Article

Big Data and the Law: a holistic analysis based on a three-step approach - Mapping property-like rights, their exceptions and licensing practices

BENHAMOU, Yaniv

Abstract

Given Big Data's increasing importance, it seems im-portant to develop a clear and coherent body of law and to organize the different relationships among all stakeholders and legal regimes that grant different property-like rights. There are already numerous stud-ies on the interactions between Big Data and intellec-tual property and/or privacy, but less that map all legal regimes, including the question of licensing and overlaps between legal regimes. In a first part specifi-cally addressed to data producers or data users active in Big Data (collectively referred to as "organizations"), this article suggests a holistic approach, where we map all legal regimes from a comparative law perspective and suggest a three-step approach, according to which each organization needs to assess whether a given data is subject to a property right; if yes, whether that rele-vant data may be used thanks to exceptions or other flexibilities notwithstanding the property right and, if not, how to license them (below II). In a second part specifically addressed to policymakers, this article ex-plores selected policy considerations, [...]

Reference

BENHAMOU, Yaniv. Big Data and the Law: a holistic analysis based on a three-step approach - Mapping property-like rights, their exceptions and licensing practices. *Revue suisse de droit des affaires et du marché financier*, 2020, no. 4, p. 393-418

Big Data and the Law: a holistic analysis based on a three-step approach – Mapping property-like rights, their exceptions and licensing practices

Yaniv Benhamou*

Given Big Data's increasing importance, it seems important to develop a clear and coherent body of law and to organize the different relationships among all stakeholders and legal regimes that grant different property-like rights. There are already numerous studies on the interactions between Big Data and intellectual property and/or privacy, but less that map all legal regimes, including the question of licensing and overlaps between legal regimes. In a first part specifically addressed to data producers or data users active in Big Data (collectively referred to as "organizations"), this article suggests a holistic approach, where we map all legal regimes from a comparative law perspective and suggest a three-step approach, according to which each organization needs to assess whether a given data is subject to a property right; if yes, whether that relevant data may be used thanks to exceptions or other flexibilities notwithstanding the property right and, if not, how to license them (below II). In a second part specifically addressed to policymakers, this article explores selected policy considerations, including the question of overlaps between legal regimes, the creation of alternative dispute resolution mechanisms to ensure that all data processes function seamlessly, and the promotion of co-regulation in developing digital standards at an age of de-regulation to increase democratic control and the participation of all stakeholders (below III).

Table of contents

I. Introduction

Big Data is subject to several definitions[1] but may be defined with reference to the three main operations: preexisting data (so-called "**inputs**") feed a Big Data engine, system, tool or device that analyzes the input (so-called **"analytics"**) to come up with an outcome (so-called **"output"**).[2] Big Data is also a way to describe business models based on a large volume of data (the more data fed into an algorithm, the more

* Lecturer University of Geneva (IP & Privacy), PhD, Of Counsel Attorney. The author sincerely thanks Prof. *Alain Strowel* and *Florent Thouvenin, Ms. Maryam Kanna, Ana Andrijevic, Hélène Bruderer, Sotiria Kechagia* and *Emmy Gijs* for their helpful comments and in particular Ms. *Justine Ferland,* who contributed to a previous version on Big Data & Intellectual Property.

1 Big Data is often used in the literature with capital letters and is traditionally defined with reference to the 4 V's: a large Volume of data produced coming from a high Variety of sources at a high Velocity and Veracity (*Tal Zarsky,* Incompatible: The GDPR in the Age of Big Data, Seton Hall Law Review (2017), Vol. 47 Iss. 4, 998–999 and the cited references). For further references and definitions, see *Philippe Meier,* Le défi de Big Data dans les relations entre privés: avec quelques réflexions de lege ferenda, in: Epiney/Nüesch (ed.), Big Data und Datenschutzrecht = Big data et droit de la protection des données, Forum Europarecht 37/2016, Zurich 2016, 50.

2 *Richard Kemp,* Legal Aspects of Managing Big Data, Computer Law & Security Review: The International Journal of Technology Law and Practice (2014), Vol. 30 Iss. 5, 482–491, 482 and 486. In this ecosystem, data may also be defined with reference to a pyramid, in the sense that "data precedes information, which precedes knowledge, which precedes understanding" (see *Alain Strowel,* Big Data and Data Appropriation in the EU, in: Aplin (ed.), Research Handbook on Intellectual Property and Digital Technologies, Cheltenham/Northampton 2018, 107–135, 107).

relevant the results will be), so that the "data" remain the main focus of the discussion.

Data may be defined by reference to their different and sometimes overlapping **legal regimes** (e.g. copyright for copyrighted data, privacy for personal data, trade secret for confidential data). These legal regimes may have absolute effects *(erga omnes),* such as is the case with the protection granted by copyright to copyrighted data (e.g. videos or photos posted on Facebook), privacy protection for personal data (e.g. name, voice, IP addresses) or other intellectual property similar rights for confidential information (e.g. customer lists or other business information), or relative effects, such as the one granted by contractual agreements for instance to raw data (e.g. technical and training data used by AI systems, research data or machine-generated data collected by sensors).

There are already numerous studies on the interactions between Big Data and intellectual property and/or privacy, but less that map all legal regimes, including the question of licensing and overlaps between legal regimes.[3] This article suggests a **holistic approach,** that maps all legal regimes from a Swiss and comparative law perspective. However, this article does not purport to analyze each issue extensively but to select few relevant issues. Also, it does not focus on a specific jurisdiction and applicable law, as legal regimes vary from one jurisdiction to another. However, particular attention will be given to EU law given its significance for Swiss based organizations, and in particular to the EU General Data Protection Regulation (GDPR) due to its extra-territorial scope and its global influence,[4] in addition to the Swiss data protection legislation (both the current act "LPD" and the revised act "nLPD").[5]

In particular, the article suggests to guide data producers or data users (collectively referred to as **"organizations"**) based on a **three-step approach,** according to which each organization active in Big Data needs to assess whether a given data is subject to a property-like right; if yes, whether that relevant data may be used thanks to exceptions or other data access flexibilities and, if not, how to license them (below II). Following this three-step approach, the article explores selected **policy considerations,** including the question of overlaps between legal regimes, the promotion of alternative dispute resolution mechanisms and of co-regulation in developing digital standards (below III).

II. Three-Step Approach

The following section aims to guide organizations with a three-step approach, according to which it is advisable (1) to assess whether a given data is subject to property-like rights and (2) to exceptions or other flexibilities, (3) before considering subjecting them to contractual terms.

1. Property-like rights

Data may be subject to several legal regimes. Legal regimes can be classified into a binary approach opposing personal and non-personal data, or opposing horizontal and vertical legal instruments.[6] We chose

[3] For an analysis on the interfaces between Big Data and Intellectual Property Rights (IPR) from a European and comparative law perspective, see *Daniel Gervais,* Exploring the Interfaces Between Big Data and Intellectual Property Law, JIPITEC 10 (1) (2019); from a Swiss law perspective, see *Yaniv Benhamou/Laurent Tran,* Circulation des biens numériques: de la commercialisation à la portabilité, sic! (2016), 571–591, 579. For an analysis on the interfaces between Big Data and data protection from a Swiss law perspective, see *Meier* (note 1), 47 ff.

[4] For the geographical extra-territorial scope of application and its enforcement abroad, see *Yaniv Benhamou/Emilie Jacot-Guillarmod,* GDPR on the Swiss territory, Cooperation with European Authorities and Enforcement of Monetary Fines (24 May 2018), Jusletter IT, 1; European Data Protection Board, Guideline 3/2018 on the territorial scope of the GDPR (Article 3), adopted on 16 November 2018. Regardless of the geographical scope, other lawmakers adopted or consider adopting similar privacy laws (e.g. Brazil, India, China, California). See *Ahmed Baladi,* Can GDPR Hinder AI Made in Europe? (10 July 2019), Cybersecurity law Report, 1.

[5] The Federal Act on Data Protection of 19 June 1992 (LPD) has been revised into a new act (nLPD), adopted by the Parliament on 25 September 2020. See *<https://www.parlament.ch/fr/ratsbetrieb/suche-curia-vista/geschaeft?AffairId=20170059>* [accessed 2 October 2020].

[6] This binary approach is the one typically adopted by EU data protection law (see Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union, which defines non-personal data as data "other than personal data" (article 3(1) Regulation (EU) 2018/1807) and has been increasingly discussed. See also *Laura Somaini,* Regulating the Dynamic Concept of Non-Personal Data in the EU: From Ownership to Portability, European Data Protection Law Review

another approach of relying on property-like rights.[7] If the scope of each property-like protection seems clear *prima facie,* the analysis shows that the eligibility for protection may be delicate in many respects. There is thus a need to clarify the scope of each one of those legal regimes.

1.1 Privacy and data protection

Data may be subject to **privacy protection** when an individual's identifying characteristics are concerned (e.g. image or voice) and to **personal data protection** when there is a processing of personal data[8] (collectively referred to as "privacy laws").[9] Many Big Data activities shall be subject to privacy laws, as most data fall under the broad definition of *"processing"* and of *"personal data".*[10] Indeed, most data might qualify as personal data, whether they are used as input (e.g. location data that feed an algorithm that analyzes the traffic flows), analytics (e.g. processing of the input in order to better understand the traffic flow) or output (e.g. understandings about the traffic flows).[11] However, purely raw data (such as failure indicators on an aircraft engine; sensors placed in a sine; weather data) or personal data which were subsequently anonymized are considered non-personal data.[12]

As a result, some consider privacy laws incompatible with Big Data, while others conversely view it as a tool to improve data quality. This contribution does not discuss this issue extensively, but the following considerations can be raised in the Big Data context. Some obligations might be in tension with Big Data activities, in particular those of **"purpose limitation"** and **"data minimisation"**, particularly when organizations collect data with no clear limit or purpose (e.g. Data Broker or Data Warehousing), the principle of **"lawfulness"** when organizations do not rely on a valid consent or other lawful grounds or

(2020), Vol. 6 Iss. 1, 84–93, 89. For a Swiss law perspective, see *Jacques de Werra,* Entreprises et Big Data: peut-on forcer les entreprises à partager leurs données non personnelles (par des licenses obligatoires ou des licences "FRAND")?, in the same issue, section I.1.

7 Property-like right refers here to both real exclusive and full enjoyment-power *("jouissance-maîtrise"),* such as for corporeal things, absolute subjective right (erga omnes), such as for intellectual property rights, as well as other control rights or claims, such as for privacy protection (e.g. access, portability, rectification, erasure, objection to processing). See *Julie Cohen,* Property as Institutions for Resources: Lessons from and for IP, Texas Law Review (2016), Vol. 94 Iss. 1, 1–57, 5 ff.; *Strowel* (note 2), 24, indicates that there is no one property but many different property-like regimes, and that the rights conferred by privacy laws come closer to property due to their similarities with copyright (in particular due to their tradable features and their similarities with non-economic moral copyright).

8 "Personal data" means *"all information relating to an identified or identifiable person"* (see articles 3 let. a LPD, 4 let. a P-LPD and 4 (1) GDPR).

9 Privacy protection is the freedom against any unjustified interference by the State or private parties (see article 7 EU Charter on fundamental rights, which provides that everyone has a *"right to respect for his or her private and family life"*), while personal data protection provides control rights of individuals and obligations of organizations processing personal data (see article 8 (1) EU Charter, granting *"the right to the protection of personal data"*). See *Strowel* (note 2), 18 and the cited references. Privacy protection has long been recognized as a human right (see article 12 of the Universal Declaration of Human Rights and article 17 of the International Covenant on Civil and Political Rights), while personal data protection is a relatively recent human right that is closely connected to the right to privacy. See *Christophe Kuner/Massimo Marelli,* Handbook on data protection in humanitarian action, Geneva, International Committee of the Red Cross (ICRC), May 2020, 24.

10 In Swiss law, see *Meier* (note 1), 55 ff. In EU law, see *Michèle Finck/Frank Pallas,* They who must not be identified – distinguishing personal from non-personal data under the GDPR, International Data Privacy Law (2020), Vol. 10 Iss. 1, 11–36; *Nadezhda Purtova,* The law of everything. Broad concept of personal data and future of EU data protection law, Law, Innovation and Technology (2018), Vol. 10 Iss. 1, 40–81, 78.

11 Personal data processed by companies are traditionally classified in accordance with the origins of: own-generated data (first party data), licensed data by a data provider (second party data) or public available data (third party data), and with the four following types of data: demographic (e.g. name, gender, age, address), geographic (e.g. IP address, current location), behavioural (e.g. sites browsed, interactions with other websites) and psychographic data (e.g. preferences, values, lifestyle). Personal data may also be shared as a product (as an end in itself) or as a by-product of another transaction (i.e. used as a means to furnish other services). See *Julien Debussche/Jasmien César,* Data-related legal, ethical and social issues, Bird & Bird, August 2019, 10.

12 *Meier* (note 1), 55; Communication from the Commission to the European Parliament and the Council, Guidance on the Regulation on a framework for the free flow of non-personal data in the European Union, COM(2019) 250 final, Brussels, 29.5.2019, 9.

process data against the will of the user (e.g. when on a social network, upon request of the user, the profile is not deleted but simply deactivated).[13] The individual's **control rights** might also be difficult to implement in certain Big Data business models, such as the rights of access or erasure, and information duties.[14] Moreover, due to the **dynamic nature** of data in a Big Data context, even raw or anonymized data[15] transform into personal data when aggregated in a dataset that allows the identification of individuals (i.e. de-anonymization of raw data via aggregation).[16] Consequently, the company that analyzes anonymized data will still have to comply to data protection standards not only when the input consists of personal data but also when the output allows to correlate the data with a determined person. Similarly, "ordinary data" may become "sensitive", which is subject to increased data protection, for instance when the data analysis reveals certain characteristics about the data subject.[17] Despite these difficulties, there are solutions to comply with privacy laws, in particular when organizations rely on anonymized data, other legitimate grounds and other flexibilities (see below II.2).

