THE RIGHT TO PROCESS DATA FOR MACHINE LEARNING PURPOSES IN THE EU

Mauritz Kop*

TABLE OF CONTENTS

I.   INTRODUCTION

Data is a primary resource. It is a means of existence for AI and a necessity of life for humanity. In its data strategy, the European Union (EU) recognizes the crucial importance of data for machine learning purposes. Machine learning is an exigency for a vibrant artificial intelligence (AI) ecosystem. In AI, machine learning algorithms and data work together in unison. Data wants to be exchanged[1], to be accessed, shared and processed.

As a third hand, AI can assist us in finding solutions to the big challenges we face, such as climate breakdown, withering natural resources, interplanetary travel, diversity, equality and inclusivity.[2] Thus, AI can help to preserve our core democratic values, prosperity and well-being as we shape the future of our society.

---

* Mauritz Kop *is Stanford Law School TTLF Fellow at Stanford University and is Managing Partner at AIRecht, Amsterdam, The Netherlands. Correspondence: advies@airecht.nl*
[1] Peter K. Yu, *The Political Economy of Data Protection*, 84 CHI.-KENT L. REV. 777, 799 (2010).
[2] Stephen Hawking, *Brief Answers to the Big Questions* (John Murray 2018).

Data sharing involves a certain amount of data openness. Optimal levels of data openness depend on many factors. According to the OECD, data openness is a continuum, where the benefits of enhancing access, sharing and re-use need to be balanced with the risks.[3] In addition, data poses myriad legal challenges.[4] A well-functioning legal-technical system in Europe that takes into account the interests of all parties involved, which offers legal certainty and a favorable investment climate, and that above all has been created with the data-driven economy in mind, does not yet exist. We are dealing here with a complex, multidimensional problem that hinders exponential innovation. Solving this problem requires legislative intervention.

A key objective of the European Strategy for Data is the creation of a legislative framework for data governance: The Data Governance Act (expected to be adopted mid-2021).[5] The goal of this article is, *de lege ferenda*, to present thoughts about ways to make access to and equitable sharing of machine learning training, testing and validation datasets work within the EU *acquis communautaire*.[6] In this context, I would like to give an interdisciplinary *impetus* for an adequate legal framework for the use of input data for machine learning purposes.[7] The principal role is played by a novel right to process data.

---

[3] OECD Report, *Enhancing Access to and Sharing of Data: Reconciling Risks and Benefits for Data Re-use across Societies*, OECD PUBLISHING, PARIS (2019), https://doi.org/10.1787/276aaca8-en.

[4] Mauritz Kop, *Machine Learning & EU Data Sharing Practices*, 1 TTLF NEWSLETTER ON TRANSATLANTIC ANTITRUST AND IPR DEVELOPMENTS STANFORD-VIENNA TRANSATLANTIC TECHNOLOGY LAW FORUM, STANFORD UNIVERSITY 2020, https://www-cdn.law.stanford.edu/wp-content/uploads/2015/04/2020-1.pdf.

[5] On November 25, 2020, the EC published its Proposal for a Regulation on European data governance (Data Governance Act), the first of a set of measures announced in the 2020 European strategy for data.

[6] For further reading about the need to ensure access to data, and the relevance of data sharing *see* Josef Drexl et al., *Data Ownership and Access to Data*, MAX PLANCK INSTITUTE FOR INNOVATION AND COMPETITION 1, 2 (2016) http://www.ip.mpg.de/en/link/positionpaperdata-2016-08-16.html; Drexl, Josef, *Designing Competitive Markets for Industrial Data - Between Propertisation and Access* (October 31, 2016), Max Planck Institute for Innovation & Competition Research Paper No. 16-13; Josef Drexl, *Legal Challenges of the Changing Role of Personal and Non-Personal Data in the Data Economy* (October 31, 2018), in A. De Franceschi, R. Schulze (eds.), Digital Revolution: Data Protection, Smart Products, Blockchain Technology and Bitcoins Challenges for Law in Practice, MÜNCHEN, BECK, 2019, 19-41; MAX PLANCK INSTITUTE FOR INNOVATION & COMPETITION RESEARCH PAPER NO. 18-23.; Jurcys, Donewald, Globocnik & Lampinen, Note, *My Data, My Terms: A Proposal for Personal Data Use Licenses*, HARV. J.L & TECH. DIG. (2020), https://jolt.law.harvard.edu/digest/my-data-my-terms and Bonnie Kaplan, *Seeing Transparency Through: Healthcare Software, Data Privacy, and Regulation* (2018), https://balkin.blogspot.com/2018/10/seeing-transparency-through-healthcare.html.

