THE ROLE OF LICENSING IN DATA FAIRIZATION

Thomas Margoni[1] *& Luca Schirru*[2]

ABSTRACT

Data FAIRization and "open" licenses such as Creative Commons have a common denominator: copyright law. In fact, most of the licenses used to reshare content are based on an underlaying copyright to operate, and when this is absent (or a certain use is outside copyright's exclusivity, such as in the case of exceptions and limitations), the license is not activated. This aspect is central in the discussion about data FAIRization since most data, just like mere facts and ideas, have been intentionally left outside the scope of copyright. In this chapter we try to offer a concise but legally sound overview of this relationship and identify areas of future reform.

Keywords: FAIR, Data, Copyright, Licenses, Creative Commons, Data Governance

[1] Research Professor of Intellectual Property Law at the Faculty of Law and Criminology, KU Leuven, and member of the Board of Directors at the Centre for IT & IP Law (CiTiP).

[2] Postdoctoral researcher at Centre for IT & IP Law (CiTiP), Faculty of Law and Criminology, University of Leuven (KU Leuven), Belgium; luca.schirru@kuleuven.be. His research is funded under the Skills4EOSC project. Skills4EOSC has received funding from the European Union's Horizon Europe research and innovation Programme under Grant Agreement No. 101058527.

ABSTRACT **1**

INTRODUCTION **2**

I. FAIR DATA 5

II. BRIEF OVERVIEW OF RIGHTS AND LIMITATIONS 8

A. Copyrights *8*

B. Sui Generis Database Rights *11*

C. Additional Regulatory Levels *13*

III. OPEN LICENSES 15

A. Creative Commons *16*

B. Other Data Access Regimes *20*

CONCLUSION **22**

INTRODUCTION

It is usually said that data is the new oil. However, despite the fascination of the 'value' component of the analogy, the same can't be said regarding the good itself, its nature. Oil, as any finite and tangible goods, is rival to consumption, which means that its usage will directly reduce the amount left for others.[3] Another characteristic of tangible goods is the ease of exclusion, meaning that it is much easier (or cheaper) to exclude someone who does not pay from enjoying a specific good (known as the "free rider" problem) when this has a physical embodiment (e.g., a music cassette), than when the information is purely immaterial (e.g., an mp3). Data may be better understood as a public good, or, as categorized by Benjamin Coriat, as a knowledge common, which has the following characteristics: is non-rival to consumption and non-excludable.[4] Due to these characteristics, it has been argued that data governance should be 'oriented not towards the *conservation* of resources' since they are not depletable, but towards their *enrichment* and *growth*.[5]

It may also be argued that the impact that data is having on our society,

---

[3] See E Ostrom, 'Private and Common Property Rights' (1999) <https://www.sfu.ca/~allen/common%20property.pdf>.

[4] B Coriat, 'From Natural-Resource Commons to Knowledge Commons: Common Traits and Differences' (2011) LEM Papers Series 2011/16, 22 <https://ideas.repec.org/p/ssa/lemwps/2011-16.html>. See also W Fisher, Theories of Intellectual Property (1987) <https://cyber.harvard.edu/people/tfisher/iptheory.pdf>. P Drahos, A Philosophy of Intellectual Property (ANU eText, 2016).

[5] B Coriat (n 3) 13-19.

on the economy, on cultural and technological advances, is in fact much more pervasive than anything else observed in modern human history. The datafication of society (and of economy, culture, technology, etc) is a recent example of how digital information, expressed now in data, influences the classical categories of knowledge and power.[6] If "everything" is becoming data, regulating data means by extension regulating "everything", or at least much more than what one may have initially considered. Perhaps, aware of this changing function in the regulation of data, recent EU legislative developments such as the Data Governance Act and the proposed Data Act seem to advance a new paradigm for the regulation of data, a paradigm that is not (or not exclusively) based on property rights such as copyright, but on some sort of private-public governance model.[7]

Remaining within the field of property rights and in particular copyright, it should be noted that despite the many differences between tangible and intangible objects, the power to exclude, the celebrated *jus excludendi alios* characteristic of property rights over tangible goods, can be artificially reconstituted by Intellectual Property Rights (IPRs) over intangible artifacts such as information and data.[8]

Yet, it is often conceptually problematic to simply transfer the same institutional structures adopted for rival and tangible goods to deal with knowledge governance. Collaborative projects like Wikipedia and Free, Libre and Open Source Software (FLOSS) offer evidence of how access and collaboration instead of exclusion and competition can be conducive to economic, cultural, and social development.[9] Accordingly, it seems at least disputable that under the framework of a knowledge and data economy, the traditional tenet of intellectual property, i.e., the artificial creation of exclusionary mechanisms, offer the optimal amount of incentive to advance socially desirable goals.[10] FAIR principles (findability, accessibility,

---

[6] UA Mejias & N Couldry, Datafication (2019) 8(4) *Internet Policy Review*[online] <https://policyreview.info/concepts/datafication> accessed: 20 Dec. 2022.

[7] See C. Ducuing, T, Margoni, L. Schirru, D. Spajic, T. Lalova-Spinks, L. Stähler, E. Bayamlıoğlu, A. Pétel, J. Chu, B. Peeters, A. Christofi, J. Baloup, M. Avramidou, A. Benmayor, T. Gils, E. Kun, E. De Noyette, and E. Biasin, White Paper on the Data Act Proposal (October 26, 2022), <https://ssrn.com/abstract=4259428> accessed 20 Dec., 2022. J Baloup, E Bayamlıoğlu, A Benmayor, C Ducuing, L Dutkiewicz, T Lalova, Y Miadzvetskaya, B Peeters, 'White Paper on the Data Governance Act' (2021) CiTiP Working Paper, 38 <https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3872703> accessed 10 October 2022.

[8] Fisher (n 3).

[9] Séverine Dusollier, 'Propriété Inclusive ou Inclusivité' in Marie Cornu; Fabienne Orsi; Judith Rochfeld (eds) *Dictionnaire des biens communs* (Presses Universitaires de France), pp.983 - 987,

[10] JE Stiglitz, 'Knowledge as a Global Public Good' in Inge Kaul, Isabelle Grunberg, Marc Stern (eds.), *Global Public Goods* (OUP, 1999) ("Initial knowledge is a key input into the production of further knowledge, and the design of the patent system can thus affect dramatically the overall pace of innovation. An excessively broad patent system (e.g. with

interoperability and reusability) seem to replicate some of the dynamics present in the knowledge commons and arguably contribute to foster access, circulation and re-use of data.[11] Or in other words, enrichment and growth.