1.2 Intellectual Property Rights (IPR)

1.2.1 Copyright protection

Copyright may protect **individual data** (e.g. images, texts, films, music) or softwares (source code, object code, associated documentation) with sufficient originality but not simple facts or information.[18] Copyright may also protect **compilation of data** (e.g. a database) if their selection or arrangement is original.[19]

13 For an analysis under the GDPR, see *Zarsky* (note 1), 995 ff.; *Reto Hilty,* Big Data: Ownership and Use in the Digital Age, in: Seuba/Geiger/Pénin (ed.), Intellectual Property and Digital Trade in the Age of Artificial Intelligence and Big Data, International Centre for Trade and Sustainable Development (ICTSD)/Centre d'Études Internationales de la Propriété Intellectuelle (CEIPI), Geneva/Strasbourg 2018, 85–94, 85. For an analysis under Swiss law, see *Meier* (note 1), 60, referring to the principles of proportionality, speciality, recognizability, as well as the requirements of information and of consent.

14 See *Meier* (note 1), 65, giving the example of the Disneyland magicband described on the company's website as an *"all-in-one device that enables the visitor to enter the parks and buy food and merchandise",* and questioning how to comply with the information duty when the device is provided in exchange of the geolocation data, purchase records, attractions visited in order to improve the traffic flow in the parc on one hand, and improve targeted advertisements on the other hand.

15 Anonymized data are not covered by data protection. See Message du 15 septembre 2017 concernant la loi fédérale sur la révision totale de la loi fédérale sur la protection des données et sur la modification d'autres lois fédérales, FF 2017 6565, 6640. See also GDPR Recital 26: *"The principles of data protection should therefore not apply to anonymous information, namely information that does not relate to an identified or identifiable natural person or to personal data rendered anonymous in such a manner that the data subject is not or no longer identifiable."*

16 In Swiss law, see *Meier* (note 1), 59. In EU law, see *Strowel* (note 2), 19.

17 *Debussche/César* (note 11), 11, considering that the use of sensitive data is restricted and prohibited in most cases. See *Strowel* (note 2), 19, recalling that Big Data aims to assemble as much data as possible, which can lead to the creation of a person's partial or complete profile, even contributing to indicate a person's sexual orientation (sensitive data) for instance, leading to the application of special norms.

18 Article 2 (1) Bern Convention for the Protection of Literary and Artistic Works *("literary and artistic works shall include every production in the literary, scientific and artistic domain, whatever may be the mode or form of its expression")*. In Swiss law, see article 2 (1) Swiss Copyright Act *("Works are literary and artistic intellectual creations with an individual character, irrespective of their value or purpose"); Vincent Salvadé/Nathalie Tissot,* La réalisation d'un site web ou l'ouverture d'un compte par le travailleur. Qui est titulaire des droits?, in: Dunand/Mahon (ed.), Internet au travail, Geneva/Zurich/Basle 2014, 227–253, 229; *Jacques de Werra/Yaniv Benhamou,* Kunst und geistiges Eigentum, in: Mösimann/Renold/Raschèr (ed.), Kultur, Kunst Recht: Schweizerisches und internationales Recht, Basle 2020, 707 ff. In EU law, see *Gervais* (note 3), N 8 ff.; *Debussche/César* (note 11), 11.

19 See article 10 (2) of the World Trade Organization's Agreement on Trade-Related Aspects of Intellectual Property Rights (TRIPS Agreement) *("Compilations of data or other material, whether in machine readable or other form, which by reason of the selection or arrangement of their contents constitute intellectual creations")*. In Switzerland, such databases can be protected as collected works defined at article 4 (1) Swiss Copyright Act *("creations with individual character with regard to their selection and arrangement")*. In the EU, article 3 (1) EU Database Directive *("Databases which, by reason of the selection or arrangement of their contents, constitute the author's own intellectual creation shall be protected as such by copyright")*. In the US, Title 17 USC s. 101 Copyright Act protecting compilation (defined as *"work formed by the collection and assembling of pre-existing materials or of data that are selected, coordinated, or arranged in such a way that the resulting work as a whole constitutes an original work of authorship"*).

With respect to the three Big Data operations (inputs, analytics, and outputs), as examples of **inputs** and **outputs,** one can think of data enriched with original information (such as data visualisations, reports, annotations, figures, charts, graphics) or structured databases (such as compilations with an original selection or arrangement of content, even though the latter is not copyrightable). As an example of **analytics (algorithms),** one can think of the underlying source or object code, recalling that copyright only applies to the original form of expression, not the idea or simple information embedded in a creative work, so that the underlying software is more likely than the individual data or database itself to receive copyright protection.[20]

Eligibility for copyright protection may be however questionable in a Big Data context. From the outset, level of investments made to create the database is insufficient to demonstrate the originality.[21] Moreover, copyright protection usually requires an intellectual human intervention and the consciousness of achieving a result. Therefore, raw data, AI-generated data by unsupervised machine-learning techniques (e.g. weather forecasts, stock quotations or sports scores) or databases automatically created by an algorithm are as matter of principle excluded from copyright protection.[22] Similarly, the simple collection of data unaltered for a new database may lack copyright protection (e.g. listing the temperatures acquired from a weather station for a specific period, top-selling songs for a certain month, customer's data).[23]

As an outcome, data collected and processed in a Big Data context will often not be copyrightable for lack of originality. Companies that have made investments to create new databases or algorithms may wish to rely on other property-like rights.

1.2.2 Database protection

Databases may be protected in many countries through competition law,[24] while they may be protected in the EU by a *sui generis* database right.[25]

In practice, this means that the database producer could prohibit the downloading of data or datasets contained therein, even if the underlying data are not copyrighted.[26] As examples of **inputs** that may be protected by competition law or the *sui generis* database right, one can think of a customer list collected by a company. As examples of **outputs,** one can think of databases generated by a company to obtain, verify and/or present the data collected.

[20] See for instance article 9 (2) TRIPS Agreement *("Copyright protection shall extend to expressions and not to ideas, procedures, methods of operation or mathematical concepts as such")*. See however *Teresa Scassa,* Data Ownership, CIGI Papers No 187, September 2018, 9, who considers that the idea/expression dichotomy (according to which copyright protection extends to the expression of ideas only) may be blurred. For instance, where the expression of a fact or an idea merges with that fact or idea (for example, where there is only one or a very limited number of ways to express it), there can be no copyright protection since the practical result of any such protection would be to give a monopoly over the fact or idea.

[21] In Switzerland, copyright protection has been denied, despite investments made, for a compendium of drugs, a telephone directory and logarithmic tables (case-law quoted by *Philippe Gilliéron,* in: de Werra/Gilliéron (ed.), Commentaire Romand de la Propriété intellectuelle, Basle 2013, N 6 ad art. 4). In the EU, see article 3 (1) of the EU Database Directive. In the US, see *Feist Publ'ns, Inc. v. Rural Tel. Serv. Co.,* 499 U.S. 340 (1991), 344.

[22] It is however important to distinguish between works created with the assistance of a computer (i.e. regular works just like books created with the assistance of a pen, or movies created with the assistance of a camera) and Computer-generated works (CGW) (i.e. works generated by computer in circumstances such that there is no human author of the work). Most AI-generated work will likely not meet the threshold to be qualified as a CGW. See below II.2.2.2.

[23] *Robert Maier/Joshua Sibble,* Big Data Handbook: A Guide for Lawyers, Wolters Kluwer Legal & Regulatory, May 2018, 23.

[24] In Switzerland, databases may be protected in certain circumstances by the Swiss Act against Unfair Competition (UCA), in particular article 5 let. c UCA prohibiting the reuse of third-party work by technical processes without corresponding investments, see *Jacques de Werra/Yaniv Benhamou,* Propriété intellectuelle et concurrence déloyale. Analyse du droit suisse et perspectives de droit allemand, in: Puttemans/Gendreau/de Werra (ed.), Propriété intellectuelle et concurrence déloyale: les liaisons dangereuses?, Brussels 2017, 183–208, 185. See below II.2.2.2.

[25] Directive 96/9/EC of the European Parliament and of the Council of 11 March 1996 on the legal protection of Databases (Database Directive), article 7 (1)–(2). The EU *sui generis* database right was developed in order to protect data producers' investments and to prevent free-riding on somebody else's investment in creating the database; see *Strowel* (note 2), 15.

[26] In Swiss law, see *de Werra/Benhamou* (note 18), N 119 ff. In EU Law, see *Strowel* (note 2), 15.

Like copyright protection, the eligibility for database protection is delicate in a Big Data context. In Switzerland, case-law tends to limit this protection, in particular when the third party user *(repreneur)* has made substantial investment or when the data producer covered his/her investments made. Consequently, in case of reuse of third party data into a derivative data or a database, the protection against unfair competition is usually refused because of the investment made by the third party user (*repreneur*) to generate his own database.[27] In the EU, following the Court of Justice of the European Union's (CJEU), the *sui generis* database right protection requires that substantial investments have been made and extends to the collection only, not to the creation of data.[28] This may exclude many Big Data and IoT businesses, as it could exclude machine-generated databases, for which it could be argued that the underlying data are "created" instead of "obtained".[29] Moreover, even if the scope of protection is quite broad,[30] the *sui generis* database right does not protect each individual data but only substantive parts of the database, so that the extraction and reuse of insubstantial part of a database remains possible.[31]

Despite these difficulties, organizations may try to claim database rights due to the amount of data and investment made in a Big Data context (e.g. IoT solutions, hardware, software, integrated apps or otherwise). For instance, in Switzerland, they may argue that no substantive investments have been made by the third party's user and that they did not cover their own investments. In the EU, they may claim such protection based on the interpretation of the conditions of protection (allowing for the protection of generated data), the subject matter (extending to individual data in case of repeated extraction), and the scope of protection (extraction and reutilization rights).[32] Nevertheless, it will become increasingly difficult to rely on this protection in the Big Data context, which does not necessarily require data to be reproduced in order to perform analytics or data mining and given the automation of the processes of obtaining, verifying and/or presenting the data.[33]

1.2.3 Patent protection

Data may be also protected by patent law. Individual data in itself is not patentable due to its factual nature but, under some circumstances, an algorithm or a software program that processes data may be protected under patent law.[34] In the Big Data context, patent protection may be particularly attractive for software programs and analytical algorithms. Unlike copyright protection, the patent will protect the tech-

27 See *de Werra/Benhamou* (note 18), N 119 ff.

28 Case C-203/02, *The British Horseracing Board Ltd e.a. v. William Hill,* [2004] ECLI:EU:C:2004:695, par. 31 ff. See *Strowel* (note 2), 15.

29 See *Strowel* (note 2), N 82 noting that "*While Art. 1 (2) of the directive defines a database as a collection of independent works, data or other materials arranged in a systematic or methodical way, the sui generis right only applies to 'the whole' or 'a substantial part' of a database (Art. 7)".* See *Debussche/César* (note 11), 201, suggesting to circumvent this difficulty by better separating the activities of data generation and of data collection, while noting that it becomes increasingly difficult to distinguish between data generation and collection in such processes.

30 Case C-202/12, *Innoweb v. Wegener,* [2013] ECLI:EU:C:2013:850. The CJEU has proposed a broad interpretation of the notion of reutilization in the context of a search engines implying that the translation of the queries from end users into the search engine for the database site "in real time" is to be considered as a reutilization.

31 There is no definite answer as to how much data exactly constitutes a "whole or substantial part" of the database, and answering this question will require a qualitative and quantitative analysis in each situation. See however, *Strowel* (note 2), 15, indicating that repeated and systematic "pumping" of individual data (which do not qualify as substantive part) could in certain conditions be prohibited under the database right (article 7 (5) Database Directive).

32 See notes 30 and 31.

33 *Debussche/César* (note 11), 56, noting that, for instance, it is unclear how techniques of enrichment, partitioning, harmonisation, homogenisation of data would fit within the criteria of obtaining, verification or presentation of the database contents.

34 In Swiss law, see article 1 (1) Swiss Patent Act *("Patents for inventions are granted for new [non-obvious] inventions applicable in industry");* for an analysis of the conditions of protection, see *Jacques de Werra,* Patents and Trade secrets in the internet age, RDS (2015), Vol. 134, 123–190, 127. In international law, WIPO, Patenting Software, available at: *<https://www.wipo.int/sme/en/documents/software_patents_fulltext.html>*. Amongst the patentability criteria that exist in most jurisdictions, five are most significant in determining patentability: (1) the invention must consist of patentable subject matter; (2) the invention must be capable of industrial application (or, in certain countries, be useful); (3) it must be new (novel); (4) it must involve an inventive step (be non-obvious); and (5) the disclosure of the invention in the patent application must meet certain formal and substantive standards.

nical application of the software, which is often what has the most commercial value.

However, the patentability of softwares and algorithms is **highly debated** at the national and international levels,[35] and a glance at the international patent landscape reveals a variety of approaches.[36] In many jurisdictions, software-related inventions either do not qualify for patent protection or have a very limited scope of protection,[37] and eligibility for patenting in those cases will be assessed on a case-by-case basis. In short, software programs and algorithms may be patented as long as the invention can be shown to have a real-world technical effect, although this result may be difficult to achieve in practice. Google's patent for MapReduce, described as a *"system and method for efficient large-scale data processing",* is an example of an innovation in data collection and analysis that has been successfully patented.[38] What is certain is that obtaining a patent in the field of Big Data analytics is a difficult process and that the wording of the claim is of the utmost importance.

