalization that *everything is data* resembles many other similar holistic claims (as Marxian economic holism or Schmittian political holism). Accordingly, this approach has its own scientific taboos or suspended questions as *what is algorithm?* or *what is data?* In the light of these stumbles of dataism, a skeptical approach may appear as more reasonable and scientifically adequate. However, it is not without its problems. Firstly, by entertaining this *calm down* mentality we may become too ignorant to the processes related to data and AI. Secondly, the skepsis is founded on another holistic paradigm – anthropocentrism – with all the corresponding claims and remaining unanswered questions.

This situation with two approaches resonates with the similar one in the context of legal philosophy – oscillation between so-called *jusnaturalism* and *juspositivism*. However, the example of legal philosophy may offer a certain way-out: replacing the question *what [the law is]* with the one *how [it functions],* otherwise called *the law in action* approach. Likewise, it would be more reasonable and practical to make the same shift of the modus of a question in relation to data and AI. We should turn our focus from *what it (data, algorithm, AI) is* to *how all this functions/ works* and *what very practical problems we should solve to make it function more efficiently and more contributing to the common good?* Practical regulatory realities should be the focus of scientists (reciprocally with politicians), starting from legal articulations of the concepts, the phenomenon of big data, tensions between privacy and openness of data, or potential of copyright and *sui generis* right to better shelter data under the umbrella of its protection.

By following this approach, we find that EU legislation defines data as digital representations of acts, facts, and information, with various categories serving specific purposes, including research, real-time updates, personal information, and energy-related data. The Data Governance Act and the proposal for the Data Act define data as any digital representation of acts, facts, or information, including sound, visual, or audiovisual recordings that form a comprehensive information repository. EU legislation classifies data into distinct categories, each subject to different legal regulations. Research data plays a fundamental role in scientific research, providing empirical evidence and validation. Dynamic data with real-time updates offers flexibility for adaptive analyses, particularly those generated by sensors, reflecting real-world changes. Personal data requires stringent privacy and security measures as it pertains to identifiable individuals and their specific attributes. Metering and consumption data play a vital role in energy policies, optimizing resource allocation and monitoring energy consumption patterns.

Additionally, data itself does not receive copyright protection within EU law, but organized databases may be eligible for certain rights. The Database Directive introduced two distinct rights, one based on originality of selection or arrangement

and the other known as the *sui generis* database right, incentivizing investment in databases. Moving away from relying solely on copyright and *sui generis* right for data is crucial. Data-as-a-service offers data providers contractual freedom, but data users must be cautious, as contractual provisions may not offer the same level of protection as copyright and the *sui generis* right. The European Union's Data Governance Act and Data Act aim to address data challenges, fostering a balance between data sharing and protection in the evolving data economy.

REFERENCES

Agamben, Giorgio. *Means without End. Notes on Politics*. Minneapolis, London: University of Minnesota Press, 2000.

Agamben, Giorgio. *The Open. Man and Animal*. Stanford: Stanford University Press, 2004.

Agreement on Trade-Related Aspects of Intellectual Property Rights, 15 April 1994, Marrakesh. Internet access <https://www.wto.org/english/docs_e/legal_e/27-trips_01_e.htm>.

Berne Convention for the Protection of Literary and Artistic Works, 9 September 1886, Paris. Internet access <https://www.wipo.int/wipolex/en/text/283698#P85_10661>.

The British Horseracing Board Ltd and Others v William Hill Organization Ltd., ECJ 9 November 2004, case C-203/02, ECR [2004] I-10415.

Commission Implementing Regulation (EU) 2023/1162 of 6 June 2023 on interoperability requirements and non-discriminatory and transparent procedures for access to metering and consumption data, OJ L 154, 2023 06 15, 10–40. Internet access <http://data.europa.eu/eli/reg_impl/2023/1162/oj>.

Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information (recast), OJ L 172, 2019 06 26, 56–83. Internet access <http://data.europa.eu/eli/dir/2019/1024/oj>.

Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases, OJ L 77, 1996 03 27, 20–28. Internet access <http://data.europa.eu/eli/dir/1996/9/2019-06-06>.

Fixtures Marketing Ltd v Organismos prognostikon agonon podosfairou AE (OPAP), ECJ 9 November 2004, case C-444/02, ECR [2004] I-10549.

Fixtures Marketing Ltd v Oy Veikkaus Ab, ECJ 9 November 2004, case C-46/02, ECR [2004] I-10365.

Fixtures Marketing Ltd v Svenska Spel AB, ECJ 9 November 2004, case C-338/02, ECR [2004] I-10497.

Football Association Premier League Ltd and Others v QC Leisure and Others; and Karen Murphy v Media Protection Services Ltd, ECJ 4 October 2011, joined cases C-403/08 and C-429/08, ECR [2011] I-9083.