The market failure argument that is often employed to justify (intellectual) property rights in the field of data also appears in need of clearer empirical support. Whereas property is traditionally justified by a need to avoid overuse of a finite and depletable pool of tangible resources, intellectual property is justified by the different need to avoid underproduction of a resource that is not depletable, but which still needs to be produced in societally optimal amounts.[12] Data, on the contrary, have been long considered mainly as an intermediate- or by-product (think of Internet of Things data for instance) and their production is often incidental to a distinct main activity. That said, whereas the creation of new data is frequently – albeit not always – incidental to a main activity, their verification or presentation may still require substantial efforts. When they do, this extra effort may very well represent the added value distinguishing the original, raw data from the verified or "cleaned" version that often demands moral or economic recognition.

Although it is true that some data under certain circumstances may be regulated by copyright, their nature as a non-rival good with the potential to self-enrich through use and circulation raises the question of how to fairly balance the multiple fundamental interests related to data access.[13] Licensing, and in particular open licensing, appears as a practicable option towards the FAIRification of data and the achievement of a data commons.[14] Yet, licensing, and in particular some of the most common open license types such as Creative Commons (CC), presupposes an underlying copyright on which they root their legal basis. As it will be explained, however, only certain data under certain circumstances trigger copyright protection.

In this chapter we shall refer to 'data' to address non-personal data, which

---

long-lived patents of broad scope) can raise the price of one of the most vital inputs into the innovative process and thus reduce the pace of follow-on innovations, even as it may provide returns to those making the original innovation. As a result, the overall pace of technical progress may be slowed.") (footnote omitted)

[11] See M Wilkinson, M Dumontier, I Aalbersberg, and others, 'The FAIR Guiding Principles for scientific data management and stewardship' (2016) 3 Sci Data 160018, 4 <https://doi.org/10.1038/sdata.2016.18> accessed 21 November 2022.

[12] Fisher (n 3). See also Coriat (n 3).

[13] Coriat (n 3).

[14] European Commission, Directorate-General for Research and Innovation, M Senftleben, 'Study on EU copyright and related rights and access to and reuse of data, Publications Office of the European Union' (2022) 65 <https://data.europa.eu/doi/10.2777/78973> ("**Considering this configuration of use privileges for scientific research, the conclusion is inescapable that there is an imbalance in the EU acquis.** Whereas EU copyright, related rights and sui generis database law offers right holder broad, robust exclusive rights, researchers must content themselves with narrow, unreliable research exceptions.").

may, or may not, be protected by copyright and cognate rights. In particular we will not discuss personal data, given the fact that they follow a rather different regulatory approach (e.g., GDPR), which is not predicated on property rights, but on a different framework (e.g., information self-determination). When it comes to licenses, we will focus on some of the most well-known models to regulate copyright-related content, as is the case of the Creative Commons Public License version 4.0 (CCPL) and Creative Commons Zero (CC0). Finally, whereas FAIR Data stands for findable, accessible, interoperable and reusable data, considering the focus of this chapter on addressing the role of licensing in the FAIRization of data, the reusability and interoperability principles will attract greater attention.[15]

The first part of this chapter will address the main aspects of FAIR and what it means for data to be findable, accessible interoperable and reusable. Then we will briefly address the relationship between data and copyright law in Part II. Part III will present some of the licensing models used for non-personal data, focusing on Creative Commons licenses, and other data access regimes.

I. FAIR DATA

FAIR data are data made available according to four main principles: findability, accessibility, interoperability and reusability.[16] As mentioned above, this chapter will focus on non-personal data and will specifically look at the relationship of non-personal data, copyright licenses and FAIR principles. The "Turning FAIR into reality"[17] analysis and other important sources, like the GO Fair initiative, are very helpful to better understand how data can comply with the principles of findability, accessibility, interoperability, and reusability, as seen below:

Principle	**Definition** (https://www.go-fair.org/fair-principles/)	**Description** (*Turning FAIR into reality: Final Report and Action Plan of FAIR Data, European Commission, 2018, 19-20 (footnotes removed)*
Findability	"F1. (Meta)data are assigned a globally unique and persistent identifier	"Data are **Findable** when they are described by sufficiently rich metadata and registered or indexed in a searchable resource

[15] A Landi and others, 'The "A" of FAIR – As open as possible, as closed as necessary' (2020) 2 Data Intelligence, 47 <doi: 10.1162/dint_a_00027>, accessed 21 November 2022.

[16] M Wilkinson, M Dumontier, I Aalbersberg, and others (n 11)

[17] S Collins and others, European Commission, European Commission Expert Group on FAIR Data, Turning FAIR into reality: Final Report and Action Plan from the European Commission Expert Group on FAIR Data (2018) <https://ec.europa.eu/info/sites/default/files/turning_fair_into_reality_1.pdf> accessed 21 November 2022.

	F2. Data are described with rich metadata (defined by R1 below) F3. Metadata clearly and explicitly include the identifier of the data they describe F4. (Meta)data are registered or indexed in a searchable resource"	that is known and accessible to potential users. Additionally, a unique and persistent identifier should be assigned such that the data can be unequivocally referenced and cited in research communications."
Accessibility	"A1. (Meta)data are retrievable by their identifier using a standardised communications protocol A1.1 The protocol is open, free, and universally implementable A1.2 The protocol allows for an authentication and authorisation procedure, where necessary A2. Metadata are accessible, even when the data are no longer available"	**"Accessible** data objects can be obtained by humans and machines upon appropriate authorisation and through a well-defined and universally implementable protocol. In other words, anyone with a computer and an Internet connection should be able to access at least the metadata. It is important to emphasise that Accessible in FAIR does not mean Open without constraint. Accessibility means that the human or machine is provided - through metadata - with the precise conditions by which the data are accessible and that the mechanisms and technical protocols for data access are implemented such that the data and/or metadata can be accessed and used at scale, by machines, across the web."
Interoperability	"I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation. I2. (Meta)data use vocabularies that follow FAIR principles I3. (Meta)data include qualified references to other (meta)data"	"**Interoperable** data and metadata are described in the FAIR principles as those that use a formal, accessible, shared, and broadly applicable language for knowledge representation. They use vocabularies which themselves follow the FAIR principles, and they include qualified references to other data or metadata. What this describes is semantic interoperability. […]It is not only semantics but also technical and legal interoperability. Technical interoperability means that the data and related information is encoded using a standard that can be read on all applicable systems. In FAIR, legal interoperability falls under the principle that data should be 'Reusable'."