Given the legal uncertainties regarding patentability of Big Data algorithms, legal disputes are likely to arise concerning the patent's validity. Moreover, companies should bear in mind that registering an algorithm under patent law implies disclosing the contents and details of said algorithm. Finally, independent from the question of patentability of algorithms remains the question of relevance of patenting Big Data algorithms. From a business perspective, the real value of analytical algorithms in the Big Data context depends on their accessibility and their ability to evolve continuously. In other words, the fixed nature of the algorithm captured in a patent may not be the most appropriate way to encourage Big Data growth and innovation. As a result, patents may be superfluous in a Big Data environment.[39] Companies may wish to keep this information secret and claim trade secret protection instead.

1.2.4 Trade secret protection

Data may be protected as a trade secret, if the data (1) is secret; (2) has commercial value because it is a secret; and (3) has been subject to reasonable steps by the rightful holder of the information to keep it secret (e.g. through confidentiality agreements and/or physical and technical restrictions on access).[40] In the Big Data context, trade secret protection appears particularly attractive for companies for two main reasons: it may protect any type of data regardless of their originality and for an unlimited period of time;[41] and there are no registration formalities, exceptions, or limitations (unlike patent or other industrial property rights).

35 See *de Werra* (note 34), 127 ff., analyzing the conditions of patent protection and showing how difficult it is to fulfil the conditions of novelty and non-obviousness in a Big Data context.

36 See the report of the United Kingdom Patent Office, Eight Great Technologies: Big Data, A patent overview, Newport 2014, which maps the players and trends on the patenting market for "big data and efficient computing" technologies. For patentability of software-enabled inventions in Europe, see European Patent Convention (EPC), notably article 52.2, and the European Patent Office (EPO)'s Guidelines for Examination. In the United States, see the two-step process established by the Supreme Court in *Alice Corporation Pty Ltd v. CLS Bank International,* 573 U.S. 208 (2014) which has rendered the process rather difficult. For examples of patents that were deemed eligible (or not) following the Alice decision, see *Maier/Sibble* (note 23), 11. In Japan, article 2 (3) (i) of the Japan Patent Act explicitly refers to computer programs as patentable subject matter and, according to the Examination Guidelines of the Japan Patent Office, a claim for a software-related invention must demonstrate that software and hardware resources work cooperatively to be patent-eligible.

37 *Ania Jedrusik/Phil Wadsworth,* Patent protection for software-implemented inventions, WIPO Magazine, February 2017.

38 It is interesting to note that any third-party big data project utilizing this framework would technically infringe on Google's patent, however Google grants general licences in that regard.

39 *Joren de Wachter,* Intellectual Property in an Age of Big Data: an Exercise in Futility?, Computer Law Review International (2014), Vol. 15 Iss. 1, 1 ff., 5.

40 In international law, see article 39 TRIPS Agreement which identifies the standards generally applicable. In Swiss law, see *de Werra* (note 34), 164. In EU law, see article 2 (1) Directive 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 (Trade Secret Directive). Also, data that have not been yet disclosed may be protected by confidential agreements or, in the absence of specific clause, by confidentiality undertakings provided by specific rules (e.g. labour law provides sometimes an obligation to keep information secret), see below II.1.3.2.

41 *Strowel* (note 2), 16; *Debussche/César* (note 11), 57.

Examples of **inputs** are data on customers and suppliers, and commercial information, such as business plans, market research and strategies. Personal data arguably could also fall under the Trade Secret Directive's scope when it is valuable because it is kept confidential, such as information about private life that celebrities keep secret to grant an exclusivity for the coverage of the event to a media company.[42] Examples of **analytics** and **outputs** as far as trade secrets are concerned include data compilations and data analysis techniques.[43] This could be the case of observed and derived data created by the data controller by virtue of the use of the service or the device by the data subject (e.g. search history, traffic data and location data that are analyzed to create a customer profile).[44] This could even be the case of trivial data, which might gain value through the new data analysis tools that find patterns and accordingly propose ads or services, and thus may qualify for trade secret protection.[45]

Like the other intellectual property rights discussed, trade secret protection has several complications in a Big Data context. First, the concept of accessibility (or non-accessibility) is affected when the information may be easily accessed by using Internet search tools and technologies.[46] Secondly, the standard of reasonableness may also be affected in the digital environment, where information are mostly stored electronically, either in-house or in the cloud with a risk of data leakage, so that the information may not be considered reasonably protected.[47] Thirdly, trade secrets are only legally protected in instances where someone has obtained the confidential information by illegitimate means (e.g. through spying, theft or bribery).[48] A trade secret holder has only a right to prohibit certain behaviors (unlawful acquisition, use or disclosure of the secret), but no exclusive rights, unlike IPR that grant exclusive rights that are legally enforceable.[49] Therefore, the trade secret holder cannot prevent competitors from copying and using the same solutions, or reverse engineering (i.e. the process of discovering the technological principles of a device, object or system through analysis of its structure, function and operation).

Consequently, trade secret protection is still uncertain in the Big Data context, so that organizations may try to rely on other protection mechanisms, in particular on contracts and technological protection measures (TPM) in order to get a kind of data exclusivity (see below II.1.3.2).

1.3 Other protection mechanisms

1.3.1 Data ownership

Because of the limitations of other legal protections, there are repeated debates around the introduction of new data ownerships, which remains however highly controversial.[50] A shift towards data access mechanisms can be observed instead.

42 *Strowel* (note 2), 23.

43 *Ibidem.*

44 Article 29 Data Protection Working Party, Guidelines on the right to data portability, adopted on 13 December 2016, last revised and adopted on 5 April 2017, WP242 rev.01, 10.

45 *Strowel* (note 2), 23, referring to the Recital 14 of the Trade Secret Directive that states that the protection applies to information that *"should have a commercial value, whether actual or* ***potential***" (emphasis added). Data out of which relevant trends are extracted by big data tools, although trivial as such, can have a potential value.

46 See article 39 par. 2 (a) TRIPS: information not *"generally known among or readily accessible to persons within the circles that normally deal with the kind of information in question"*. See *Sasqua Gr., Inc. v. Courtney and Artemis,* No. CV-10–528, 2010 WL 3613855 (E.D.N.Y. Aug. 2, 2010); *de Werra* (note 34), 176.

47 See article 39 para. 2 (c) TRIPS Agreement; *de Werra* (note 34), 176.

48 *Debussche/César* (note 11), 58.

49 Under the Trade Secret Directive, the trade secret protection is seen *"as a complement or as an alternative to intellectual property rights"* (Recital 2) which *"in the interest of innovation […] should not create any exclusive right to know-how or information"* (Recital 16). See however *Strowel* (note 2), 23, indicating that the contractual practice in certain countries relating to trade secrets shows a stronger association with property (e.g. common law countries using terms such as "assignment", "sale" or "asset transfers" for trade secrets) and that the Trade Secret Directive has the remedial aspect of a property-like protection (largely built on the IPR civil enforcement measures).

50 In the EU, one of the options discussed is the creation of a data producer's right for non-personal or anonymized data, which is heavily criticized by scholars, see Commission Staff Working Document, Evaluation of Directive 96/9/EC on the legal protection of databases, SWD(2018) 146 final, Brussels, 25.4.2018 (hereafter Commission Staff Working Document, Evaluation of Directive 96/9/EC on the legal protection of databases), 39 and the references made thereto. In the US, the option of a new data producer's right is also discussed and criticized, see for instance *Peter K. Yu,* Data Producer's Right and the Protec-

Data ownership remains **controversial.** First, the term "ownership" is already controversial. To define ownership, legal scholars usually rely on the exclusive and full enjoyment-power over a subject matter, while businesses rely instead on accountability and liability principles.[51] Secondly, many stakeholders may claim data ownership (in particular when they cannot rely on other exclusive rights) because, for instance, they generate, compile, structure, re-format, enrich, analyze or add value to the data.[52] Thirdly, there is no specific data-related legislation that explicitly recognizes ownership in data and case-law is inconclusive. For instance, certain courts consider that the deletion of data amounts to be a criminal theft, acknowledging indirectly that such data may be owned, while most courts do not consider data to be property and therefore hold that it cannot be stolen (e.g. data stored on a data carrier, such as a company-owned laptop, or the content of an email).[53] Consequently, most scholars seem to agree that there is no data ownership *de lege lata*.[54] The debate revolves around the introduction of a new ownership right *de lege ferenda,* at least at the EU level. Recent communications show that neither authorities nor businesses are in favour of a new "data ownership" type of right, on grounds that *"the crucial question in business-to-business sharing is not so much about ownership, but about how access is organized",*[55] which is largely *"left to contractual solutions".*[56]

This short overview does not allow us to go into detail, but some high-level considerations can be made. Recognizing ownership in data may seem logical for some reasons. In particular, data have the same nature as intangible assets protected by IPR (intangible and ubiquitous) and have a confirmed commercial value. Data ownership could offer better protection and enforcement mechanisms *(erga omnes),* as opposed to contractual arrangements.[57] Moreover, although property rights relate traditionally to material objects,[58] their subject matters are evolving: the legal system defines the subject matter of a property right, which is characterized by a *de facto* control

tion of Machine-Generated Data, Tulane Law Review (2019), Vol. 93, 859–929, 859, 860 ff.

51 For a recent in-depth analysis of data ownership, see *Florent Thouvenin/Rolf H. Weber/Alfred Früh,* Elemente einer Datenpolitik, Zurich 2019. See also *Rolf H. Weber/Florent Thouvenin,* Dateneigentum und Datenzugangsrechte – Bausteine der Informationsgesellschaft?, ZSR I (2018), 43–74, 43; *Benhamou/Tran* (note 3), 579 ff.

52 See *Hilty* (note 13), 83, giving the examples of traffic information apps: *"who should be the 'owner?' Should it be the car producer, the supplier of the sensor or control unit, the app producer, the service provider, the car driver – or even perhaps another party? Or would such a complex setting entail a kind of 'co-ownership?' What would 'ownership' mean, say, in the context of traffic information apps?".* See also *Debussche/César* (note 11), 72, giving the example of autonomous vehicles: *"The on-board computing systems present in connected and autonomous vehicles will allow for the transfer of substantial amounts of information, including about the driver and its location. At the current stage, it is still unclear who will 'own' this information among the many different actors involved; i.e. the driver who the personal data relates to, the owner of the vehicle (if different from the driver), the manufacturer of the vehicle, insurance companies, navigation service providers, the government, or any other third party. Any data ownership claim may have a far-reaching impact on the further implementation of the technology concerned. In any event, the personal data protection rules will need to be respected."*

53 *Debussche/César* (note 11), 73, and the references to case law.

54 Alternatively, some scholars recognize an exclusive right to control instead, or a non-exclusive right to property in data. For a non-exclusive right to property in data (eventually paired with data sharing obligations in the context of the EU-funded H2020 project TOREADOR), see *Debussche/César* (note 11), 75. For an exclusive right to control, see *Josef Drexl et al.,* Data Ownership and Access to Data: Position Statement of the Max Planck Institute for Innovation and Competition of 16 August 2016 on the Current European Debate, Max Planck Institute for Innovation and Competition Research Paper No. 16–10, 16 August 2016. Contra (i.e. in favour of the creation of an ownership right), see *Herbert Zech,* Information as Property, 6 (2015) JIPITEC 192.

55 *Debussche/César* (note 11), 75. Current debates tend to distinguish between personal data and non-personal data. With respect to non-personal data (e.g. industrial or raw data collected by sensors).

56 Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, Towards a common European data space, COM(2018) 232 final, Brussels, 25.4.2018, 9.

57 See *Yaniv Benhamou,* L'immatériel et les biens, Rapport suisse, in: L'immatériel: Journées internationales de l'Association Henri Capitant 2014, Brussels 2015, 520 and the cited references.

58 *Paul-Henri Steinauer,* Les droits réels: Tome I, 5th ed., Bern 2012, N 61; *Heinz Rey,* Die Grundlagen des Sachenrechts und das Eigentum, 3rd ed., Bern 2007, N 81.

(maîtrise de fait),[59] and data may be precisely protected by certain control rights (i.e. via the *numerus clausus* of IPR). Finally, data are non-rival and can be duplicated and transferred at zero cost worldwide. Consequently, there is a need to stimulate investment in the creation of data and databases; otherwise, such investments could decline.[60]

Nonetheless, recognizing ownership in data should be rejected. Data in the Big Data analytics lifecycle are dynamic in nature, which makes it impossible to adequately define a delimited subject matter.[61] It would create an additional *enclosure,* which could stifle innovation, bearing in mind that the introduction of other property-like rights in data, such as the *sui generis* database right, has not necessarily stimulated investment in the creation of databases.[62] With respect to personal data, data ownership could lead data subjects to lose control over their personal data when selling them to companies (e.g. through contracts concluded by a click that grant permissions to use their data).[63] That said, personal data protection comes close to property in some respects and may be seen as a type of property, so that there seems to be no one property but many different property-like regimes.[64] As an outcome, there is **no one property** but **several ownership rights,** or other concepts leading to different data spaces (e.g. industrial data space, humanitarian data space, mobility data space).[65]

1.3.2 Contracts and technological protection measures (TPM)

In the absence of a recognized property right in data and because of the limitations of other legal protections as explained above, organizations tend to rely on other protection mechanisms, in particular on contractual restrictions and technological protection measures (such as access control, passwords or encryption) in order to obtain a sort of data exclusivity.[66]

These mechanisms may apply alternatively or cumulatively to the other protection mechanisms. There are however several downsides, such as the difficulty in regulating ownership issues in the Big Data contexts where there is a multitude of data sources and actors; the risk of interdependencies between all stakeholders; the unenforceability vis-à-vis third parties; and imbalances in bargaining power. Policy reforms are neceesary to remedy these downsides (see below III.3).