Freeman, M.D.A. *Lloyd's Introduction to Jurisprudence*. 8th Edition. London, Cornwall: Thomson Reuters / Sweet & Maxwell, 2008.

Graux, Hans. Data sharing as a service: Will data services remove intellectual property rights from the picture, and at what cost? Luxembourg: Publications Office of the European Union, 2021. Internet access <https://op.europa.eu/en/publication-detail/-/publication/a6cdccee-315c-11ec-bd8e-01aa75ed71a1/language-en> [retrieved 2023 06 19].

Harari, Yuval Noah. *Homo Deus. A Brief History of Tomorrow*. London: Penguin Random House / Vintage, 2017.

Harari, Yuval Noah. Yuval Noah Harari argues that AI has hacked the operating system of human civilisation. *The Economist*, 2023 04 28. Internet access <https://www.economist.com/by-invitation/2023/04/28/yuval-noah-harari-argues-that-ai-has-hacked-the-operating-system-of-human-civilisation> [retrieved 2023 06 19].

Hugenholtz, Bernt. Against 'Data Property.' In *Kritika: Essays on Intellectual Property*, vol. 3. Eds. Hans Ullrich, Peter Drahos and Gustavo Ghidini. Edward Elgar, 2018, 48–71. Internet access <https://doi.org/10.4337/9781788971164.00010> [retrieved 2023 08 12].

Lumet, Sidney. *Network* [film]. Metro-Goldwin Mayer, 1976.

Popper, Karl Raimund. *The Logic of Scientific Discovery*. London and New York: Routledge, 2005. Internet access <http://philotextes.info/spip/

IMG/pdf/popper-logic-scientific-discovery.pdf> [retrieved 2023 08 12].

Proposal for a Regulation of the European Parliament and of the Council on harmonised rules on fair access to and use of data (Data Act), COM/2022/68 final, 2022 02 23. Internet access <https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=COM:2022:68:FIN>.

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), OJ L 119, 2016 05 04, 1–88. Internet access <http://data.europa.eu/eli/reg/2016/679/2016-05-04>.

Regulation (EU) 2022/868 of the European Parliament and of the Council of 30 May 2022 on European data governance and amending Regulation 2018/1724 (Data Governance Act) OJ L 152, 2022 06 03, 1–44. Internet access <http://data.europa.eu/eli/reg/2022/868/oj>.

Wacks, Raymond. *Understanding Jurisprudence. An Introduction to Legal Theory*. 3rd Edition. Oxford University Press, 2012.

WIPO Copyright Treaty, 20 December 1996, Geneva. Internet access <https://www.wipo.int/wipolex/en/text/295166>.

Tomas Berkmanas, Saulė Milčiuvienė
Vytauto Didžiojo universitetas, Lietuva

DUOMENYS TEISĖS FILOSOFIJOJE IR PRAKTIKOJE: ŠIUOLAIKINIAI SAMPRATOS IR REGULIAVIMO POKYČIAI

SANTRAUKA. Straipsnyje analizuojama duomenų – kaip neatsiejamo skaitmeninio amžiaus atributo – sampratos ir reguliavimo raida pastaraisiais metais. Aptariant remiamasi dviem kraštutiniais požiūriais. Pirmasis, vadinamas *deitaizmu*, pasižymi spekuliatyviu, dažnai radikaliu ir šokiruojančiu holizmu. Kitas ne mažiau radikalus – skepticizmas, inertiškas ir problemiškas. Straipsnyje duomenų (kaip reiškinio) atžvilgiu siūloma alternatyvi filosofinė pozicija, kuri būtų analogiška sociologinės jurisprudencijos ar teisinio realizmo prieigoms teisės filosofijoje. Analizuojami atskiri aktualūs reguliavimo aspektai – pradedant nuo bandymų teisės aktuose apibrėžti „duomenis" ir baigiant galimais tektoniniais lūžiais autorių ir *sui generis* teisių paradigmoje. Pabrėžiama, kad nors apie duomenų turėtojų turtinių teisių apsaugą diskutuoja tik nedidelė mokslininkų grupė, šiuo metu duomenys įgyja vis didesnę ekonominę vertę ir yra vis svarbesni ekonomikos augimui ir gerovei. Straipsnyje kviečiama ieškoti interesų pusiausvyros tarp duomenų turėtojų ir jų naudotojų bei visuomenės. Dėl to reikėtų reikšmingų pokyčių reguliuojant ir autorių, ir *sui generis* teises, nes duomenų reguliavimas labai retais atvejais tampa autorių teisių ir *sui generis* teisių apsaugos objektu.

RAKTAŽODŽIAI: duomenys, deitaizmas, autorių teisės, skaitmenizacija.