[7] Since comparison of similarities and differences between legal systems is a rewarding informer of legal advancements and legal reform, I will provide both EU and U.S. perspectives.

II. MACHINE LEARNING & THE EUROPEAN DATA GOVERNANCE ACT

Machine learning can be many things. A machine learning system us not programmed, but trained.[8]

Currently, most machine learning techniques require training datasets (*corpora*) to achieve valuable results with regard to predictive analysis and optimization. The data needs to be processed to enable model learning. Training refers to the process of determining the ideal parameters comprising an AI-model.[9] Once the algorithmic model has been trained it can be applied to new datasets to produce new correct answers to prediction (inference) and optimization problems. High quality training datasets are required to take full advantage of AI's capabilities: the quality of the output is contingent with the quality of the input.[10]

Data are not a monolithic entity, but heterogenous digital assets.[11] This makes precise classification of data a challenge. Important aspects for a useful and valid data taxonomy are the personal nature and identifiability of data[12], the domain of the data[13] and the manner data originates.[14] Machine learning training datasets also come in many forms.[15] From a legal point of view these datasets

---

[8] We can distinguish 5 tribes of machine learning who each have a different philosophy on how to approach AI related problems: Connectionists (deep learning), Analogizers (reinforcement learning), Evolutionaries (evolutionary algorithms), Symbolists (inverse deduction), Bayesians (quantum computing). *See* What Is Machine Learning (The Dawn of Artificial Intelligence), https://youtu.be/E2IAhfD1WYM, and Domingos, P. *The master algorithm: How the quest for the ultimate learning machine will remake our world*. (BASIC BOOKS, 2015).

[9] *See* Google, Training and Test Sets: Splitting Data, https://developers.google.com/machine-learning/crash-course/training-and-test-sets/splitting-data.

[10] This principle is referred to as "garbage in garbage out" (GIGO). *See, e.g.*, Jeff Gorke, AI And Machine Learning In Healthcare: Garbage In, Garbage Out, Forbes (2020). https://www.forbes.com/sites/jeffgorke/2020/06/18/ai-and-machine-learning-in-healthcare-garbage-in-garbage-out/. To prepare data for machine learning, it needs to be pre-processed, cleaned, transformed and normalized, *see* Ciklum, Garbage In, Garbage Out: How to Prepare Your Data Set for Machine Learning, (2019), https://www.ciklum.com/blog/garbage-in-garbage-out-how-to-prepare-your-data-set-for-machine-learning/ Model retraining is used to produce more accurate results. That process also requires data.

[11] OECD Report, *supra* note 3.

[12] *Id.* Such as healthcare or financial data.

[13] *Id.* Private, public or personal domain.

[14] *Id.* Such as satellite data or machine generated data.

[15] *Corpora* can be centralized (in supervised machine learning) or decentralized (in federated learning) datasets, structured or unstructured (i.e. raw and fragmented), hand labelled, machine labelled, unlabelled, or a combination of these types of data. During the learning process training, testing and validation subsets are used. The location of the data can be in repositories like a server, in the cloud or on edge devices.

can contain private domain, public domain, personal domain data, or a mixture.[16] Besides that, data can be open or enclosed. It can be industrial, consumer, government or scientific data. It can be commercial or non-commercial. The diverse and overlapping typology of data complicates legal classification and qualification. It also confuses informed policy debates about data-governance frameworks.[17] Legal terminology with regard to AI and data hasn't matured yet and should be harmonized.

The European Commission's (EC) new Digital Strategy is titled "Shaping Europe's Digital Future."[18] An essential part of this five-year roadmap is the European Data Strategy agenda.[19] The EC has the ambition to regain data sovereignty, and Europe must become an international data hub. Further, the EC aims to set the worldwide standard for horizontal and vertical data sharing through the Towards Common Data Spaces initiative. As the General Data Protection Regulation (GDPR)[20] became the international standard for data protection. The EU has an enormous amount of high quality industrial data that is eagerly waiting to be exchanged.