59 *Wolfgang Wiegand,* in: Geiser/Thomas/Wolf (ed.), Zivilgesetzbuch II: Art. 457–977 ZGB und Art. 1–61 SchIT ZGB, 5th ed., Basle 2015, Vor. ZGB 641 ff., N 6.

60 See the extensive discussion on the introduction of an EU sui generis database right, i.e. on whether such right has stimulated investment (Commission Staff Working Document, Evaluation of Directive 96/9/EC on the legal protection of databases [note 50]).

61 *Strowel* (note 2), 65 (the fluid nature of data may pose problems to *"adequately define the subject matter, scope of protection and ownership of a possible new right on industrial data"); Benhamou* (note 57), 520.

62 See the two evaluations of the Database Directive: the 2018 Commission Staff Working Document, Evaluation of Directive 96/9/EC on the legal protection of databases (note 50); the 2005 European Commission, First Evaluation of Directive 96/9/EC on the Legal Protection of Databases, DG Internal Market and Services Working Paper, 12.12.2005.

63 *Benhamou/Tran* (note 3), 579.

64 See references in note 8.

65 See the EU Data Strategy that intends to support the creation of *"common European data spaces"* in strategic sectors and domains of public interest (Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, A European strategy for data, COM(2020) 66 final, Brussels, 19.2.2020 (hereafter European strategy for data), 21). With respect to humanitarian law, see for instance the recent Draft Resolution, Restoring Family Links while respecting privacy, including as it relates to personal data protection, of October 2019, 33rd International Conference of the Red Cross and Red Crescent, Geneva, 9–12 December 2019, 33IC/19/12.4DR rev1, which leads to the creation of humanitarian data space and urges the states not to require personal data to the ICRC. With respect to personal data for instance, instead of a real property right or data ownership, one can speak of control rights (e.g. access, portability, rectification, erasure, objection to processing), which are unwaivable (see below II.3.2) and shall in turn lead to an obligation of "privacy by design" or "access by design" imposed to the data controllers/processors (see note 7).

66 See the Commission Staff Working Document, Evaluation of Directive 96/9/EC on the legal protection of databases (note 50), 31 concluding that, following a public consultation, contracts are the most widespread form of database protection (52.5% answered that they always rely on contracts, 78.6% in the publishing sector, 80% in IT sector and 100% in the finances and banking sector). See also *Strowel* (note 2), describing contracts and technological protection measures as other data appropriation mechanisms.

2. Exceptions and other data access flexibilities

In some cases, there are exceptions and data access flexibilities allowing the access to inputs, algorithms or outputs without the rightholder's authorization, or other mechanisms forcing the rightholder to lock-out its data (below collectively referred to as "flexibilities").

2.1 Privacy and data protection

There are flexibilities which may address the compliance difficulties associated with data protection rules.

From the outset, pure **raw data** (e.g. meteorological data generated by a building sensor), **anonymized data** (i.e. data where all personal identifiers have been removed irreversibly) and **pseudonymized data** (i.e. data where all personal identifiers have been substituted) can facilitate compliance, as they are excluded respectively release organizations from certain data protection obligations.[67] Consequently, data providers may consider whether anonymization or pseudonymization is appropriate prior to any data transfer, and data users may consider whether the use of such data is sufficient to achieve their goals.[68] For example, an organization may hold a dataset containing personal data in one data store, and share an anonymized version for analytic purposes in a separate area (e.g. for apps that track traffic flows, the underlying geolocation data may be anonymized before being sent to the research organization).[69]

However, these techniques carry their own complications, as anonymized data are requalified as personal data when they can be traced back to the data subject thanks to the combination of other data and databases.[70] In this respect, most regulators and experts tend to rely on a risk-based approach (i.e. when the risk of re-identification can be mitigated so it is no longer significant), as opposed to a zero risk-based approach (i.e. when the risk is eliminated taking into account all available data and means).[71] Even with this practical risk-based approach, the dynamic nature of data could transform an anonymized data into a personal data depending on future techniques and practices. Assuming that risk-based approach re-

67 See above II.1.1.

68 In Switzerland, the question of anonymization keys was however unclear, until recently. The Swiss Data Protection Authority considered in 2013 that, for banking data sent abroad, both the standpoint of the data provider and the data recipient should be considered, so that all information transferred to a third party would qualify as personal data, even if the data recipient has no access at all to the content of the data. See Préposé fédéral à la protection des données et à la transparence (PFPDT), 22e Rapport d'activités du PFPDT – 2014/2015, 18.06.2015, 67 ff. This statement seems now outdated further to a recent decision by the Swiss Supreme Court (*Tribunal fédéral* or TF) considering that pseudonymized data are excluded from data protection rules if pseudonymization effectively prevents the data user to identify the data subject (TF, case 4A_365/2017, judgment of 26 February 2018). See also *Meier* (note 1), 55.

69 See the UK Information Commissioner's Office (ICO), Big data, artificial intelligence, machine learning and data protection, March 2017 (hereafter UK ICO Report (2017)), 59, giving the example of Telefonica's Smart Steps that uses geolocation data of mobile phones to track the movement of crowds of people, and many other examples of the use of anonymized data. The personal data is stripped out before the analysis and the anonymized data is aggregated to gain insights about the population and combined with market research data from other sources.

70 Some commentators and studies have pointed to examples of where it has apparently been possible to identify individuals in anonymized datasets, such as a recent MIT study that looked at records of three months of credit card transactions for 1.1 million people and claimed that, using the dates and locations of four purchases, it is possible to identify 90 percent of the people in the dataset. On the other hand, other experts have found shortcomings in these studies and pointed out that more robust anoymization may be possible too. See UK ICO Report (2017) (note 69), 59, and the cited references.

71 In Switzerland, see TF, case 4A_365/2017, judgment of 26 February 2018; *Célian Hirsch/Emilie Jacot-Guillarmod,* Les données bancaires pseudonymisées: du secret bancaire à la protection des données, RSDA (2020), 151–167. In Europe, see UK ICO Report (2017) (note 69), 59; Case C-582/14, *Patrick Breyer v Bundesrepublik Deutschland,* [2016] ECLI:EU:C:2016:779, paras. 45–46 (anonymized data *"if the identification of the data subject is prohibited by law or practically impossible on account of the fact that it requires a disproportionate effort in terms of time, cost and man-power, so that the risk of identification appears in reality to be insignificant"*). There may be also sector-specific regulations. In the US, see the US Health Insurance Portability and Accountability Act (HIPAA), which qualifies "de-identified health information" (DHI) as data that "does not identify an individual" when 18 identifiers have been removed.

mains the trend, organizations should focus on mitigating the risks to the point where the chance of reidentification is extremely remote.

As to the principles, there is room for flexibility.[72] For instance, in order to comply with the transparency principle, organizations may offer multilayered privacy notices (transparency) and, with the minimization principle, they may try new approaches, such as assessing whether a new purpose is fair compared to the initial purpose.[73] With respect to the lawfulness principle and when organizations rely on the consent, they may try new approaches, such as seeking graduated consent, in which the terms of granting consent will vary with the severity of processing (e.g. processing activities leading to data transfer and several purposes should lead to a higher level of consent) and/or in which people consent to different uses of their data throughout their relationship with a service provider, rather than having a simple binary choice at the start ('just in time' notifications).[74] There may also be automation models of consent for both the collection and withdrawal of consent.[75] Organizations must however bear in mind that the consent may be withdrawn at any time by the data subject, so that other lawful grounds may be contemplated, if not preferred. However, the lawful basis of data processing should be decided in advance and could not be switched from consent to another lawful ground.[76] When organizations rely on other legitimate interests, they may argue their own interests or third parties' interests (e.g. preventing fraud or misuse of services), which nevertheless requires a balancing of these interests and those of the data subject.[77]

Finally, privacy laws, such as the GDPR, may provide other **flexibilities** with possible exceptions for research or archival purposes, allowing Member States to limit certain individual rights under certain circumstances.[78]

2.2 Intellectual Property Rights

2.2.1 Input

With respect to input data, a distinction must be made between two situations at least, depending on whether the input is protected or not by copyright or other similar intellectual property rights.

The first situation is when the input is **protected by copyright.** There is a need to know whether the use of the input for Big Data analytics is covered by copyright. Most jurisdictions, including Switzerland, seem to agree that the use of inputs systematically triggers the reproduction right, whether or not there

72 For compliance solutions in Big Data in Swiss law, see *Meier* (note 1), 69 ff.; in EU law, see UK ICO Report (2017) (note 69), 59.

73 See *Meier* (note 1), 70.

74 UK ICO Report (2017) (note 69), 29 *("Obtaining meaningful consent is often difficult in a big data context, but novel and innovative approaches can help")* and 59, giving the example of a point when an app wants to use mobile phone location data or share data with a third party, the user can be asked to give their consent. We must bear in mind that a "regular consent" is difficult to apply to Big Data situations as it is generally not known which data exactly will be used, in which combinations and by whom. Those aspects make it more difficult to evaluate the risks for the personnality of the data subject. Also, consent is often limited in time, so that the reuse of data after a predetermined period of time needs a renewed declaration from the data subject (*Meier* [note 1], 67 and 74).

75 UK ICO Report (2017) (note 69), 59, giving the example of software agents providing consent on user's behalf based on the properties of certain applications, as well as sensors, smart devices and other types of user positive actions, which could constitute consent (e.g. gesture, spatial patterns, behavioral patterns, motions).

76 See Article 29 Data Protection Working Party, Guidelines on consent under Regulation 2016/679, adopted on 28 November 2017, last revised and adopted on 10 April 2018, WP259 rev.01 (hereafter Working Party 29, Guidelines on consent), 22 ff. *("it is very important that controllers assess the purposes for which data is actually processed and the lawful grounds on which it is based prior to collecting the data (e.g. customer data may be based on contract, consent and legal obligation […] In cases where the data subject withdraws his/her consent and the controller wishes to continue to process the personal data on another lawful basis, they cannot silently migrate from consent (which is withdrawn) to this other lawful basis")*. For example, it is not allowed to retrospectively utilize the legitimate interest basis in order to justify processing, where problems have been encountered with the validity of consent. Any change in the lawful ground for processing must be notified to a data subject in accordance with the information requirements in articles 13 and 14 and under the general principle of transparency.

77 See Article 29 Data Protection Working Party, Opinion 06/2014 on the notion of legitimate interests of the data controller under Article 7 of Directive 95/46/EC, adopted on 9 April 2014, WP 217.

78 For research purpose, see article 89 (2) combined with articles 5, 16, 18 and 21 and Recitals 156 "data minimization" and 159. For archiving, see article 89 (3) combined with articles 15, 16, 18, 19, 20 and 21 and Recital 156.

is identical or partial reproduction of the input. This is linked to the broad interpretation of the reproduction right, which covers identical, partial, direct or indirect reproduction by any means, in whole or in part, i.e. when the input is used as training data (e.g. plagiarism detection software), or is recognizable in the output (e.g. AI-generated music song or painting containing thousands of inputs).[79] A question to policy-makers would be whether to shift from this broad interpretation to a perceptibility approach or an economic approach, and to reject copyright protection, at least when the input is not recognizable or even inexistent in the output, as there would be no economic exploitation of the input (economic approach), or as the public would not perceive the input (perceptibility approach).[80]

There may nevertheless be exceptions allowing the use of copyrighted input. One noteworthy exception in a Big Data context is the exception for text and data mining (TDM). TDM techniques usually involve copying of third-party data or databases in order to drive them through algorithms,[81] so that they are likely to fall under copyright protection and render these extractions and copying acts unlawful if they are carried out without permission from the relevant right holders or without an exception.

In **Switzerland,** the TDM exception is now provided by the Swiss Copyright Act, which makes the copy of data lawful if they can be lawfully accessed and for the purposes of scientific research.[82] If this exception is a welcome development for authorizing Big Data activities, it is subject to three important restrictions at least: *(i)* it is limited to copy for scientific purposes, which excludes uses for primarily commercial purposes; *(ii)* it is limited to acts of reproduction, which excludes the communication or commercialization of output; *(iii)* input data shall be lawfully accessed, which excludes data or databases subject to contractual restrictions or TPM, i.e. all trade secrets and proprietary databases.[83]

In **Europe,** the TDM exception is now expressly provided by the Directive on Copyright in the Digital Single Market (DSM Directive), which applies to both copyright and *sui generis* database subject matters.[84]

79 In Swiss law, see article 10 (2) let. a Swiss Copyright Act (the reproduction right covers the right to *"produce copies of the work, such as printed matter, phonograms, audiovisual fixations or data carriers");* see *de Werra/Benhamou* (note 18), 753, giving the examples of the Edmond de Balamy portrait based on 15'000 preexisting portraits or Google Dream trained on open access images. In EU law, see article 2 Copyright Directive 2001/29/CE (the reproduction right covers the *"exclusive right to authorise or prohibit direct or indirect, temporary or permanent reproduction by any means and in any form, in whole or in part");* Case C-5/08, *Infopaq International vs Danske Dagblades Forening,* [2009] ECLI:EU:C:2009:465, para. 51; see *Strowel* (note 2), 12, indicating that the reproduction right may also apply, when the use relates to a raw data embedded in a copyrighted file, as the underlying raw data often overlaps, if not merges with the embedding copyrighted file for which copyright protection applies.

80 For instance, compare the preexisting painting "Your World Without Paper" by Amel Chamandy and the painting "85.81%_match" by Adam Basanta. Both paintings contain similarities that are invisible to the naked eye, while an AI detects 85% similarities. This assessment of public perceptibility should be based on a human-centered approach, as a machine-centered approach (e.g. an AI or an "augmented human" using Google glasses trained to detect similarities) would lead to recognize similarities.