Reusability	"R1. (Meta)data are richly described with a plurality of accurate and relevant attributes R1.1. (Meta)data are released with a clear and accessible data usage license R1.2. (Meta)data are associated with detailed provenance R1.3. (Meta)data meet domain-relevant community standards"	"For data to be **Reusable**, the FAIR principles reassert the need for rich metadata and documentation that meet relevant community standards and provide information about provenance. This covers reporting how data was created (e.g. survey protocols, experimental processes, information about sensor calibration and location) and information about data reduction or transformation processes to make data more usable, understandable or 'science-ready'. As shown in the example of the DOBES case study (Fig. 3), open community-endorsed formats also play a key role in reusability. The ability of humans and machines to assess and select data on the basis of criteria relating to provenance information is essential to data reuse, especially at scale. Reusability also requires that the data be released with a 'clear and accessible data usage license': in other words, the conditions under which the data can be used should be transparent to both humans and machines."

Among the four principles mentioned above, two are central in the analysis developed for this chapter: reusability and interoperability. Regarding the former, reusability usually refers to the possibility to reuse the licensed material, which, in the field of copyright law commonly triggers the rights of reproduction, distribution, adaptation and communication to the public. Regarding interoperability, it is worth clarifying that 'Interoperability' may assume different forms and concepts according to the context used (technical, legal etc).[18] We will limit our focus on legal interoperability.[19] In particular, we employ the definition of 'Legal Interoperability' adopted in the study 'Legal Interoperability and the FAIR Data Principles', which is the following:

> (Interoperability) concerns the ability to combine datasets from multiple sources without conflicts among the restrictions that each dataset carries (i.e., support of one restriction inherently negates

[18] Ohad Graber-Soudry and others, 'Legal Interoperability and the FAIR Data Principles (1.0)' (2021) 19 <https://doi.org/10.5281/zenodo.4471312>

[19] European Commission Expert Group on FAIR Data (n 17)

> support of another). […] Legal interoperability also concerns situations where regulatory or policy measures restrict the disclosure of data, or that datasets may be made available only in certain jurisdictions or under certain conditions. Examples include legal restrictions based on intellectual property law, national security, the protection of endangered species or privacy regulations, such as the GDPR.[20]

It appears evident that reusability is often, if not always, a precondition to interoperability. However, considering the potential overlaps between legal interoperability and reusability, this study will be addressing them together, unless otherwise noted, through the lenses of the exclusive rights (and their exceptions) that may cover non-personal data, when these data are protected by copyright or selected related rights.[21] Therefore, the next section will provide a brief overview of these exclusive rights, mainly copyrights and Sui Generis Database Rights (SGDR), and their limitations.

II. BRIEF OVERVIEW OF RIGHTS AND LIMITATIONS

A. Copyright

Under the umbrella of Intellectual Property Rights (IPRs) there are several exclusive rights that apply to different products, processes or activities. Signs that may distinguish goods and services from different undertakings can be considered trademarks.[22] The outer appearance of products of industrial design or applied art (like the design of a phone) may be subject to design rights.[23] Inventions, either products or processes that are "new, involve an inventive step and are capable of industrial application" can be patented.[24] Information that is secret and has commercial value due to its

---

[20] Graber-Soudry and others (n 18) 19 (footnotes omitted)

[21] Wilkinson, Dumontier, Aalbersberg, and others (n 11) 4.

[22] World Trade Organization, Agreement on Trade-Related Aspects of Intellectual Property Rights (15 April 1994, Annex 1C) 1869 UNTS. 299 (TRIPS Agreement), art 15. ("Protectable Subject Matter. 1. Any sign, or any combination of signs, capable of distinguishing the goods or services of one undertaking from those of other undertakings, shall be capable of constituting a trademark. Such signs, in particular words including personal names, letters, numerals, figurative elements and combinations of colours as well as any combination of such signs, shall be eligible for registration as trademarks. Where signs are not inherently capable of distinguishing the relevant goods or services, Members may make registrability depend on distinctiveness acquired through use. Members may require, as a condition of registration, that signs be visually perceptible.")

[23] Id, art 25 ("Requirements for Protection. 1. Members shall provide for the protection of independently created industrial designs that are new or original. Members may provide that designs are not new or original if they do not significantly differ from known designs or combinations of known design features. Members may provide that such protection shall not extend to designs dictated essentially by technical or functional considerations.").

[24] Id, art 27 ("Patentable Subject Matter. 1. Subject to the provisions of paragraphs 2 and 3, patents shall be available for any inventions, whether products or processes, in all fields of technology, provided that they are new, involve an inventive step and are capable of industrial application").

secrecy may qualify as trade secrets.[25] In this chapter the focus will be on copyright which protects original expressions in the literary and artistic fields. These include both traditional creations such as books, poems, pamphlets, musical compositions, paintings, sculptures, or works of photography, as well as more modern productions such as computer software, databases or digital transmissions.

It is important at this point to note, that "data" in its scientific sense does not necessarily translate as "data" in the legal sense. From the point of view of the law, only works of authorship are protected by copyright. These include the aforementioned original literary and artistic expressions. A category connected to copyright, called related rights or neighboring rights, has the objective to offer a – usually thinner – degree of protection to deserving investments in an act somehow connected to the creative process, but distinct. This is the case of sound (and in the EU, film) recordings, broadcasting, performances (although the latter is often considered closer to copyright), among others.

Accordingly, it is often said that ideas, facts and data as such – which do not qualify as original expressions – are not protected by copyright and only eventually and to a limited extent by related rights. [26] In particular, the so called idea-expression dichotomy (and the connected fact-expression dichotomy) represent the balance identified by the legislator between public access and private initiative, between creative reuse and return on investments, between the ability to freely learn from the basic brinks of knowledge and the need to protect creative activity from lavish imitations.

Databases deserve closer scrutiny. The TRIPs agreements provide in art. 10(2) that databases can be considered as copyrighted works if the selection or arrangement of their content is considered as an "intellectual creation". The second part of art. 10(2) provides that "[s]uch protection [...] shall not extend to the data or material itself" and "shall be without prejudice to any copyright subsisting in the data or material itself".[27] Databases covered by art. 10 (2) of TRIPS and, in Europe, under the first part of Directive 96/9/EC ("Database Directive"),[28] are commonly referred to as "original databases" because, in order to be protected by copyright law, their arrangement or selection must be original. Illustratively, arranging a database only in

[25] Directive (EU) 2016/943 of the European Parliament and of the Council of 8 June 2016 on the protection of undisclosed know-how and business information (trade secrets) against their unlawful acquisition, use and disclosure, OJ L 157/1 (Trade Secret Directive).