More specifically, an essential policy objective is to implement permission for horizontal B2B, G2B, B2G and C2B data sharing, and between European Member States. The data can then be annotated, analyzed and processed with the purpose of teaching an AI model how to learn. In addition, each vertical, or economic sector e.g. Logistics, Energy, Creative Industry and Life Sciences & Health, has specific challenges in the field of sectoral and cross-sectoral sharing and re-use of machine learning *corpora*. This set of policy goals requires a modern, tailor-made legal framework in the form of the European Data Governance Act, which is expected to be adopted in the course of 2021.

In my opinion it is crucial that the Data Governance Act contains an explicit right to data processing.

---

[16] Such as domain specific industrial data, government satellite data, anonymized edge device data or sensitive, hi-risk data.

[17] OECD Report, *supra* note 3. For example commercial versus non-commercial data.

[18] On February 19, 2020, the EC published 3 important policy documents: (1) Communication "Shaping Europe's Digital Future", (2) Communication "A European Strategy for Data" and (3) White Paper "On Artificial Intelligence – A European Approach to Excellence and Trust".

[19] Important elements of this agenda are the construction of a secure cloud in the edge including Rulebook and Marketplace and the launch of federated cloud platform GAIA-X.

[20] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

III. LEGAL PROBLEMS REGARDING DATA ACCESS, SHARING AND RE-USE

Elsewhere I wrote that issues concerning intellectual property and ownership of data, data protection and privacy are obstacles to the ability to naturally share high quality data between citizens, companies, research institutions and the government.[21]

The first legal hurdle that must be overcome when sharing data is copyright. Datasets consisting of copyrighted works such as books, photos, film fragments and music have to be cleared before they can be reproduced by our intelligence machines into a usable model.

Second, third party (*sui generis*) database rights can rest on (parts of) the training dataset. In Europe, augmented machine learning training datasets are still protected by a (*sui generis*) database right.[22] In case the training data contains copyrighted works that are not yet in the public domain, or information protected by database rights -and no TDM exception applies- prior permission to use and process must be obtained from the rights holders (for scientific, commercial and non-commercial training purposes) in the form of a written user-license.[23]

Third, companies will make every effort to protect their investments in training an AI model and to generate revenue themselves with their AI applications.[24] They will do everything they can to keep the datasets a trade secret or to protect the entire database whether hand-coded or machine-generated- through contracts or technological measures.[25]

The fourth problem is the fear of the GDPR. While the GDPR protects the personal data of European citizens, in some cases it

---

[21] Kop, *supra* note 4.

[22] For *sui generis* regimes in the context of AI and data, see: Péter:Mezei, *From Leonardo to the Next Rembrandt – The Need for AI-Pessimism in the Age of Algorithms* (May 4, 2020), https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3592187.

[23] Kop, *supra* note 4. For further reading on data licenses, see: Jurcys et al., *supra* note 6.

[24] For further reading on cross-border investments and its relation to intellectual property rights, see: Siegfried Fina and Gabriel Lentner, *The New Generation of International Investment Agreements Adopted by the EU and Its Implications for the Protection of Intellectual Property Rights* (February 1, 2017). JOURNAL OF WORLD INVESTMENT & TRADE (2017) 271 – 305. and Geiger, Christophe, Intellectual Property and Investment Law: An Introduction (November 14, 2019). Christophe Geiger (ed.), *Research Handbook on Intellectual Property and Investment Law*, CHELTENHAM (UK)/NORTHAMPTON, MA (USA), EDWARD ELGAR PUBLISHING, 2020, CENTRE FOR INTERNATIONAL INTELLECTUAL PROPERTY STUDIES (CEIPI) RESEARCH PAPER NO. 2019-10.