81 Although TDM operations vary in complexity and sophistication, three common steps are generally followed: (1) access to content, (2) extraction and/or copying of content and (3) mining of text and/or data and knowledge discovery. For a summary of the steps and the possible legal issues, see *Christophe Geiger et al.,* The Exception for Text and Data Mining (TDM) in the Proposed Directive on Copyright in the Digital Single Market – Legal Aspects, Centre for International Intellectual Property Studies (CEIPI), Research Paper No. 2018-02, February 2018.

82 Article 24*d* Swiss Copyright Act *("1. For the purposes of scientific research, it is permissible to reproduce a work if the copying is due to the use of a technical process and if the works to be copied can be lawfully accessed")*.

83 Message relatif à la modification de la loi sur le droit d'auteur, à l'approbation de deux traités de l'Organisation Mondiale de la Propriété Intellectuelle et à leur mise en œuvre du 22 novembre 2017, FF 2018 559 (hereafter Message P-LDA), 594 f.

84 In Europe, other existing exceptions could be argued to facilitate the use of original data, such as the exception for private use (which is limited to strict private uses, thus excluding research and business organizations); the right to extract and/or re-utilize insubstantial parts of a database protected by sui generis right; the transient or incidental copying exception of article 5 (1) of the Directive 2001/29/EC of the European Parliament and of the Council of 22 May 2001 on the harmonisation of certain aspects of copyright and related rights in the information society (InfoSoc Directive); and the reverse engineering exception of the Directive 2009/24/EC of the European Parliament and of the Council of 23 April 2009 on the legal protection of computer programs (Software Directive) (however limited to temporary reproductions, transient or incidental). For an in-depth analysis of all exceptions pos-

However, this exception is also subject to important restrictions: *(i)* it is limited to *"research organisations"* and *"for the purposes of scientific research"* (art. 3 DSM),[85] which excludes independent researchers who are not affiliated with a research organization or when it is conducted in a commercial context or with the involvement of private parties (at least, to grant them access and use of the data retrieved from TDM); *(ii)* TPM can override this exception (although it is a mandatory exception and cannot be overridden by contract).[86]

In the **United States,** TDM activities could be subject to the fair use exception, which can be raised in relation to a large number of different factual circumstances based on four factors.[87] When assessing the first of the four factors (purpose and character of the use), the more transformative the use, the more likely it is *not* an infringement of the original copyright. In the Big Data context, the key question will be whether the output (derivative work) is transformative enough compared to the pre-existing input.[88] The answer is, however, not clear-cut and will depend on the facts of each case. The data user may argue that the Big Data analysis sufficiently alters the original dataset, while the data supplier may argue that the transformative threshold is not met.

The second situation is when the input is not protected by copyright, such as training and technical data. In the absence of such protection, such data can be in principle freely used. However, their use can be limited in certain situations. First, data producers can impose contractual restrictions or TPM, creating a kind of data exclusivity.[89] Moreover, other laws can restrict their use in certain circumstances. In Switzerland, the UCA, in particular article 5 let. c UCA, prohibits the reuse of third-party work by technical processes without corresponding investments. As previously stated, the Swiss Supreme Court refuses to restrict competitors based on this provision when the third party user *(repreneur)* undertook certain investments or when the initial data producer has amortized expenses. Therefore, in two leading decisions, the data producer of a database could not prohibit third party users *(repreneur)* from reusing the underlying data to build their competitive website, as the data producer had already amortized his/her expenses at the time of the competitive database respectively as the third-party (*repreneur*) undertook programming efforts to extract and restructure the data.[90]

sibly applicable to Big Data activities, see *Geiger et al.* (note 81).

85 "Research organization" means any entity *"the primary goal of which is to conduct scientific research or to conduct scientific research and provide educational services"* (article 2.1 DSM Directive). For an in-depth analysis, see *Strowel* (note 2), 9, explaining that this exception also excludes the use by independent researchers who are not affiliated with a research organization and by private partners of public-private partnerships making the functioning of those consortia quite complex (despite the Recital 10 providing that research organizations can *"benefit from the exception when they engage into public-private partnerships"*).

86 *Strowel* (note 2), 14, noting that a broader exception is provided by article 4, which applies to commercial uses, but which can be overridden by contracts and TPM. The author also notes that broader TDM exceptions have been introduced in some EU Member States (e.g. the UK, France, Germany, Estonia) and are part of current discussions on copyright reforms (e.g. Singapore, Australia), but most of them are likely not applicable to TDM in a Big Data context for business activities, as they are usually limited to non-commercial research.

87 See *Yu* (note 50), 913, referring to "the modalities of access" (in lieu of "exceptions").

88 *Maier/Sibble* (note 23), 22. In Switzerland, see *Benhamou/Tran* (note 3), 583.

89 Such contractual restrictions have been considered as valid by the CJUE in the *Ryanair v PR Aviation decision* (Case C-30/14, *Ryanair v PR Aviation,* [2015] ECLI:EU:C:2015:10, Recital 39: *"it is clear from the purpose and structure of Directive 96/9 that Articles 6 (1), 8 and 15 thereof, which establish mandatory rights for lawful users of databases, are not applicable to a database which is not protected either by copyright or by the sui generis right under that directive, so that it does not prevent the adoption of contractual clauses concerning the conditions of use of such a database"); de Werra* (note 34), 173.

90 ATF 134 III 166 = JdT 2008 I 399 (*Compendium* case), c. 4.3 *("Une disproportion évidente entre les investissements du demandeur et ceux du défendeur n'existe plus, lorsque le demandeur a pu amortir ses coûts. À ce moment la protection légale de l'article 5, lit. c LCD cesse, et il n'est plus illicite de reprendre la prestation");* ATF 131 III 384 (*Such-Spider* case), c. 4.4.2 *("programmation d'un système pour reprendre les données présentes dans des bases appartenant à autrui et pour les préparer, du moins lorsque le résultat du travail d'autrui est d'abord démembré et réassemblé différemment [investissement qui] n'est pas minime au point que la reprise et l'exploitation de la prestation d'autrui se fassent telles quelles").*

2.2.2 Output

With respect to output, the main question is to know whether such output deserves protection. Most jurisdictions, including Switzerland, tend to consider that machine-generated data fall out of the scope of copyright protection.[91] Indeed, a certain level of creativity is required (subject matter), thus only natural persons with control over the creation can create copyrighted works.[92] This would exclude machine-generated data, which do neither rely on human creativity nor on the control or supervision of the author. A distinction must however be made between computer-assisted-works (i.e. works created with human intervention) which are copyrighted works granted to the producer, and computer-generated-works (CGW) (i.e. works created without human intervention) which remain unprotected in the public domain due to a human-centered approach.[93] This distinction depends on the causal link between the developer and/or the user selecting the training data and the resulting output, which may be considered as created by them, only to the extent that they had a creative added-value in the output. The more autonomous the AI, the less likely it is to have a causal link between the developer and/or the user.[94]

In sum, due to the broad interpretation of the reproduction right and apart from CGW that are in the public domain, the use of inputs and outputs falls within the scope of copyright protection and is subject to the authorization of the rightholder, unless an exception applies.

2.3 Other flexibilities

Other data access flexibilities can be found in certain jurisdictions. This is particularly the case in Europe with sector-specific or horizontal instruments that aim to grant a greater access to data,[95] in particular with the free flow of non-personal data,[96] the non-protection of public sector information (e.g. geographical information, statistics, weather data, data from publicly funded research projects and digitised books from libraries)[97] and government access to pri-

[91] For an analysis, see *de Werra/Benhamou* (note 18), 756 ff.

[92] In Switzerland, this arises from the notion of copyrighted works, defined as *"literary and artistic **intellectual** creations with an individual character, irrespective of their value or purpose"* (article 2 (1) Swiss Copyright Act), as well as the notion of author, defined as *"the **natural** person who has created the work"* (emphasis added); ATF 130 II 168; Message P-LDA (note 83), 587 *("On parle de création de l'esprit lorsqu'elle est le fait d'un être humain. Par conséquent, l'auteur […] doit nécessairement être une personne")*. For an analysis in Swiss law, see *de Werra/Benhamou* (note 18), 756 ff. In the EU, see *Infopaq International A/S v Danske Dagblades Forening* (note 79), which stipulates that copyright protection requires the author's own intellectual creation and needs a human-intervention. In the US, see *Feist Publications vs Rural Telephone Service Company, Inc.,* 499 U.S. 340 (1991) (copyright protects only intellectual efforts based on the intellectual creative power). In Australia, see *AU-Acohs Pty Ltd vs Ucorp Pty Ltd* [2012] FCAFC 16 (no protection for a computer-generated work).

[93] In Swiss law, see *de Werra/Benhamou* (note 18), N 123–125. In the EU, see *Ryan Abbott,* in: Aplin (ed.), Research Handbook on Intellectual Property and Digital Technologies, Cheltenham/Northampton 2018, 338–362.

[94] For an analysis, including the different stakeholders involved in AI and Big Data activities, see *de Werra/Benhamou* (note 18), 757.

[95] For recent developments, see the European strategy for data (note 65), with the aim to create a single market for data, both personal and non-personal. This document shall be read in conjunction with, and is accompanied by a White Paper on Artificial Intelligence (White Paper On Artificial Intelligence – A European approach to excellence and trust, COM(2020) 65 final, Brussels, 19.2.2020) and the Digital Strategy (Shaping Europe's Digital Future, available at <*https://ec.europa.eu/commission/presscorner/detail/en/fs_20_278*> [accessed 2 October 2020]). See also, the legislative framework for the governance of common European data spaces, which aims to create sectoral European data spaces on health, skills and mobility <*https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12491-Legislative-framework-for-the-governance-of-common-European-data-spaces*> [accessed 2 October 2020]. For an analysis of the situation in Switzerland, in particular for the access to non-personal data, see the report of *de Werra* (note 6), 364 ff.

[96] In the EU, see the Regulation (EU) 2018/1807 of the European Parliament and of the Council of 14 November 2018 on a framework for the free flow of non-personal data in the European Union (FFD Regulation), that further increases the cross-border exchange of non-personal data, defined by opposition to personal data. See also *Strowel* (note 2), 15–16. In Swiss law, see *de Werra* (note 6), 364 ff.

[97] The open data movement (data whose access, reuse, or redistribution is free, subject at most to the requirement of attribution) has developed over the last decades. Combined with the increase of transparency within public administration, the open data movement has led to the adoption of many laws on the freedom of information or the access to public documents. See the 2008 call of the OECD for governments to open up their data, the 2009 initiative by the U.S. government regarding the data.gov website, or the EU and Member States policies on open data. In the

vately-held data (e.g. machine-generated data with IoT).[98] Competition law may also contain further restrictions, as it might allow to avoid lock-in effects and impose specific access requirements, such as data access, data portability and platform interoperability.[99] As an example of data access requirements, the Court of Milan recently ruled that Ryanair's refusal to grant access to its database to the online travel agency Viaggiare S.r.l. amounted to an abuse of dominant position in the downstream market of information and intermediation on flights.[100] As an example of interoperability requirements, the EU Commission ordered Microsoft to disclose, within 120 days, complete and accurate interface information which would allow rival vendors to interoperate with Windows, and to make that information available on reasonable terms.[101] Particularly noteworthy is the fact that the requisite degree of interoperability was a difficult and disputed issue, and that Microsoft has been subsequently fined for charging unreasonable prices for access to interface information.[102] To promote competition, the European Commission suggested the use of fair, reasonable and non-discriminatory (FRAND) licenses to facilitate the use and dissemination of machine-generated data.[103] Companies should also pay attention to industrial standards and the possibility of being subject to FRAND terms in cases of standard essential patent (i.e. invention that is necessary to use to comply with technical standards). If a standard essential patent is not in the public domain or shared by the owner, complying with technical standards will expose companies to infringement liability,[104] while the owner may be subject to scrutiny under competition law; measures have been taken to prevent such potentially abusive conduct.[105]

Finally, to ensure the effectiveness of these exceptions, some jurisdictions provide a "no-contractual-override" provision (i.e. unenforceability of contrary contractual provisions that circumvent the safeguards provided by these exceptions). Provisions protecting some exceptions from contractual override may be found for instance in the Database Directive (articles 6 (1) and 15), the Software Directive (articles 5 (2) and 8) and the TDM exception (article 3 (2) DSM Directive).[106]

EU, see the recent 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 (known as the 'Open Data Directive'), which entered into force on 16 July 2019 and replaces the Public Sector Information Directive 2003/98 (known as the 'PSI Directive'). The Open Data Directive aims to encourage Member States to facilitate the reuse of public sector data with minimal or no legal, technical and financial retrictions. See also the EU Guidance on private sector data sharing (available at *<https://ec.europa.eu/digital-single-market/en/guidance-private-sector-data-sharing>* [accessed 2 October 2020]), that intends to provide a framework which supports the supply of private sector data to public bodies under preferential conditions for reuse. See *Strowel* (note 2), 11; *Yu* (note 50), 914 and the cited references.

98 See Commission Staff Working Document on the free flow of data and emerging issues of the European data economy, SWD(2017) 2 final, Brussels, 10.1.2017, 10: *"Public sector bodies may also have a legitimate interest in obtaining access to certain data. This has relevance for the provision of statistical information, urban planning, environmental protection, civil protection, etc. In most situations, public sector bodies would need aggregate information only."*

99 For instance, the refusal to grant access to data or to transfer data could constitute, under specific circumstances, an abuse of dominant position, in particular when the dominant undertaking prevents the emergence of innovative competitors because customers are locked-in (portability) or because competitors have no access to the data (data access), which is an essential condition to entering or efficiently operating in a specific market. On this basis, it is sometimes argued that a powerful undertaking could maintain its dominant position in the market due to its possession of a large amount of information that is difficult to duplicate (essential facility doctrine) and abuse this position by refusing to share that information and thus excluding competitors. See *Adrien Alberini/Yaniv Benhamou,* Data portability and interoperability: an issue that needs to be anticipated in today's IT-driven world, Expert Focus (2017), No. 8, 518 ff., 520; *Adrien Alberini,* Pouvoir de marché dans le secteur numérique: l'accès à de larges quantités de données est-il suffisant?, RSDA (2019), 31–42, 31 ff.