[26] Berne Convention for the Protection of Literary and Artistic Works (adopted 14 July 1967, entered into force 29 January 1970) 828 UNTS 221 (Berne Convention), art. 2(1) and 2(3).

[27] In other words, if it is a database of photographs, copyrights on the photos shall subsist – and be managed by, e.g., the photographer – regardless of the specific legal treatment granted to the database.

[28] Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of databases [1996] OJ L 77/20 (Database Directive).

chronological or alphabetical order should not suffice, because there is no intellectual creation behind this structure. However, a database where the selection or arrangement follow the original choices of the author may very well be protected by copyright. It should be noted however that in this case it is the structure of the database which is protected not the content. So, for instance, absent any other legal entitlement, it would be lawful to reuse the elements (the data) from the database without infringing copyright, as long as the structure (the selection or arrangement) is not copied. In the EU, a *sui generis* database right (SGDR) offers an additional layer of protection in similar cases and is discussed in the next section.

Another fundamental characteristic of copyright is territoriality, meaning that, despite international and regional agreements intended to coordinate and harmonize rules, copyright remains an object regulated by national law. An example of this principle is that, whereas international sources set the time of protection for copyright to a minimum of 50 years *pma*, national (or EU) law can determine within this framework to extend protection beyond that, ie., to 70 years, as in the US and EU. Once protection elapses, works enter the public domain and are free for everyone to be used (although moral rights may endure).

Copyright identifies specific rights of economic exploitation (usually the right to reproduce, adapt, distribute and communicate to the public), as well as certain limitation or exceptions. Facts and news of the day are some of the examples of what is expressly excluded from the scope of copyright.[29] Limitations and exceptions, similarly to the case of the term of protection, may vary according to the national legislation, but some of them are mandatory by the signatories of certain international agreements, for instance the Berne Convention and the quotation exception.[30] The same may happen when it comes to the EU directives, that can provide mandatory or optional exceptions to be adopted in national law, as is the case of Directive 2001/29/EC which provides for one mandatory and 20 optional exceptions and limitations, or Directive (EU) 2019/790 on Copyright and Related Rights in the Digital Single Market and the (mandatory) exceptions for Text and Data Mining (TDM).[31]

---

[29] Berne Convention (n 26) art. 2(8) "The protection of this Convention shall not apply to news of the day or to miscellaneous facts having the character of mere items of press information.". See Thomas Margoni, Martin Kretschmer, A Deeper Look into the EU Text and Data Mining Exceptions: Harmonisation, Data Ownership, and the Future of Technology 71(8) *GRUR International* 685, 689-690 (2022) https://doi.org/10.1093/grurint/ikac054

[30] Berne Convention (n 26) art. 10(1) "(1) It shall be permissible to make quotations from a work which has already been lawfully made available to the public, provided that their making is compatible with fair practice, and their extent does not exceed that justified by the purpose, including quotations from newspaper articles and periodicals in the form of press summaries."

[31] Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives

B. Sui Generis Database Rights (SGDR)

In the EU, the Database Directive also provides for a *sui generis* right for databases, independent and cumulable with copyright.[32] Whereas copyright is granted upon the original arrangement or selection of the contents, "the object of this sui generis right is to ensure protection of any investment in obtaining, verifying or presenting the contents of a database for the limited duration of the right".[33] Recital 40 of the Database Directive further explains that the investment "may consist in the deployment of financial resources and/or the expending of time, effort and energy". Recitals of the Database Directive[34] and decisions by the CJEU as the ones in the *Fixtures Marketing*[35] and *The British Horseracing Board*[36] cases help to provide some additional light on what would configure as an investment in the obtainment, verification, and presentation of the contents of the database.

An important remark from the *British Horseracing Board* case refers to the real purpose of the Database Directive, which is (was?) "to promote the establishment of storage and processing systems for existing information and not the creation of materials capable of being collected subsequently in a

---

96/9/EC and 2001/29/EC [2019] OJ L 130/92 (CDSM), art 3. See Margoni & Kretschmer (n 29)

[32] Database Directive (n 28) art. 7(1) ("Object of protection 1. Member States shall provide for a right for the maker of a database which shows that there has been qualitatively and/or quantitatively a substantial investment in either the obtaining, verification or presentation of the contents to prevent extraction and/or re-utilization of the whole or of a substantial part, evaluated qualitatively and/or quantitatively, of the contents of that database.").

[33] Recital 40 of the Database Directive (n 28).

[34] See e.g., Recital 19 of the Database Directive (n 28): ("(19) Whereas, as a rule, the compilation of several recordings of musical performances on a CD does not come within the scope of this Directive, both because, as a compilation, it does not meet the conditions for copyright protection and because it does not represent a substantial enough investment to be eligible under the sui generis right;").

[35] Case C-338/02 *Fixtures Marketing Ltd v Svenska Spel AB* [2004] ECR I-10497, paras 27-28. ("The expression 'investment in … the … verification … of the contents' of a database must be understood to refer to the resources used, with a view to ensuring the reliability of the information contained in that database, to monitor the accuracy of the materials collected when the database was created and during its operation. The expression 'investment in … the … presentation of the contents' of the database concerns, for its part, the resources used for the purpose of giving the database its function of processing information, that is to say those used for the systematic or methodical arrangement of the materials contained in that database and the organisation of their individual accessibility. 28Investment in the creation of a database may consist in the deployment of human, financial or technical resources but it must be substantial in quantitative or qualitative terms. The quantitative assessment refers to quantifiable resources and the qualitative assessment to efforts which cannot be quantified, such as intellectual effort or energy, according to the 7th, 39th and 40th recitals of the preamble to the directive.").

[36] Case C-203/02 *The British Horseracing Board Ltd and Others v William Hill Organization Ltd* [2004] ECR I-10415, para 31.

database".[37] In other words: it is clear that the Database Directive does not aim to foster the creation - and legal protection - of data itself.