[25] *Id. See also* Daniel J. Gervais, *Exploring the Interfaces Between Big Data and Intellectual Property Law* 10, J. INTELL.L PROP., INFO.TECH.Y & ELEC.COM. L. (JIPITEC) 22; VANDERBILT LAW RESEARCH PAPER NO. 19-36 (March 26, 2019). For further reading about trade secrets and AI, see: Camilla Alexandra Hrdy and Mark A. Lemley, *Abandoning Trade Secrets* (February 7, 2020), 73 STAN. L. REV. (Forthcoming), and Sonia Katyal, *The Paradox of Source Code Secrecy* (June 25, 2019), CORNELL L. REV. (Forthcoming),.

demonstrably hinders the rapid roll-out of AI and data start-ups[26] and scale-ups within the European internal market.[27] In other words, there is a risk of tension between the GDPR, intellectual property, policy aims such as sharing training data and open innovation.[28] Here, lack of knowledge and legal uncertainty causes risk-averse behavior. In any case, it does not produce spectacular data-driven AI applications and European unicorns that can compete with digital champions from Silicon Valley and Shenzhen.[29]

The fifth roadblock is legal uncertainty about legal ownership of data. *De facto*, economic ownership of data exists.[30] Legal ownership, however, does not exist, in the sense of property law. Data, of course, is a digital asset and does have characteristics of property. Legal ownership (or real property) is not the same as intellectual property.[31] Moreover, data ownership rules vary per country, even within the EU. German law excludes intangibles, where French law allows full rights with regard to intangibles.[32] Digital data has not yet been recognized as legal object that can be possessed.[33] In other words, digital assets have not yet been included in the *numerus clausus* of subject matter eligible for ownership.[34] Thus, there can be no formal property rights in data. Due to confusion about *de facto* economic and legal ownership, stakeholders -understandably- have no clue who the rightful owner of data is and what a data holder can and cannot do with it. Let alone that ownership interests are clear in more

---

[26] To reinforce the single market for digital services, modernize the legal framework for digital services and help provide smaller businesses with the legal clarity and level playing field they need, the EC has put forward a new Digital Services Act and a Digital Market Act on December 15, 2020, *see*, The Digital Services Act package and Europe fit for the Digital Age: Commission proposes new rules for digital platforms.

[27] See also: OECD Report, *supra* note 3.

[28] Timo Minssen, Rajam Neethu & Marcel Bogers, *Clinical Trial Data Transparency and GDPR Compliance: Implications for Data Sharing and Open Innovation,* In: Katerina Sideri & Graham Dutfield (eds.), OPENNESS, INTELLECTUAL PROPERTY AND SCIENCE POLICY IN THE AGE OF DATA DRIVEN MEDICINE, SPECIAL ISSUE OF SCIENCE AND PUBLIC POLICY (2019).

[29] On November 25, 2020 the EC adopted its Action Plan for Intellectual Property, to increase Europe's ability to develop next generation technologies and reflect advances in data and AI. *See also* the IP Action Plan's Factsheet.

[30] Drexl *supra* note 6, Jurcys et al., *supra* note 6. The Behemoth social media platforms can be considered *de facto* consumer data owners in the economic sense of the word, since they possess/control valuable user data. For *de facto* ownership regimes in the U.S., see: Jorge Contreras, Property Rules and Liability Rules for Genetic Data (2015), https://balkin.blogspot.com/2015/03/property-rules-and-liability-rules-for.html.

[31] *See also* Mark A. Lemley, Property, *Intellectual Property, and Free Riding*. 83 TEX. L. REV 1031 (2015).

[32] van Erp, S. *Ownership of data: the numerus clausus of legal objects*. 6 BRIGHAM-KANNER PROP. RTS. CONF. J., 235-257 (2017), https://scholarship.law.wm.edu/propertyjournal/6/.

[33] *Id*. The absence of legal ownership rights on data as intangible goods is should be maintained.