100 *Viaggiare S.r.l. vs Ryanair Ltd,* Court of Milan, Decision of 4 June 2013.

101 See in particular the summary of the case and remedies, in: European Commission, Commission concludes on Microsoft investigation, imposes conduct remedies and a fine (24 March 2004), IP/04/382, available at *<https://ec.europa.eu/commission/presscorner/detail/en/IP_04_382>* [accessed 2 October 2020].

102 *Alberini/Benhamou* (note 99), 520, and the cited references.

103 *Yu* (note 50), 927.

104 *Maier/Sibble* (note 23), 30.

105 *De Werra* (note 34), 147 and the cited references.

106 Unlike other exceptions that can be contractually overridden, such as temporary acts of reproduction, the extraction of insubstantial parts of a database and the national

In sum, depending on the facts at stake and on the jurisdiction concerned and in order to avoid conflicts, organizations may prefer, when feasible, to negotiate licensing agreements with third parties to access data or databases in which they do not hold any right.

3. Contracting Big Data

Data sharing (or access) is done mainly through contractual agreements.[107] It is therefore required to carefully assess the multiple agreements governing the data sharing between the various actors, taking into consideration the type of data involved.

3.1 Nature of transfer: assignment, license or sale

Data may be voluntarily shared with third parties, provided that the data is free of right or that the transferor owns the rights.[108] This is usually done via **"data sharing agreement"** (DSA), which can be defined as an agreement between two or more parties concerning the sharing of information of any kind.[109] DSAs usually refer to a broad typology of agreements.[110] Their qualification is important, in particular to determine the legal provisions from which the parties cannot derogate and/or to remedy possible gaps.[111] DSAs are subject to two main types of contract: assignment or license.

– **Assignment** is the definitive transfer of right: the assignor loses the control over his right definitively.[112] Relatedly, the assignee becomes the owner of the assigned right, which may be freely used (e.g. assigned or licensed to other third parties).[113]

  Assignment is possible for any type of data,[114] except for personal data which are not assignable in most jurisdictions, at least in those providing that data subjects have a withdrawal right at any time (what can be called the "unwaivability of data protection").[115]

– A **license** is a temporary usage right:[116] the licensor remains the right owner and consents to the use of his right by the licensee. The licensee may not transfer, sub-license or assign the right to third parties unless expressly authorized.[117] If the data is protected by an exclusive right (copyright, personal data), consent simultaneously

research exceptions (limited to non-commercial purposes and to the "sole purpose of illustration for teaching or scientific research") cannot be contractually circumvented. See *Yu* (note 50), 917 referring to the UK legislation (2014 Regulations for Copyright and Rights in Performances, Quotation and Parody), which renders contractual terms that purport to restrict lawful acts, such as quotation, caricature, parody and pastiche, unenforceable.

107 See also Conseil Supérieur de la propriété littéraire et artistique, Mission Intelligence artificielle et Culture, Rapport final, 27 January 2020, 47, that reaches an equivalent conclusion in the matter of AI-generated art.

108 *Aude Peyrot/Sevan Antreasyan,* Succession 2.0: les biens numériques, Revue de droit privé et fiscal du patrimoine – Not@lex (2016), 20–33, 23.

109 Data Transfer and Use Agreements (DTUA), or Data Usage Agreements (DUA) are also used in practice. *Debussche/César* (note 11), 77.

110 *Benhamou/Tran* (note 3), 579, listing several forms and circumstances of DSA, e.g. whether the data transfer is subject to a reciprocal exchange of data; free of charge or against payment; as part of a principal or ancillary service; as part of licensing agreements; consultancy or work contracts; joint ventures; or service agreements.

111 The agreement *"should explain why the data sharing is necessary, the specific aims [the company] has and the benefits [it] hope to bring to individuals or to society more widely"* (UK's Information Commissioner Office [ICO], Data Sharing: Code of Practice, 2019, 26).

112 It has an absolute effect *(erga omnes)* contrary to the licence who only bears a relative effect *(de Werra/Behnamou* [note 18], 25).

113 *Jacques de Werra,* Commentaire romand, Propriété intellectuelle, Basle 2013, N 7 ad Art. 16 LDA; *Denis Barrelet/Willi Egloff,* Le nouveau droit d'auteur: commentaire de la loi fédérale sur le droit d'auteur et les droits voisins, 3rd ed., Bern 2008, N 10 *ad* Art. 16 LDA.

114 E.g. for trade secrets, it is common practice to use terms close to property, ownership, such as "assignment", "asset transfers" or "sale" (as opposed to other countries relying on tort law, such as France). See *Strowel* (note 2), N 130.

115 See however in Switzerland the decision of the Supreme Court which accepted the validity of an irrevocable contractual commitment in a decision concerning a model assigning her image rights of a photograph to an agency: ATF 136 III 401 = JdT 2011 II 508 (those are personality rights that are not part of the intangible core of the human essence *(noyau intangible de l'essence humaine)*; e.g. voice or image according to the importance their commercialization has taken for the past few years); *Benhamou/Tran* (note 3), 579.

116 A license expires at the agreed expiration of the contract. In the absence of such agreed expiration, license may be usually terminated by either party for cause. See *Benhamou/Tran* (note 3), 579.

117 *De Werra* (note 113), N 7 *ad* Art. 16 LDA.

constitutes a waiver to claim protection.[118] When the licensor grants a perpetual license (e.g. a definitive usage right by downloading the file against a one-off lump sum payment), the license corresponds to a sale.[119]

Licensing is possible for any data, whether protected by an exclusive right (copyrighted or personal data) or not (raw data).[120] For instance, for personal data, despite the unwaivability of data protection, persons commonly trade their data in exchange for ("free") services through contracts concluded by a click, as confirmed by the recent Digital Content Directive.[121] This leads to a generalized (cross-)licensing scheme, according to which users become licensor over their data and, at the same time, licensees for the services (e.g. apps or other software they use) as counterparts.[122] For raw data, data producers protect their investment via licenses, and detail the modalities of access and reuse, in particular when the content is a trade secret or when the content is fragmented, with the value of the data found in the way it is structured into a dataset.

3.2 Limitations to the transfer

Parties to a DSA are bound to comply with mandatory rules arising from the applicable law(s), in particular with general rules of contract law, specific rules applying to each type of data and possible specific **sector-specific regulations** (e.g. for financial or health data).

With respect to **general rules of contract law,** most of such general rules concern the formation and execution of an agreement. In particular, the contracting parties must have validly agreed to the terms (e.g. choice of law and/or court)[123] and these terms must be enforceable (e.g. no excessive commitments by the data transferor). There may also be formal requirements[124] or mandatory termination rights.

With respect to specific rules, some **copyright laws** provide for the **unwaivability of moral copyrights:** while economic rights are transferrable in whole or in part, moral rights are more difficult to transfer.[125] Unlimited contractual waivers are held

118 *Julien Rouvinez,* La licence des droits de la personnalité: Étude de droit privé suisse, Recherches juridiques lausannoises, Faculté de droit de l'Université de Lausanne, Zurich 2011, N 170.

119 In the EU, the CJUE considered that *"from an economic point of view, the sale of a computer program on CDROM or DVD and the sale of a program by downloading from the internet are similar. The on-line transmission method is the functional equivalent of the supply of a material medium"* (Case C-128/11, *Usedsoft GmbH vs Oracle International Corp.,* [2012] ECLI:EU:C:2012:407, para. 61). In Canada, a decision of 12 July 2012 *(Entertainment Software Association vs Société canadienne des auteurs, compositeurs et éditeurs de musique,* 2012 CSC 34, [2012] 2 R.C.S. 231) considered that with the download of games online *"there is no practical difference between the purchase of a copy in a shop, receive the copy by mail or download the same copy online. Internet is just a technological mean (taxi) ensuring the delivery of the durable copy to the end user"*. See also *Nicolas Rouiller,* L'immatériel et le contrat, in: Rapport suisse à l'Association Capitant des amis de la culture juridique française, Paris 2014, 12: rules on sale are applicable in case of supply of turnkey softwares *(livraison "clef en main" de logiciel)*, the obligations being fully performed in exchange of services. *Benhamou* (note 57), 520 noting that the sale however only grants a control over a copy of the data and is different from the assignment and license: unlike the assignment, the control right remains by the transferor; unlike the license, the usage right is perpetual.

120 *Benhamou/Tran* (note 3), 579 and the cited references.

121 See article 3 para. 3 of the Directive (EU) 2019/770 of the European Parliament and of the Council of 20 May 2019 on certain aspects concerning contracts for the supply of digital content and digital services: *"This Directive shall also apply where the trader supplies or undertakes to supply digital content or a digital service to the consumer, and the consumer provides or undertakes to provide personal data to the trader, […]"*. Either **directly from the consumer** (data subject) who usually consents to share personal data via standard terms of use on online platforms when registering or using the services (e.g. social network, search engine, content streaming), or **indirectly from data suppliers** (including data brokers, such as Axciom, Corelogic, Datalogix, Bluekai). See also *Strowel* (note 2), 9.

122 *Strowel* (note 2), N 120, noting that the number and scope of the data licenses has the effect that, despite the unwaivability of data protection, large segments of private life are traded in practice (think about Facebook).

123 In particular with the general terms and conditions, sometimes imposed by data providers to data users, such as internet users. See *de Werra* (note 34), 174, referring to a decision where the choice-of-court has been made on the basis of a reference made in an email to the general terms and conditions, which was admitted by the Swiss Federal Supreme Court (ATF 139 III 345, c. 4.4).

124 For instance when the written form is mandatory (e.g. article 32 GDPR: data processing agreement between a controller and a processor).

125 *Yaniv Benhamou,* Posthumous replications: rights and limitations, notion of original and copies, in: Mosimann/Schönenberger (ed.), Kunst & Recht 2017/Art & Law 2017: Referate zur gleichnamigen Veranstaltung der Ju-

valid in certain jurisdictions, such as under UK and US copyright laws, but other jurisdictions limit such waivers. Under German and Swiss law, moral rights may be subject to contractual waivers provided that they are not excessive.[126] For instance, an author may agree in advance to specific known and determinable modifications by third parties, but not to any modification. Under French law, blanket waivers to moral rights are deemed null and void, even if the contractual waiver is deemed valid under other applicable laws.[127] For instance, French courts tend to apply French copyright law as a matter of public policy *(ordre public)* or as the applicable law *(lex protectionis)* according to which moral rights could not validly be waived (at least to the broad extent which is allowed under UK copyright law),[128] even if there is a valid contractual waiver under UK law. From this perspective, companies should account for the fact that some waivers to moral rights could be valid under the law of one country but struck down by protective rules of another country.[129]

Some **privacy laws** provide for the **unwaivability of data protection:** while personal data seem tradable like any other IPR, some privacy laws consider that (certain) aspects of personality may not be transferred or waived, for instance when the consent to the use of data may be withdrawn at any time.[130] For instance, under the GDPR, data subjects have the right to withdraw consent at any given time, without undue effort (usually with the same electronic means, such as one mouse-click, swipe, or keystroke)[131] and without detriment (in principle free of charge or without lowering service levels).[132] If consent is withdrawn, controller must **stop processing** the data operations based on consent and, if there is no other lawful basis justifying the processing (e.g. further storage), **delete** the data (article 17 (1)(b) and (3) GDPR).[133] Moreover, a data subject may even request erasure of data that is processed on another lawful basis. Then, controllers are obliged to assess whether continued processing of the data in question is appropriate, even in the absence of an erasure request by the data subject (see article 5 (1)(e) GDPR).[134]

3.3 Contractual good practices

Within the above mandatory limitations, parties to a DSA are free to agree on **additional terms and conditions applicable** to the data sharing, such as time of disclosure; warranties, accuracy and completeness of data; obligations of the receiving party to manage the data according to specific rules and to apply certain security measures to protect the data; right or prohibition of the receiving party to transfer onward/disclose the data to a third party; ownership of the data and intellectual property rights; payment of any consideration for the data sharing; confidentiality obligations; audit of the receiving party by the disclosing party or by the authorities; warranties on the power to disclose and receive data; duration of the agreement; governing law; and competent court.[135]

Such terms and conditions are almost unlimited, depending on the needs of the parties, the evolution of business models and the bargaining power of one party. Without being exhaustive, the following con-

ristischen Fakultät der Universität Basel vom 16. Juni 2017, Bern 2017, 149–168, 158 and the cited references.

126 Assignment of moral rights is debated by some scholars, see *de Werra* (note 113), N 16 ff. ad Art. 16 LDA.

127 As was confirmed in a recent decision of the French Cour de Cassation, which stated that the unwaivability of the right of integrity is a principle of public policy (ordre public), recalling the equally strong protective position that was adopted in the famous *Asphalt* case about colorization applying moral rights irrespective of any contractual waivers.

128 *Jacques de Werra,* The moral right of integrity, in: Derclaye (ed.), Research handbook on the future of EU copyright, Cheltenham 2009, 267–285, 267 and 276.

129 *Benhamou* (note 57), 158.

130 See *Benhamou/Tran* (note 3), 578.

131 Working Party 29, Guidelines on consent (note 76), 21: indicating that for instance *"when consent is obtained via electronic means through only one mouse-click, swipe, or keystroke, data subjects must, in practice, be able to withdraw that consent equally as easily*".