An important, and often misunderstood aspect of databases is that the data contained in a database are not protected by the (eventual) copyright in the database. Data are also not directly protected by the SGDR. In fact, the single datum or insubstantial amounts of data do not fall in the scope of the SGDR. It is only when a substantial investment in the obtaining, verification or presentation of the data has been made, that substantial amounts of data in the database are protected against extractions and reuse. Created data (or better data where the eventual substantial investment has been made in the creation phase) are also not protected, in order to avoid anti-competitive situations in information markets (such as the so-called single source databases). Only if data have been collected (and arguably already existed) and the investment is accordingly placed, protection arises.[38]

All this is of course without prejudice to any eventual copyright already existing in the individual elements constituting the database. Think of a database of journal articles, which usually attract copyright protection in their own right. This would be a potentially complex situation where on the same database might insist three different layers of protection: the copyright on the database structure, the SGDR on the substantial investment, the copyright or related rights on the individual elements (the journal articles) of the database. Situations where each one of these three layers of protection follow different ownership, protection and reusability routes, cannot be excluded. Consequently, it becomes paramount that licenses, especially those granting generous reusability rights in order to advance FAIR principles in data, duly address this complexity.

C. Additional regulatory levels

From what was previously presented in this chapter, in addition to the existing protections to databases (copyright(s) and SGDR), it should be highlighted that data-intensive businesses and research organizations have to deal with several layers of protection, either legislative, contractual or technological.[39] On the legislative layer, IP rights are the main element, since

---

[37] Id.

[38] See Margoni & Kretschmer (n 29) 699 ("However, whereas it has always seemed the correct reading of the SGDR that it cannot offer protection to machine generated data to the extent that it is generated data […], machine generated data acquired by a third party through a substantial investment (e.g. payment of money), could become protected by an SGDR that rewards not the one who invested in the creation, but the third party who invested in the obtaining of this data. This has arguably been the proper reading of the creation versus obtaining dichotomy, whereby the 1996 legislator and the subsequent CJEU case law have attempted to avoid anti-competitive situations such as those originated by so-called single-source databases.").

[39] AR Souza, L Schirru and M Alvarenga, 'Copyright and text mining in the fight against

use of data and databases may be subject to copyright and SGDR as seen, but also by other IP rights, most likely patents,[40] and trade secrets.[41] The technical layer would be better represented by Technological Protection Measures (TPMs) and Digital Rights Management (DRMs).[42] A contractual layer is represented not only by licenses, which will be discussed later, but also by so called Terms of Use that, beyond the specific underlying rights, may be used to limit or prohibit the access and/or use of data.[43]

While it seems an already complex framework of exclusivity-based rules, with dubious effects on research and innovation,[44] until relatively recently, the idea to extend or create ex novo a right protecting machine generated data (which currently are likely excluded on the basis of the creation v. obtaining dichotomy) was explored.[45] While these data may be of value for certain data intensive industries, they may fall far away from the subject of protection of IP. And not only because they are machine-generated, but also because they will probably represent mere facts and data, which are excluded categories.[46] Even putting the IP rational on the side for a moment, it is far from clear that from an industrial organization perspective restricting access to such data via property rights may be the optimal choice, considering the importance of

---

Covid-19 in Brazil' (2020) 16 (2) Liinc em Revista. AR Souza, L Schirru and M Alvarenga, 'Covid-19, Text and Data Mining and Copyright: The Brazilian Case' in Y Pai, J van Weelde and W Zaman (eds) *WIPO-WTO Colloquium Papers* 11 (WIPO Academy and WTO IP, Government Procurement and Competition Division, 2020).

[40] MW Carroll, 'Sharing Research Data and Intellectual Property Law: A Primer' (2015) 13 (8) PLoS Biol e1002235, 2 ("Patents may apply to some forms of data, but the more common issue is that data sharing may have implications for the acquisition of patent protection in inventions that arise from research."), <doi:10.1371/journal.pbio.1002235> accessed 21 November 2022.

[41] H Zech, 'A legal framework for a data economy in the European Digital Single Market: rights to use data' (2016) 11 (6) Journal of Intellectual Property Law & Practice 460, 6 ("Although data may qualify as trade secrets, the protection does not lead to a real data use right. Especially with Big Data matters this leads to problems."), < https://doi.org/10.1093/jiplp/jpw049>.

[42] See Souza, Schirru and Alvarenga, 'Copyright and text mining in the fight against Covid-19 in Brazil' (n 36) and Souza, Schirru and Alvarenga, 'Covid-19, Text and Data Mining and Copyright: The Brazilian Case' (n 39).

[43] Id.

[44] See Souza, Schirru and Alvarenga, 'Copyright and text mining in the fight against Covid-19 in Brazil' (n 36). Souza, Schirru and Alvarenga, 'Covid-19, Text and Data Mining and Copyright: The Brazilian Case' (n 39). G Dosi and JE Stiglitz, 'The Role of Intellectual Property Rights in the Development Process, with Some Lessons from Developed Countries: An Introduction' (2013) LEM Working Paper Series, No. 2013/23, 22 <https://www.econstor.eu/bitstream/10419/89516/1/771928769.pdf>.

[45] PB Hugenholtz, 'Against "data property"' in H Ullrich, P Drahos, G Ghidini (eds), *Kritika: Essays on Intellectual Property* (Edward Elgar Publishing, 2018).

[46] See Berne Convention (n 26) art. 2 (8) ("The protection of this Convention shall not apply to news of the day or to miscellaneous facts having the character of mere items of press information."). Agreement on Trade-Related Aspects of Intellectual Property Rights, art. 9 (2) (as amended by the 2005 Protocol amending the TRIPS Agreement) ("Copyright protection shall extend to expressions and not to ideas, procedures, methods of operation or mathematical concepts as such.")

anticompetitive effects that data monopolies have on trade and innovation.[47]

The incentive theory is a strong business justification for the creation and enforcement of IPRs. However, empirical literature shows that different sectors behave differently and that stronger IP rights, beyond a certain threshold, harm rather than benefit markets and innovation.[48] But even considering a specific IPR like patents or copyrights , it is difficult to assume that a one-size-fits-all solution will work.[49] For instance, the use of data on the health sector is subject to different norms and ethical standards than the use of data in seismology research.[50] Property rights, given their absolute characteristics, are not well equipped to address this sector specific nuances.