[34] *Id*.

complex cases where a dataset has been created by different companies from different locations over a certain period of time. Or when a dataset consists of a mix of personal information, government data and industrial data. There is a strong and urgent need for legal certainty about these matters.[35]

IV. LEGAL SOLUTIONS

For context and clarity, I will briefly repeat several previously proposed solutions to the identified legal problems surrounding data.[36]

To make machine learning possible, we need exceptions and limitations to copyrights. First, the implementation of a broad, mandatory text and data mining (TDM) exception in the Copyright Directive[37] that covers all types of data in Europe.[38] Second, the creation of an online one-stop-shop clearinghouse with mandatory or statutory licensing for machine learning training datasets alike a pan-European, multi-territorial collective rights agency.[39] Third, the acknowledgement and implementation of a right to machine legibility.[40] A right to fair learning is not possible in the foreseeable future in the EU because the European intellectual property rights system does not allow fair use.[41] In general, copyright law should allow the reproduction of works for non-expressive purposes.[42]

---

[35] This is called a "mixed dataset". For practical guidance on how to process mixed datasets, *see* https://ec.europa.eu/digital-single-market/en/news/practical-guidance-businesses-how-process-mixed-datasets. For a strict conceptual separation of personal data from personal information, *see* Václav Janeček, *Ownership of Personal Data in the Internet of Things* (Dec. 1, 2017). 34 COMPUT.L. & SEC. REV, 1039-105 (2018).

[36] Kop, *supra* note 4.

[37] Directive (EU) 2019/790 of the European Parliament and of the Council of 17 April 2019 on copyright and related rights in the Digital Single Market and amending Directives 96/9/EC and 2001/29/EC (CDSM Directive).

[38] Christophe Geiger, Giancarlo Frosio, & Oleksandr Bulayenko, *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 (Mar. 2, 2018). See also: Sean Flynn, Christophe Geiger & João Quintais et al., *Implementing User Rights for Research in the Field of Artificial Intelligence: A Call for International Action,* EUROPEAN INTELL. PROP. REV. 2020, ISSUE 7 (April 20, 2020).

[39] *Id. See also* Kop, *supra* note 4. Cross-border Extended Collective Licensing (ECL), or Collective Licensing with Extend Effect (CLEE) would also be an option. See: Axhamn, Johan and Guibault, L., *Solving Europeana's Mass-Digitization Issues Through Extended Collective Licensing?* (Dec. 20, 2011). 6 NORDIC INTELL. PROP. L. REV., 509 (2011).

[40] Rossana Ducato & Alain M. Strowel, *Limitations to Text and Data Mining and Consumer Empowerment: Making the Case for a Right to Machine Legibility*, CRIDES WORKING PAPER SERIES (2018).

[41] Mark A. Lemley & Bryan Casey, *Fair Learning* (Jan. 30, 2020). *See also* Hugenholtz, P. Bernt and Senftleben, Martin, *Fair Use in Europe: In Search of Flexibilities* (Nov. 14, 2011).

[42] *Id.* For further reading on unanticipated future benefits of emerging technologies in the longer term, or "spillover effects", and why it is important that copyright regimes do not

Concerning the scope of the EU Database Directive 96/9/EC[43], the EU Commission should revise it to prevent machine generated data, including data generated by connected edge devices from qualifying for *sui generis* database rights protection. Edge computing data should not be handicapped by exclusive rights for this hinders access to and sharing of machine learning training datasets.[44]

The definition of the Trade Secrets Directive[45] should be clarified and narrowed, so that the derived and inferred data cannot be classified as a trade secret.[46] Similarly, non-personal machine generated data ought not be eligible for protection under the Trade Secrets Directive: it should be open, public domain data.[47]

With regard to the GDPR, I would like to share the following three observations. First, there is no dichotomy between privacy and (personal) data protection. These are two sides of the same medal.[48] The protection of (personal) data doesn't necessarily conflict with other policy goals such as safeguarding privacy; they can easily coexist with each other.[49] Moreover, data protection goes hand in

---

stifle innovation, *see* Mark A. Lemley & R. Anthony Reese, *Reducing Digital Copyright Infringement Without Restricting Innovation*, 56 STAN. L. REV. 1345 (2004), https://law.stanford.edu/publications/reducing-digital-copyright-infringement-without-restricting-innovation/.

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

[44] Kop, *supra* note 4. *See also* Julia Johnson, *Database Protection a Reality? How the Professional and Fantasy Sporting World Could Benefit from a Sui Generis Intellectual Property Right*, INTELL. PROP. J. 256 (2015).

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

[46] Kop, *supra* note 4. Note that "individual data" should not qualify as trade secret, due to the requirements of "secrecy" and "commercial value", see: Drexl. Et al., *supra* note 6.