132 Working Party 29, Guidelines on consent (note 76), 21: *"Data subject should be able to withdraw his/her consent without detriment. This means, inter alia, that a controller must make withdrawal of consent possible free of charge or without lowering service levels."*

133 Working Party 29, Guidelines on consent (note 76), 21.

134 Working Party 29, Guidelines on consent (note 76), 21.

135 See *Alberini/Benhamou* (note 99), 521, concluding that it is highly advisable to anticipate in the formation of IT contracts issues like data retrieval, portability and interoperability, and recommending good practices based on a checklist, as the law is in many instances insufficient to address these issues.

siderations should be in our view carefully addressed:[136]

- The contract should **define the data** as precisely as possible, including the level of updates to be expected in the future, the quality of the data and compliance with third party rights, including intellectual property rights.
- The contract should define as precisely as possible the **scope of use,** i.e. who has a right to access, to reuse, and to sublicense the data, and under which conditions.[137]
- The contract should also define and allocate **the ownership** over both initial data (input) and derived data (output), especially since IP rights in derived data may be complex and even self-contradictory.[138] In the absence of a specific clause, ownership depends on the nature of the data, as the applicable legal regime provides different ownership rules:
  - On **initial data, copyright** provides that the original owner is the person who created the work, solely or jointly (creator principle) and the derivative owner the person whose copyrights have been transferred via license or assigment. **Data protection** provides that the original rightholder is the data subject and the derivative rightholder the person to whom the rights have been transferred via license. For **other data,** the original owner is the person who had the initial *de facto* control over the data (trade secret holder, database producer) and the derivative holder the person to whom the control has been granted (e.g. access to a database).[139] In the case of employment relationships, data generated by the employee are usually presumed to belong to the employer.
  - On **derived data** (i.e. when third parties use the data, compile, structure, re-format, enrich, analyze, take a license, or add value to the data), **copyright** provides that the derivative owner is the person who created the derivative work, which requires the consent of the initial owner (usually contained in the license agreement which defines who has the right to develop work products and who owns them). For **personal data,** one could argue that the same rules apply. For **other data,** the derivative owner is the person who created the derivative work, which does not specifically require the consent of the initial owner (i.e. unless otherwise expressly indicated, such as ownership or specific modalities of use, the derivative owner shall have the right to reuse).
- The contract should determine the **technical means and modalities for data access and/or exchange,** including the frequency of data access, maximum loads, IT security requirements and service levels for support.
- The contract should also include **liability provisions** to cover situations of supply of erroneous data, disruptions in data transmission, low quality interpretative work if shared with datasets, or the destruction/loss or alteration of data (if unlawful or accidental) that may potentially cause damages.
- Companies are also advised to define **audit rights** regarding the respect of the mutual obligations.

136 For further details, see the guidance issued by the European Commission with a list of considerations to help companies in the preparation and/or negotiation of DSAs among private companies (i.e. B2B) and between privates companies and the public sector (i.e. B2G) (Commission Staff Working Document, Guidance on sharing private sector data in the European data economy, SWD(2018) 125 final, Brussels, 25.4.2018, 18); UK's Information Commissioner Office (ICO), Data Sharing: Code of Practice, 2019, 26.

137 For instance, whether the right to access is limited to members of a certain group, or affiliates of a certain company, or limit the right to reuse to certain specific purposes. Companies should moreover consider if and how data may be licensed for reuse and include the necessary specifications in this regard. Sub-licensing may also be considered in the sense that it should either be expressly excluded, or the conditions under which it is allowed should be clearly stipulated. See *Debussche/César* (note 11), 81.

138 Defining output ownership is important even if the DSA concerns the use of input. Indeed, the data user could argue that their analysis transforms an initial non-copyrighted dataset into a derivative copyrighted work, while on the other hand the data supplier may argue that the change is not so great to transform the dataset into a copyrighted work.

139 See for instance, article 2 (2) of the Trade Secret Directive (a trade secret holder is *"any natural or legal person lawfully controlling a trade secret"*).

– The duration of the contract and possibilities for **termination** should of course be carefully considered, as well as the applicable law and dispute settlement options.

III. Selected policy considerations

This section contains selected policy considerations, it being specified *(i)* that it does not purport to analyze them in detail but simply to raise a few elements and *(ii)* that it addresses solutions at the national level mainly, while policy solutions should be contemplated at both the national/regional and international levels, so that the elements contained herein remain open for further research and discussion.

1. Overlaps between legal regimes

Data may be subject to several legal regimes at the same time, which leads to a **fragmentation of legal regimes.**[140] In particular, there may be overlaps among the subjective absolute rights (e.g. privacy and intellectual property rights) (Figure 1) or between a subjective absolute right and the contract (Figure 2).

1.1 Overlaps between privacy and intellectual property rights

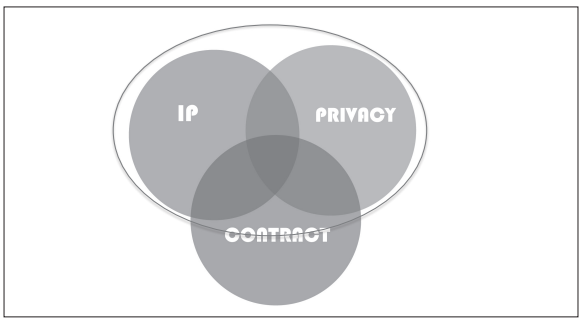

Figure 1: overlaps between privacy and intellectual property rights

In case of overlap between privacy and intellectual property, there is a need to know which legal regime applies. It is thus necessary to **disentangle the data** and apply the legal regime to each part of the data or dataset taken separately. For instance, among an email drafted by an employee, it is necessary to distinguish between the parts of the email containing copyrighted works, texts and original concepts subject to copyright, and parts of the email containing personal information subject to privacy laws.

If disentanglement is impossible (e.g. in case of mixed databases, where elements are inextricably linked), there shall be a cumulation of rights, i.e. multiple and often competing layers of rights covering the same subject matter.[141] This can lead to regime clashes, which may undermine and frustrate the balance of the legal ecosystem, so that there is a need to find solutions.[142]

One solution could be to submit the **prevalence** of one regime over another, for instance by considering that one right is hierarchically superior to another, by applying the right with the most proximity to the case, or by applying exclusively the legal regime

[140] See *Benhamou/Tran* (note 3), 579, referring to cases of termination of the data provider or death of the data user, and showing that solutions may vary between all applicable regimes, ranging from copyright, contract law, personality right, to enforcement and bankruptcy law respectively succession law. See also *Hilty* (note 13), 93, indicating, in relation to the question of data ownership, that the nature of data decisively impacts the applicable legal regime and thus the question of "ownership".

[141] See Communication from the Commission to the European Parliament and the Council, Guidance on the Regulation on a framework for the free flow of non-personal data in the European Union, COM(2019) 250 final, Brussels, 29.5.2019 (Guidance on the FFD), 9, defining mixed datasets as situations when disentanglement, data extraction may *(i)* not be technically feasible and *(ii)* result in depleting the dataset's value, and admitting that these mixed datasets are not only common, but often the most valuable datasets relevant in the data economy, especially in the contexts of big data analytics.

[142] For a discussion on overlapping rights and their effect on the IP ecosystem, see *Estelle Derclaye/Matthias Leistner,* Intellectual Property Overlaps: A European Perspective, Oxford 2011; *Mark Lemley,* Dealing with Overlapping Copyrights on the Internet, 22 U. Dayton L. rev. 547 (1997); *Anette Kur,* Cumulation of IP Rights Pertaining to Product Shapes: An "Illegitimate Off-spring" of IP Law?, in: Ghidini/Genovesi (ed.), Intellectual Property and Market Power: ATRIP papers 2006–2007, Buenos Aires 2008, 613 and 614; *Vina R. Moffat,* Mutant Copyrights and Backdoor Patents: The Problem of Overlapping Intellectual Property Protection, Protection Berkeley Technology LJ (2004), Vol. 19, 1473–1532, 1473 and 1496; *Antoon Quaedvlieg,* Overlaps/Relationships between copyright and other intellectual property rights, in: Derclaye (ed.), Research handbook on the future of EU copyright, Cheltenham 2009, 480–516.

of the prevailing part within the data or dataset.[143] This prevalence rule does not seem to be a viable option, in particular because it would violate international law when the rights belong to two different persons.[144] In situations leading to the irrelevance of data protection rights, it would also be contrary to the fundamental right status of personal data protection.[145] That is why, in case of mixed dataset, the EU legal framework subjects the entire dataset to the GDPR regardless of how much of personal data are included in mixed datasets, so that one can speak of a "data protection by default" rule. This echoes the contaminating effect of open licenses that subject proprietary data to open licence terms.[146] Consequently, a large number of mixed datasets will be subject to the GDPR.[147] Data not originally intended for GDPR application will fall within its scope, so that the GDPR will apply beyond its subject matter, with all the consequences this would entail in terms of compliance. Given the significance and unwaivable nature of data protection, this option should be in our view preferred.

A third remaining solution would be to admit that no legal regime prevails over the other and that there is **coexistence of legal regimes.** Each legal regime has its own scope of protection and applies autonomously. Such coexistence remains open for further research and discussion, but the following typical example can be given: the right of access provided by certain privacy laws (e.g. articles 8 LPD; 23 P-LPD;15 GDPR) which may conflict with trade secrets or intellectual property and in particular the copyright protecting the software.[148] Therefore, the right of access shall apply in the sense that it may lead to explanations about the general logic of a decision-making process with no access to the entire object and source-code of the software.[149]

The legal regime applicable to a certain type of data may also change during the data's lifecycle due to its **transformative nature.** For instance, raw data shall be treated as personal data when aggregated in a dataset that allows the identification of individuals (i.e. de-anonymization of raw data via aggregation) or as a copyrighted work when the aggregated dataset has a sufficient level of originality (e.g. as a compilation or as a software). Similarly, personal data can be transformed into business data when shared along with service contracts, which shows that the legal silos separating data protection and trade secrets laws, for instance, should be questioned.[150] Therefore, some scholars suggest that policy makers and courts

143 Such a solution can be found in intellectual property law, see in particular article 9a al. 4 Swiss Patent Act, providing that, in case of a complex multifaceted product protected by a patent and other IPR at the same which provide different exhaustion regimes (national vs international exhaustion), patent right only applies if the functionality of the product is of primary importance. It can be also found in article 642 Swiss Civil Code, according to which the "accessory follows the principal" when constituent parts cannot be detached from the main part. See *Benhamou/Tran* (note 3), 574, giving the example of a Facebook or LinkedIn status, containing both elements protected by copyright and by data protection, which is indivisible, would be subject to data protection (exclusively) if the Facebook user uses the service mainly to provide his personal data (e.g. personal information to his network of friends), while he would be subject to copyright (exclusively) if the Facebook user conceived his account mainly to transmit his copyrighted works (e.g. photographs, texts, drawings).

144 *Derclaye/Leistner* (note 142), 9.

145 Article 8 of the EU Charter of Fundamental Rights; Case C-131/12, *Google Spain SL and Google Inc. v Agencia Española de Protección de Datos (AEPD) and Mario Costeja González,* [2014] ECLI:EU:C:2014:317, para. 38; Case C-362/14, *Maximillian Schrems v Data Protection Commissioner,* [2015] ECLI:EU:C:2015:650, para. 39.

146 See *Benhamou/Tran* (note 3), 579, giving, as example of contamining effect, article 5 of GNU GPL 3.0: all programs based on GNU GPL licence are subject to the terms of this license, unless *"identifiable sections of that work are not derived from the Program, and can be reasonably considered independent and separate works in themselves"* (article 2 GNU GPLv2).

147 See Guidance on the FFD (note 141), 9 *("if the non-personal data part and the personal data parts are 'inextricably linked', the data protection rights and obligations stemming from the General Data Protection Regulation fully apply to the whole mixed dataset, also when personal data represent only a small part of the dataset")*.

148 See Recital 63 GDPR: *"That right [of access] should not adversely affect the rights or freedoms of others, including trade secrets or intellectual property and in particular the copyright protecting the software". See also article 2 (2) FFD ("In the case of a dataset composed of both personal and non-personal data, this Regulation applies to the non-personal data part of the dataset")* and the Guidance on the FFD (note 141), 9 (*"the Free Flow of Non-Personal Data Regulation applies to the non-personal data part of the dataset; the General Data Protection Regulation's free flow provision applies to the personal data part of the dataset")*.

149 See Recital 63 GDPR *("the result of those considerations should not be a refusal to provide all information to the data subject")*.

150 *Strowel* (note 2), 23.

abandon the piecemeal approach to separating all legal regimes in silos, and suggest a holistic approach based on data governance and/or to reason in terms of trade and investment or property rights and appropriation instead.[151]

1.2 Overlaps between a property right and a contractual right

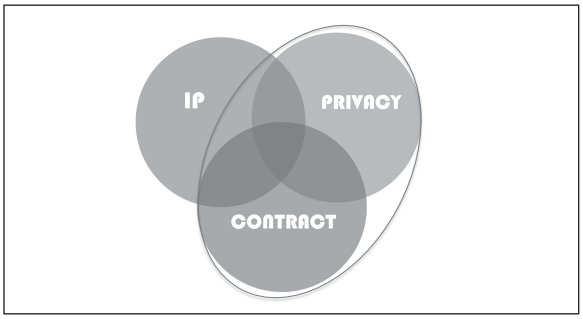

Figure 2: overlaps between an absolute right and contract

In case of overlap between privacy and contract, there may be conflict between unilateral **withdrawal rights** granted to the data subjects under certain privacy laws with the **contractual obligations** of the parties (e.g. long-term license agreement over personal data). To resolve this conflict, the solutions below may be considered.