Furthermore, it may be argued that there is no need to create a new property right to incentivize the production of data as it was until recently proposed. This is not only because of the lack of literature evidencing a clear market failure,[51] but also because it goes against the fundamental rationale behind the data economy.[52] On the latter, Drexl et al (2017, p.6) propose that

---

[47] J Drexl and others, 'Technical Aspects of Artificial Intelligence: An Understanding from an Intellectual Property Law Perspective' (2019) Max Planck Institute for Innovation and Competition Research Paper Series – Research Paper No. 19-13, 7 ("In the case of supervised learning*, the model is 'told' during the optimisation process what the training data it is confronted with represents. Hence, it knows whether the prediction it made in the optimisation process was right or wrong. For doing that, it necessitates labelled data, for which additional investments (for instance human involvement) are required."). Id, at. 5 ("A machine learning model might be used to recognise cats in pictures. In the case of supervised machine learning, the model is trained on a data set containing labelled data (i.e., each picture is accompanied by the information whether there is a cat in the picture), allowing it to become more accurate. Once the training is completed, the model should in principle be capable of recognising from an unlabelled picture whether a cat appears in it (output). This model could finally be implemented in a self-driving car, allowing it, for instance, to brake when confronted with a cat (application).") <https://ssrn.com/abstract=3465577>.

[48] Dosi and Stiglitz (n 46) .

[49] Id.

[50] Id.

[51] J Drexl and others, 'Position Statement of the Max Planck Institute for Innovation and Competition of 26 April 2017 on the European Commission's "Public consultation on Building the European Data Economy"' (2017) Max Planck Institute for Innovation & Competition Research Paper No. 17-08, 9 ("...investment by itself does not justify a new IP right. Any new IP right would be in need of being justified by an identifiable market failure to which it responds.") <https://ssrn.com/abstract=2959924>.

[52] W Kerber, 'A New (Intellectual) Property Right for Non-Personal Data? An Economic Analysis' (2016) 11 GRUR Int 989, 8-9 ("In the discussion about data property so far, nobody has claimed that we have a general incentive problem in the digital economy as regards the collection, production, and analysis of data. To the contrary, the empirical fact of the massive and often simple production of huge amounts of data and their analysis seems to be one of the most important characteristics of Big Data and the digital economy"). Id, at 19 ("The basic idea of Big Data applications is to use data from many different sources, combine them, analyze them, derive new data, which again can be used for further analyses in combination with other data. One of the characteristics of the data economy is that data can often be used for analyses in many different contexts and for solving many different problems. It has therefore been claimed that in order to develop innovations in the digital

"[a] new property rights system should only be introduced if such a right will improve the functioning of the data economy."[53] The most recent EU legislative developments in this area (some still in a draft phase) point towards a different direction, as exemplified by Art. 35 of the proposed Data Act which limits the operativity of SGDR by clarifying that it does not apply to IoT data. It could be argued that the road identified by the most recent EU legislative and policy documents, rather than in the direction of data property, seem to be in the direction of European common data spaces.[54]

III. OPEN LICENSES

When "data" is protected by copyright and/or SGDR as outlined above, there are two main avenues to reusability: (i) reuse within the scope of statutory permissions, e.g, exceptions and limitations to copyright provided by law; or (ii) reuse within the scope of a contractual permission, e.g., copyright licenses.

Regarding item (ii), a common issue connected with licenses is that they may need to be individually negotiated by users (licensees) and right holders (licensors), a situation that often is affected by power imbalances in the contractual transaction (think of business to consumer contracts and the field of consumer law intended to protect the weak party) and the ability to negotiate fair contractual terms, especially when the ultimate goal is to obtain permissions that allows reuse and interoperability as mandated by FAIR principles.

So called "open" licenses purports to assist creators, users, developers and producers ("prosumers") to achieve a fairer, transparent and efficient bargain by offering "pre-packaged" and modular (in the case of CC) public

---

economy it is essential to have easy access to many different kinds of data and that therefore all impediments to the "free flow of data" should be eliminated as far as possible. From this perspective new exclusive IPRs on data might lead to additional barriers to this "free flow of data" and hamper competition and innovation in the digital economy") <https://ssrn.com/abstract=2858171>. See also A Gärtner and K Brimsted, 'Let's talk about Data Ownership' (2017) 39 (8) E.I.P.R 461, 464 ("The economic discussion is underway, but so far it seems that no one sees an urgent need for creating a right to data. One reason for granting IP rights is to incentivise the creator to innovate, leading to the generation and disclosure of more intellectual property.35 However, there is no apparent need to incentivise the collection and analysis of data.36").

[53] See also Kerber (n 54) 3 ("This paper will draw the conclusion that - on the basis of our current preliminary knowledge - a new IPR on data is not necessary (especially due to the lack of an incentive problem for producing and analyzing data). On the contrary, its introduction can be even dangerous for innovation and competition in the digital economy, because it might lead to considerable legal uncertainty, the monopolisation of information, and impediments for the free flow of data that is so crucial for the digital economy")

[54] C. Ducuing, T, Margoni, L. Schirru, D. Spajic, T. Lalova-Spinks, L. Stähler, E. Bayamlıoğlu, A. Pétel, J. Chu, B. Peeters, A. Christofi, J. Baloup, M. Avramidou, A. Benmayor, T. Gils, E. Kun, E. De Noyette, and E. Biasin, White Paper on the Data Act Proposal (October 26, 2022), <https://ssrn.com/abstract=4259428> accessed 20 Dec., 2022,

licenses for anyone to use. Let's see how this is done in the Creative Commons project.

A. Creative Commons

Wikipedia, Flickr, Youtube, Google Images and more. What all these portals have in common? All of them work, in a way or another, with Creative Commons Licenses.[55] In the image search engine of Google Images and Flickr you can limit your results only to content that may, for example, be used for both commercial and non-commercial purposes, under the only condition to give the correct attribution to its author. The same applies to searching videos on YouTube. In Wikipedia, if you scroll down to the end of the page, you will see that Wikipedia´s entries are licensed under a CC BY-SA 3.0 license.

In defiance of the "all rights reserved" paradigm and proposing a "some rights reserved" perspective,[56] Creative Commons has as one of its main goals "to increase the amount of openly licensed creativity in "the commons" — the body of work freely available for legal use, sharing, repurposing, and remixing."[57]

The possibility of limiting the results of a search to works licensed under a CC license as mentioned above, thus favoring **F**indability and **A**ccessibility, is due to the fact that these licenses are built in three layers: a human-readable deed, with a brief summary of the main duties of the user under that license; a lawyer-readable legal code, with all the legal information, scope and limits of the license granted and; a machine-readable metadata.[58] The latter operates through the Creative Commons Rights Expression Language (CC REL) metadata, which "may be embedded in a variety of filetypes".[59]

---

[55] On the relationship between the author, the users and the role of CC licenses, see S Dusollier, 'The Master's Tools v. The Master's House: Creative Commons v. Copyright' (2006) 29 Columbia Journal of Law & Arts, 271 < https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2186187> accessed 21 Dec 2022.