[47] Autonomously AI generated creations and inventions that have no causal relationship with human authorship or inventorship anymore, should be public domain. This machine generated output should be *res publicae ex machina* (public property from the machine). *See* Mauritz Kop, *AI & Intellectual Property: Towards an Articulated Public Domain*, 29 TEX.INTELL. PROP. L. J. (2020). See also: Daniel Gervais, *The Machine As Author* VANDERBILT LAW RESEARCH PAPER NO. 19-35, 21 (2019); Ana Ramalho, *Will robots rule the (artistic) world?* 21 JOURNAL OF INTERNET LAW 22 (2017); Pamela Samuelson, *Allocating Ownership Rights in Computer Generated Works* (1986), 47 U. PITT. L. REV. 1185, 1224, Jane C. Ginsburg, and Luke Ali Budiardjo, *Authors and Machines* (August 5, 2018). COLUMBIA PUBLIC LAW RESEARCH PAPER NO. 14-597; 34 BERKELEY TECH. L. J. (2019)., Reto Hilty, Jörg Hoffmann & Stefan Scheuerer, *Intellectual Property Justification for Artificial Intelligence*, in Artificial Intelligence & Intellectual Property, (J.-A. Lee, K.-C. Liu, R. M. Hilty eds., OXFORD UNIVERSITY PRESS, 2020); MAX PLANCK INSTITUTE FOR INNOVATION & COMPETITION RESEARCH PAPER NO. 20-02. and Mauritz Kop, *Computer generated works: wie of wat is eigenaar?*, VERDERDENKEN, CENTRUM VOOR POSTACADEMISCH JURIDISCH ONDERWIJS (CPO), RADBOUD UNIVERSITY 1 APRIL 2020, https://www.ru.nl/cpo/verderdenken/columns/computer-generated-works-eigenaar/

[48] *See also* Towards Common European Data Spaces - EU digital policy interview with Yvo Volman – Head of Unit, Data Policy and Innovation, DG CONNECT, European Commission// CSBXL20, https://www.youtube.com/watch?v=v642ICQBHh4.

[49] Yu, *supra* note 1, 797.

hand with data responsibility.[50] Second, based on rules set by the GDPR for personal data, and the Free Flow of Data Regulation (FFD regulation)[51] for non-personal data, mixed datasets can move freely within the European Union.[52] The GDPR therefore offers space and flexibility to start sharing personal data and process it. Fear of its complexity is understandable but not necessary. Third, for the avoidance of doubt the GDPR needs to be elucidated by the EC, so that the GDPR cannot be interpreted as allocating partial and/or limited ownership rights of personal data to any legal subject.[53]

V.   NORMATIVE JUSTIFICATIONS FOR OPEN INNOVATION

Let us return to ownership of data and legal certainty about proprietary aspects of data (the fifth problem). As mentioned above, high quality machine learning data input is an important precondition for the birth and evolution of AI. There is a lack of a legal framework for data-driven markets.[54] The digital revolution demands a rethinking of classical property law in the virtual world.[55] For every new disrupting tech that creates a new market and value network[56], policy makers tend to create a new layer of exclusive rights.[57]

We need not to think in terms of exclusive, private property on data, but in terms of rights and freedoms to use, (modalities of) access, process and share data. Data has become a primary resource that should not be enclosed or commodified *per se*, but used for the

---

[50] Jacqueline Deborah Lipton, *Information Property: Rights and Responsibilities*. 56 FLA. L.REV 135, (Jan 2004); Yu, *supra* note 1.

[51] 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).

[52] *See* Kop, *supra* note 4. Andrea Renda calls the free circulation of non-personal data the "Fifth Freedom." *See* Andre Renda, *Single Market 2.0: the European Union as a Platform*, RESEARCH PAPER IN LAW, 02/2020, DEPARTMENT OF EUROPEAN LEGAL STUDIES, COLLEGE OF EUROPE, https://www.coleurope.eu/study/european-legal-studies/research-activities/research-papers-law.