Organizations subject to a DSA with the data subjects (e.g. the research institution using the data collected directly from the user) may rely on other lawful grounds (instead of consent), recalling that the organizations must decide from the beginning the lawful ground and may **not switch from consent to another lawful basis.** In the absence of such lawful grounds, they shall stop using the data to comply with the unwaivability of personal data protection. Certain fees could potentially be charged when the withdrawal right conflicts with the contractual obligations of the parties.[152]

The withdrawal right may not only affect the licensee's interests (e.g. a service provider that provides services in exchange of personal data) but could potentially **affect further sublicensees** who may process the same personal data, creating a dependency situation of sublicensees regarding the main contract between the licensor and the licensee. Of course, sublicensees could claim a breach of contract against the licensee when the withdrawal right is triggered. This does not alter the fact that no more personal data can be processed by both the licensee and its sublicensees. To remedy this dependency situation, German and Swiss courts tend to grant to the sublicensees, at least for copyrighted works, an **independent right to continue using the copyrighted work.**[153] If this case law relates to copyright, there are good grounds to apply a similar reasoning to personal data or other IPR, at least when data are tradable similarly as copyrighted works.

2. Promoting alternative dispute resolution mechanisms

Because of the limits of other legal protections, as explained above, contractual arrangements seem to be the prevailing system for Big Data activities: they are sufficiently flexible to regulate data flows between all stakeholders and property-like rights.[154] There are, however, several downsides which in turn lead to an **increase of data disputes**.

Typical data disputes may arise around **data ownership**.[155] Data may belong to several data pro-

151 *Strowel* (note 2), 23; *Yu* (note 50), 927.

152 The withdrawal right does not exclude, in our view, the right of the other contracting parties to impose some fees and expenses. See Working Party 29, Guidelines on consent (note 76), 21 *("Data subject should be able to withdraw his/her consent without detriment. This means, inter alia, that a controller must make withdrawal of consent possible free of charge or without lowering service levels")*.

153 *Jacques de Werra,* Perspective "Inside-Out": Défis du droit d'auteur dans un monde connecté, sic! (2014), 194–211, 206, quoting the *Reifen Progressiv* case (BGH, GRUR 2009, 946); the Federal Court (BGH) granted to the sublicensee *"the right to continue using the intangible good which form part of the sublicensee (softwares and musical works) despite the termination of the main license between the licensor and the licensee (licensor of the sublicense)".* Swiss Supreme Court (TF), sic! (2013), 603, consid. 5.5; however, in this decision, instead of granting a real independent right to use, the court refused the termination of the main licence by the licensor in order to take into account the interest of sublicensees. See also, *Benhamou/Tran* (note 3), 579.

154 See *de Werra* (note 153), 207, stating that contracts play a fundamental role in regulating these interactions; *Benhamou/Tran* (note 3), 579.

155 For other examples of data disputes, see *Jacques de Werra,* From Intellectual Property (Data-Related) Disputes to Data Disputes: Towards the Creation of a Global Dispute

ducers: under copyright laws, the output generally receives its own copyright protection, distinct from that granted to the underlying input. This may create the complex situation of a rightholders' chain or joint ownership in the output. When this copyright protection is combined with the *sui generis* database right, ownership may be more complex, as each piecemeal regime has its own definition of the rightholder. The copyright holder is usually the person who created the work, while the *sui generis* data producer is the person who bears the risk of investment.[156] Even if the organizations try to contractualize the question of data ownership, this will be practically impossible with full legal certainty with emerging technologies (Big Data, Internet of Things and AI-based systems) because of the multitude of data sources and actors who may want to claim ownership in the data concerned. It is extremely difficult to define the concept of "data" and "data ownership" clearly, since no legal definitions exist, which leads to a diversity of possible interpretations in different agreements without any harmonized view. As a result, it is frequent that stakeholders try to **define** the notion of "data ownership" as broadly as possible, and to capture the permitted **actions** in a highly restrictive manner.[157] This can block the whole Big Data analytics chain and make the Big Data analytics unworkable, as users would be severely restricted.

The multitude of actors and of data sources may also lead to an intricate chain of DSAs with the risk of **interdependencies** between all stakeholders which remain uncertain until case law is clear.[158] Contractual agreements are also **not enforceable vis-à-vis third parties,** they only generate rights and obligations for the parties to such agreements. In practice, this means that there is no recourse available against third parties that obtain unjustified access to and/or misuse the data. Finally, the frequent **transnational character** of Big Data and the imbalanced bargaining power may lead to conflicting contract interpretations (e.g. when a service provider imposes terms and conditions that are unclear to the user of the service, which may be in turn a consumer).

Alternative dispute resolution mechanisms (ADR) may be an appropriate tool to solve data disputes which takes into account the entire Big Data ecosystem and to grant a greater access to justice.[159] Ongoing initiatives to develop governance bodies in this respect are noteworthy, but **many** of them are implemented by the **industry** with no state intervention, which could compromise transparency and enforcement mechanisms. It could be therefore interesting to implement an efficient and transparent governance mechanism, for instance via the creation of an independent *Digital Ombudsman* that would help to solve data disputes, who/which could be appointed by and assist the relevant Data Protection authority in relation to data disputes.[160]

Resolution Ecosystem for Data Disputes in the Digital Era, in: Zeiler/Zojer (ed.), Resolving IP Disputes: A Selection of Contemporary Issues, Graz 2018, 87–109, 92 ss.

156 Database Directive, Recital 41; *Maier/Sibble* (note 23), 24. Although there is still no clear definition available in EU law or in case law, it can be said that the database producer will be the entity who: (1) takes the initiative in obtaining, verifying or presenting the contents of the database; and (2) assumes the risk of investing in that obtaining, verifying or presenting, to the exclusion of subcontractors.

157 E.g. prohibiting downstream uses or other actions such as reverse engineering, merging, enriching, decompiling, structuring, cleansing, altering, displaying, reproducing, visualising, communicating, loading, running, transmitting, storing, observing, studying, testing.

158 See III.1.2 with references to case law which tends to recognize sublicensees' interests when dealing with the main licence.

159 See *de Werra* (note 155), 87. For concrete proposals on how to build a robust ADR framework on internet-related disputes, see <*https://geneva-internet-disputes.ch*> [accessed 2 October 2020]; on the digitization of museum collections, see <*www.digitizationpolicies.com*> [accessed 2 October 2020].

160 For other approaches, see *Jacques de Werra,* Using Arbitration and ADR for Disputes about Personal and Non-Personal Data: What Lessons from Recent Developments in Europe?, American Review of International Arbitration 2/2019, 195–217, arguing that arbitration or other ADR would be an interesting global approach which takes into account the entire IP ecosystem and which shall not (too) narrowly focus on the bilateral relationships between licensor and licensee. See also, *Chris Reed,* Data trusts: legal and governance considerations, Pinsent Masons, April 2019, considering that data trust structures might be the adequate structure that can include alternative dispute resolution mechanisms and discussing the legal structures to achieve it, from corporate to contractual structures.

3. Promoting co-regulation to develop digital ethical standards

As the current legal framework tends to **increasingly rely on self-regulation** (i.e. norms enacted by private bodies, such as professional norms, certification or codes of ethics),[161] there is a need to increase **the participation of the civil society** and the level of democratic control in the norm-setting process.[162] This participation is necessary, even more so to avoid self-regulation of being a marketing tool, sometimes called "ethics washing".[163]

Regulating self-regulation ("co-regulation") might be the appropriate response to democratic control[164] and has recently made good progress,[165] albeit in a fragmented way, with several legal uncertainties. For instance, in the norm-setting process, an overarching question is how to ensure the participation of all stakeholders and transparency. For instance, with the EU Code of conduct on countering illegal speech online, who has been consulted and who may appeal against these norms? And once the code is out, another question is what is the **liability** for the norm-producers and endorsers? Should a company complying with a code of conduct be exempted from any liability?[166]

Policy makers could contemplate a multilayered approach to these questions:

[161] Advantages of self-regulation put forth are mainly flexibility, cost saving and prompt compliance by the stakeholders involved in the process. Consequently, there is a proliferation of self-regulation, leading to a normative transfer from public bodies to businesses, and to a normative content of internet composed of 40% of private rules, see *Boris Barraud,* Le renouvellement des sources du droit: Illustrations en droit de la communication par internet, Aix-Marseille 2017.

[162] See, the UN High-level Panel report, recommendation 3C: *"Audits and certification schemes should monitor compliance of AI systems with engineering and ethical standards, which should be developed using multi-stakeholder and multilateral approaches. Life and death decisions should not be delegated to machines."* See *Alexandre Flückiger,* (Re)faire la loi, Traité de légistique à l'ère du droit souple, Bern 2019, 325, who speaks about the "input legitimacy" when the norm-setting process complies with certain procedural guarantees and the participation of all stakeholders.

[163] *Urs Gasser/Carolyn Schmitt,* Normative Modes: Professional Ethics, in: Dubber/Pasquale/Das (ed.), The Oxford Handbook of Ethics of AI, Oxford 2019 *("the cynic can argue that these are merely marketing ploys or modes of influencing public perception, or a form of 'ethics washing'")*.

[164] For critics about self-regulation, see Report of the UN Secretary-General's High-level Panel on Digital Cooperation, The age of digital interdependence, June 2019, available at <*https://digitalcooperation.org/*> [accessed 2 October 2020] (UN Panel), 14: *"Critics counter that an overly hands-off approach has led to a concentration of market power in large firms and abuses of privacy that have sparked public and government concern."* For a call for regulating self-regulation, UN Panel, 7: *"The dynamic digital world urgently needs improved digital cooperation"; Hilty* (note 13); *Christophe Busch,* Self-Regulation and Regulatory Intermediation in the Platform Economy, in: Gamito/Micklitz (ed.), The Role of the EU in Transnational Legal Ordering: Standards, Contracts and Codes, Cheltenham 2020, 115–134. See also for instance the Swiss expert group report on data processing and security, 17 August 2018, 58 ff., suggesting to rely on standards and certifications (4.4.4 Standards und Zertifizierungen von Produkten: *"Empfehlung: Der Bund prüft in Abstimmung mit der Entwicklung im Ausland, ob und in welchen Bereichen Standards und Zertifizierungen zu einer Voraussetzung für den Marktzugang von IKT-Komponenten erklärt werden müssen, und welche gesetzlichen Rahmenbedingungen dafür nötig sind"*).

[165] See for instance the GDPR (2018) with codes and certification schemes, the Code of conduct on countering illegal hate speech online (2016) issued following a cooperation between the European Commission and IT companies (such as with Facebook, Microsoft, Twitter and YouTube), the 2019 Regulation on the free flow of non-personal data, consisting in self-regulation facilitated by the Commission so as to define best practices for switching provider and for enhancing contractual transparency.

[166] The certified companies could be held liable even when complying with the code, as codes may have different normative effects (from mere inspiration, to compliance-presumption or compliance-fiction). Consequently, the liability of auditors and SDI may result from the certification itself *(Vertrauenhaftung; responsabilité fondée sur la confiance,* ATF 128 III 324 *and* ATF 130 III 345*),* which depends on the content of the code and the level of expectations raised by the code.

- **Self-regulation layer:** civil society, academia and governmental representatives should participate in **drafting standards,** for instance **in cooperation with standardization organizations,** such as the International Organization for Standardization (ISO), which tries to give consumers an active role in developing standards through consumer representation on the boards of international NGOs and membership in national technical committees.[167]
- **Co-regulation layer:** these standards should be then, if possible, implemented into **certifications** submitted to the relevant authority for approval, so as to increase the democratic control and/or legal certainty regarding liability (with compliance-fiction or presumption). Finally, such submission / approval should be imposed, at least in **risky areas** (such as the use of AI in justice, cyberhealth and facial recognition).[168]

IV. Conclusion

There is **no one type of data, but several types of data,** which are subject to different legal regimes, ownership rights or other concepts and which lead to different data spaces[169]. Each legal regime defines its own scope of protection. However, if the scope of each property-like protection seems clear *prima facie,* the analysis shows that the **eligibility for protection is delicate** in many respects. As an outcome, **contractual arrangements** remain the **prevailing system** in the Big Data context. Contractual arrangements have in turn their shortcomings.

Consequently, it is important to develop a clear and **coherent body of law,** in particular to define the boundaries of each regime according to the nature of the relevant data, and to abandon a piecemeal approach and to suggest a holistic approach based on all legal regimes and interests at stake. A **three-step approach** may be a practical approach that guides organizations before embarking in Big Data activities (above II). **Policy solutions** may be also contemplated, including for the question of overlaps between all legal regimes, the promotion of alternative dispute resolution mechanisms which may arbitrate data disputes and of the use of co-regulation as a tool to regulate the technologies while taking into account the democratic control and the participation of all stakeholders (above III).

[167] ISO brings together members which can choose whether they want to be part of a particular technical committee (TC) (there are 250 technical committees that represent every sector) and their level of involvement (Observer-members observing the standards that being developed and offering comments and Participant-members participating actively by voting on the standard at various stages). Technically, to provide understandable descriptors for its addressee, we suggest to shape the labels modeled on the Creative Commons "three-layer design" that provides *(i)* a consumer-oriented human readable part (the CC Common Deed), *(ii)* an auditable specification for experts (the CC Legal Code) and *(iii)* a technological (machine-readable) (the CC Digital Code).

[168] See, for instance, the Council of Europe's project, which is currently exploring the feasibility of a certification mechanism for AI products used in judicial systems (Project for the certification of artificial intelligence products, following the adoption of the 2018 European Ethical Charter on the use of artificial intelligence in judicial systems, available at <*https://www.coe.int/en/web/cepej/home/*> [accessed 25 September 2019].

[169] See the European strategy for data (note 65) and legislative framework for the governance of common European data spaces (note 95).