[56] See MW Carroll, Creative Commons as Conversational Copyright (2007) in PK Yu (ed) *Intellectual Property and Information Wealth: Issues and Practices in the Digital Age* (Praeger, 2007), Villanova Law/Public Policy Research Paper No. 2007-8 <https://ssrn.com/abstract=978813> accessed 21 Dec 2022.

[57] 'Use & Remix' (*Creative Commons*) <https://creativecommons.org/use-remix/> accessed 21 November 2022. On the role of CC licenses in promoting Open Access, see T Margoni, D Peters, 'Creative Commons Licenses: Empowering Open Access' (2016) <https://ssrn.com/abstract=2746044> accessed 21 Dec 2022.

[58] 'What does it mean that Creative Commons Licenses are "machine-readable"?'(*Creative Commons*) <https://creativecommons.org/faq/#what-does-it-mean-that-creative-commons-licenses-are-machine-readable> accessed 21 November 2022.

[59] 'CC REL' (*Creative Commons Wiki*, 15 September 2021) <https://wiki.creativecommons.org/wiki/CC_REL>.

Through the option of using a License Chooser mechanism,[60] the licensor may opt for different combinations of uses that can be made of a work under the selected license. The most permissive license is the CC-BY, that "allows [licensees] to distribute, remix, adapt, and build upon the material in any medium or format, so long as attribution is given to the creator"[61] both for commercial and non-commercial uses. All the other licenses may present a combination of the attribution (BY) requirement with other additional requirements, like: NC – Noncommercial use;[62] SA – Share alike;[63] and ND – No derivatives allowed.[64] An additional tool offered by CC is the CC0 waiver, "which allows creators to give up their copyright and put their works into the worldwide public domain. CC0 allows re-users to distribute, remix, adapt, and build upon the material in any medium or format, with no conditions."[65]. Finally, there is the Public Domain Mark, which is simply a tool to identify works already in the public domain.[66]

It is important to notice that Creative Commons licenses have smart mechanisms built in the license to adjust to national copyright law implementations (remember territoriality?). Taking the CC0 waiver as an example, its legal code explicitly provides that the waiver of all copyrights shall be limited to "the greatest extent permitted by, but not in contravention of, applicable law (…)".[67] In other words, in jurisdictions where waivers of copyright are not contemplated, CC0 will operate as a license with an as close as possible legal effect. Similarly, in countries where moral rights exists (virtually every country in the world with the partial exception of the US) and cannot be waived (moral rights are not transferrable, but in certain jurisdictions they can be abandoned), CC0 waiver will, for example, be limited to the economic rights. [68] Creative Common licenses have evolved over time and the current version 4.0 includes not only copyright as such but also neighboring rights and, importantly for the present analysis, the SGDR. Additionally, Attribution (BY), which was initially (version 1.0) an option,

---

[60] 'License Chooser' (*Creative Commons*) <https://creativecommons.org/choose/> accessed 21 November 2022.

[61] 'About CC Licenses' (*Creative Commons*, 2019) <https://creativecommons.org/about/cclicenses/>.

[62] 'About CC Licenses' (*Creative Commons*, 2019) ("Only noncommercial uses of the work are permitted"), <https://creativecommons.org/about/cclicenses/>.

[63] 'About CC Licenses' (*Creative Commons*, 2019) ("Adaptations must be shared under the same terms"), <https://creativecommons.org/about/cclicenses/>.

[64] 'About CC Licenses' (*Creative Commons*, 2019) ("No derivatives or adaptations of the work are permitted") <https://creativecommons.org/about/cclicenses/>.

[65] 'About CC Licenses' (*Creative Commons*, 2019) <https://creativecommons.org/about/cclicenses/>.

[66] 'Public Domain Mark 1.0.' (*Creative Commons*) < https://creativecommons.org/publicdomain/mark/1.0/> accessed 20 December 2022.

[67] 'CC0 1.0. Universal' (*Creative Commons*) <https://creativecommons.org/publicdomain/zero/1.0/legalcode> accessed 21 November 2022.

[68] See, eg, Brazil, Law n. 9.610 of February 19, 1998 (Law on Copyright and Neighboring Rights), art. 27 ("Moral rights are inalienable and irrevocable.").

has since become a mandatory element including for US based licensors. [69] Finally, CC licenses are intended for all copyright creation except for software, since for the latter, FLOSS licenses have been in use (and in fact have inspired CC) for many years.[70]

In the following we replicate the original table on FAIR requirements and descriptions, with the addition of how CC licenses elements reflect FAIR requirements.

Principle	**Definition** (https://www.go-fair.org/fair-principles/)	**Description** (*Turning FAIR into reality: Final Report and Action Plan of FAIR Data, European Commission, 2018, 19-20 (footnotes removed)*	**CCPL 4.0**
Findability	"F1. (Meta)data are assigned a globally unique and persistent identifier F2. Data are described with rich metadata (defined by R1 below) F3. Metadata clearly and explicitly include the identifier of the data they describe F4. (Meta)data are registered or indexed in a searchable resource"	"Data are **Findable** when they are described by sufficiently rich metadata and registered or indexed in a searchable resource that is known and accessible to potential users. Additionally, a unique and persistent identifier should be assigned such that the data can be unequivocally referenced and cited in research communications."	CCPL is expressed in legal, human and machine (metadata) language. Works properly licensed are therefore easily findable, although CC does not in itself provide a default repository. However, Openverse (https://wordpress.org/openverse) allows to search for 600M works licensed under CC and other open licenses.
Accessibility	"A1. (Meta)data are retrievable by their identifier using a standardised communications protocol	**"Accessible** data objects can be obtained by humans and machines upon appropriate authorisation and through a well-defined and universally implementable protocol. In other words, anyone with a computer and an Internet connection should be able to	For reasons similar to the above, CCPL works are very often (although not necessarily) accessible via the Internet and the conditions for access

[69] See, eg, 'CC Attribution 4.0. International', Sec 2(b)(1)(2) (*Creative Commons)* ("Other rights. 1. Moral rights, such as the right of integrity, are not licensed under this Public License, nor are publicity, privacy, and/or other similar personality rights; however, to the extent possible, the Licensor waives and/or agrees not to assert any such rights held by the Licensor to the limited extent necessary to allow You to exercise the Licensed Rights, but not otherwise. 2. Patent and trademark rights are not licensed under this Public License.") < https://creativecommons.org/licenses/by/4.0/legalcode> accessed 21 November 2022.