[53] Jurcys et al., *supra* note 6. Compare to: Nestor Duch-Brown, Bertin Martens, Frank Mueller-Langer, *The Economics of Ownership, Access and Trade in Digital Data,* 12 (JOINT RESEARCH CENTER, WORKING PAPER 2017-01, 2017), https://ec.europa.eu/jrc/sites/jrcsh/files/jrc104756.pdf and Josef Drexl et al., *supra* note 6.

[54] *See also* Richter, H., Slowinski, P.R. *The Data Sharing Economy: On the Emergence of New Intermediaries*. IIC 50, 4–29 (2019), https://doi.org/10.1007/s40319-018-00777-7.

[55] Van Erp, *supra* note 32, 256.

[56] Joseph L. Bower, and Clayton M. Christensen, *Disruptive Technologies: Catching the Wave*, HARV. BUSINESS REV. (1995).

[57] Kop, *supra* note 47; Rochelle Dreyfus and Susy Frankel, *From Incentive to Commodity to Asset: How International Law Is Reconceptualizing Intellectual Property*, 36 MICH. J. INT'L L., 557 (2015). *See also* Lemley, *supra* note 31.

common good. Further, data is a non-rivalrous resource[58], which should be non-excludable.[59] Empirical study shows that overgrazing and underinvestment is not a concern.[60] In the context of data, there is no need to prevent overuse or congestion, and to internalize externalities such as social benefits.[61] Conferring data ownership rights to certain categories of stakeholders bears significant risks for competition.[62] Currently, a justification for the allocation of exclusive property rights over data does not exist.[63] More exclusive rights - either property rights[64], data producer rights[65] or IP rights on data- may have an adverse effect on innovation.[66]

That said, the idea of property, its theory and justifications can be helpful when developing checks and balances within a novel

---

[58] Lawrence Lessig, *The Architecture of Innovation*, 51 DUKE L. J. 1783 (2002); Charles Jones & Christopher Tonetti, *Nonrivalry and the Economics of Data,* NBER WORKING PAPER NO. W26260 (Sept. 2019).

[59] For an explanation of the used economic terminology see: Roy D. Adams and Ken McCormick, *Private goods, club goods, and public goods as a continuum*, 45 REV. SOC. ECON. 192 (1982), www.jstor.org/stable/29769372.

[60] Kristofer Erickson, *Defining the Public Domain in Economic Terms – Approaches and Consequences for Policy* (May 9, 2016). ETIKK I PRAKSIS. NORDIC J. APPLIED ETHICS 61 (2016). *See also* Garret Hardin, *The Tragedy of the Commons*, SCIENCE (1968); Carol M. *Rose, The Several Futures of Property: Of Cyberspace and Folk Tales, Emission Trades and Ecosystems,* 83 MINN. L. REV., 964 (1998); Wolfgang Kerber, *A New (Intellectual) Property Right for Non-Personal Data? An Economic Analysis*, GEWERBLICHER RECHTSSCHUTZ UND URHEBERRECHT, INTERNATIONALER TEIL (GRUR INT), 11/2016, 989-999 (Oct.24, 2016). *See also* William M. Landes & Richard A. Posner, 18 *An Economic Analysis of Copyright Law*, THE JOURNAL OF LEGAL STUDIES 325 (1989), www.jstor.org/stable/3085624.

[61] Lemley, *supra* note 31.

[62] Jurcys et al., *supra* note 6.

[63] Drexl et al. *supra* note 6, Jurcys et al., *supra* note 6.

[64] Kenneth J. Vandevelde, *The New Property of the Nineteenth Century: The Development of the Modern Concept of Property*, 29 BUFF. L. REV. 325 (1980), https://digitalcommons.law.buffalo.edu/buffalolawreview/vol29/iss2/2; Van Erp *supra* note 32 and Lipton, *supra* note 50.

[65] Ivan Stepanov, *Introducing a property right over data in the EU: the data producer's right – an evaluation.* INT'L REV. OF LAW, COMPUT. & TECH 1 (2019).