[70] On FLOSS, see, eg., T Margoni, M Perry, 'Free-Libre Open Source Software as a Public Policy Choice' (2010) 3 (3,4) International Journal on Advances in Internet Technology, 212.

	A1.1 The protocol is open, free, and universally implementable A1.2 The protocol allows for an authentication and authorisation procedure, where necessary A2. Metadata are accessible, even when the data are no longer available"	access at least the metadata. It is important to emphasise that Accessible in FAIR does not mean Open without constraint. Accessibility means that the human or machine is provided - through metadata - with the precise conditions by which the data are accessible and that the mechanisms and technical protocols for data access are implemented such that the data and/or metadata can be accessed and used at scale, by machines, across the web."	and reuse are clearly established in the metadata, in the human readable copyright notice as well as in the full license.
Interoperability	"I1. (Meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation. I2. (Meta)data use vocabularies that follow FAIR principles I3. (Meta)data include qualified references to other (meta)data"	"**Interoperable** data and metadata are described in the FAIR principles as those that use a formal, accessible, shared, and broadly applicable language for knowledge representation. They use vocabularies which themselves follow the FAIR principles, and they include qualified references to other data or metadata. What this describes is semantic interoperability. […]It is not only semantics but also technical and legal interoperability. Technical interoperability means that the data and related information is encoded using a standard that can be read on all applicable systems. In FAIR, legal interoperability falls under the principle that data should be 'Reusable'."	Interoperability for license purposes may be conceptualized as the ability of two different licenses to operate together and allow the underlying works to be lawfully combined, reused and reshared. Open licenses like CCPL have normally high degrees of interoperability, however, when two licenses are strong copyleft this may lead to incompatibility and thus lack of interoperability (the usual example is GPLv2 and GPLv3)[71]. which being both strong copyleft do not allow the redistribution of derived works). CCPL employs specific language (or "equivalent license") that favours interoperability.
Reusability	"R1. (Meta)data are richly described with a plurality of accurate and relevant attributes R1.1. (Meta)data are released with a clear and	"For data to be **Reusable**, the FAIR principles reassert the need for rich metadata and documentation that meet relevant community standards and provide information about provenance. This covers reporting how data was created (e.g. survey protocols, experimental processes,	Reusability for legal purposes seems to refer to a step that precedes interoperability. Reusability refers to the possibility to reuse the licensed material (thanks to a license

[71] 'GPL-Compatible Free Software Licenses' (*GNU Operating System*, 12 Jan 2022) < https://www.gnu.org/licenses/license-list.html#GPLCompatibleLicenses> accessed 21 Dec. 2022.

	accessible data usage license R1.2. (Meta)data are associated with detailed provenance R1.3. (Meta)data meet domain-relevant community standards"	information about sensor calibration and location) and information about data reduction or transformation processes to make data more usable, understandable or 'science-ready'. As shown in the example of the DOBES case study (Fig. 3), open community-endorsed formats also play a key role in reusability. The ability of humans and machines to assess and select data on the basis of criteria relating to provenance information is essential to data reuse, especially at scale. Reusability also requires that the data be released with a 'clear and accessible data usage license': in other words, the conditions under which the data can be used should be transparent to both humans and machines."	that allows reuse, i.e., reproductions, adaptations and redistributions or public communications). CCPL always allows these acts, although certain conditions may limit them (NC or ND). The BY condition ensures that provenance of the work, authorship and chain of alterations is preserved.

B. Other data access regimes

The issue of data reusability may not be limited to an analysis focused on property rights. As we saw, property rights in general, and copyright in particular, are not well designed to protect data. In fact, there are clear industrial and innovation policy considerations that for a long time have excluded – or strictly limited – the proprietary protection of data. Only under the strict conditions sketched above, data may find housing within the copyright building. That said, the apparent contradiction of a category (data) that is gaining unprecedented economic and social value, but which cannot be properly governed by the legal tools (copyright and contracts) that we often – wrongly – rely on, may have contributed to direct the EU legislator towards a different route. We will not discuss these alternative access regimes in detail, as this would exceed the scope – and space – of this chapter. Nevertheless, it may be offer valuable insight to indicate them, even briefly.

Research data. The Open Data Directive (ODD) defines research data as "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"[72].

[72] Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June

The Directive requires EU member states to ensure, via national Open Access policies, that publicly funded research data be openly available (via 'open access policies'), following the principle of 'open by default' and compatible with the FAIR principles (Art. 10 ODD). Art. 10 further clarifies that research data must be re-usable both for commercial as well as for or non-commercial purposes and researchers, research performing organisations or research funding organisations have already made them publicly available through an institutional or subject-based repository.

High-value datasets. The same ODD defines high-value datasets as documents the re-use of which is associated with important benefits for society, the environment and the economy, in particular because of their suitability for the creation of value-added services, applications and new, high-quality and decent jobs, and of the number of potential beneficiaries of the value-added services and applications based on those datasets.[73] High-value datasets have to be made available for re-use in machine-readable format, via suitable APIs and, where relevant, as a bulk download and free of charge for the final user. Chapter V of the ODD specifies a set of rules for this important category of data.

IoT Data. The proposed Data Act defines Internet of Things (IoT) data as 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 audio-visual recording produced by a tangible, movable item, including where incorporated in an immovable item, that obtains, generates or collects, data concerning its use or environment, and that is able to communicate data via a publicly available electronic communications service and whose primary function is not the storing and processing of data. Users of these products have the right to access free of charge in real time and to share these data with third parties subject to certain conditions.[74]

CONCLUSION

Research data, High-value datasets and IoT data are only some of the examples of a new approach in EU law to the regulation of data access and

2019 on open data and the re-use of public sector information (recast) [2019] OJ L 172/56, art 2(9).

[73] Id, art. 1(10).

[74] Commission, '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 ('Data Act proposal'), arts. 3-5.

reuse based on a different paradigm than the proprietary one seen above. This new paradigm, as discussed elsewhere[75], suggest a novel private-public governance model of data predicated on the ambitious project of European common data spaces. To the extent that the future of data governance is far from the traditional proprietary field (e.g., copyright) it will be interesting to re-explore how new contractual or technological forms will be employed to regulate access and reuse. It is clear that the current ones, based on copyright rules that often do not apply to data as such, will be hardly available to this new world of data.

[75] Ducuing, Margoni, Schirru et al, White Paper on the Data Act Proposal (October 26, 2022).