[66] Yochai Benkler, *Free as the Air to Common Use: First Amendment Constraints on Enclosure of the Public Domain*, 74 N.Y.U. L. REV. (1999) 354; James Boyle, *The Second Enclosure Movement and the Construction of the Public Domain*, 66 L. CONTEMP. PROBS. 33, WINTER-SPRING 2003; James Bessen, *Patent Thickets: Strategic Patenting of Complex Technologies* (SSRN ELECTRONIC JOURNAL 2004) 10.2139/ssrn.327760; Drexl, Josef (2016) *supra* note 6; Yonida Koukio, The (R)evolutionary Impact of AI-Generated Work and Big Data on Intellectual Property Law and Commercialization (IP Osgoode, 2018); Thomas Hoeren, *A New Approach to Data Property?* (AMI 2018 / 2); Begoña Gonzales Otero, *Evaluating the EC Private Data Sharing Principles: Setting a Mantra for Artificial Intelligence Nirvana?* 10 JIPITEC 87 PARA 1. (2019), https://www.jipitec.eu/issues/jipitec-10-1-2019/4878; Hugenholtz, Data Property: Unwelcome Guest in the House of IP (2017); Ramalho, *Data Producer's Right: Power, Perils & Pitfalls* (Paper presented at Better Regulation for Copyright, Brussels, 2017), Max Planck Institute, *Position Statement of the Max Planck Institute for Innovation and Competition of 26 April 2017 on the European Commission's "Public consultation on Building the European Data Economy",* MAX PLANCK INSTITUTE FOR INNOVATION AND COMPETITION and Yu, *supra* note 1.

legal framework for data.[67] Questions about balancing general public
interests and legitimate private interest in machine learning training
data need to be addressed.[68] As is the case with copyright and training
data, we could think of principles that inform the development of
exceptions and limitations to *de facto* information property rights.[69]
Of user rights instead of exclusive ownership of key digital assets
necessary for a thriving, sustainable AI-ecosystem. Of digital
commons that effectively shelter general public interests in
information. A robust public domain fosters the fundamental human
right to freedom of speech and expression and promotes democracy,
innovation, cultural diversity, prosperity and a participative society.[70]
We should utilize basic dynamics from property theory to protect
public interest and revitalize the public domain.[71] In practice, the
public interest seems best served by data access, via the introduction
of a new right to process data (i.e. a freedom) in the context of AI and
machine learning, as an exclusion or limitation of (formally non-
existing) ownership rights in data.

   Accordingly, property theory and its normative, economic
and philosophical justifications can be a source of inspiration for the
political and scholarly debate about the nature and scope of a novel
right to process data for machine learning purposes.

   The main three species of property arrangement are common
property, collective property, and private property.[72] It should be
noted that these are not all-or -nothing alternatives. As stated by
canonical thinkers such as Plato, Aristotle, Rousseau, Locke, Marx
and Engels, Hobbes, Kant, Hegel, Hume, Rawls en Mill, and recently

---

[67] Lipton, *supra* note 50.

[68] *Id*.

[69] *See also* Lipton, *supra* note 50.

[70] James Boyle, *The Public Domain: Enclosing the Commons of the Mind*, ORANGE GROVE BOOKS 2008 236; David Lange, *Reimagining the Public Domain*, 66 LAW AND CONTEMPORARY PROBLEMS 463 (WINTER 2003), https://scholarship.law.duke.edu/lcp/vol66/iss1/13; Netanel, Neil Weinstock, *Why Has Copyright Expanded? Analysis and Critique in New directions in copyright law*; Fiona Macmillan, ed., EDWARD ELGAR, 2008; UCLA SCHOOL OF LAW RESEARCH PAPER NO. 07-34; Yochai Benkler*, Through the Looking Glass: Alice and the Constitutional Foundations of the Public Domain,* LAW AND CONTEMPORARY PROBLEMS 66 (2003); Paul Goldstein & Bernt Hugenholtz, *International Copyright: Principles, Law, and Practice* (3RD EDN, OUP 2013) 4, 5; Pamela Samuelson*, Challenges in Mapping the Public Domain,* in Lucie Guibault & Bernt Hugenholtz (eds), *The Future of the Public Domain: Identifying the Commons in Information Law*, KLUWER LAW INTERNATIONAL 7 (2006); Fiona Macmillan, *Arts festivals: Property, heritage or more?*; K Bowrey & M Handler (eds), *Law and Creativity in the Age of the Entertainment Franchise*, CAMBRIDGE UNIVERSITY PRESS (2013); Fiona Macmillan, *Many analogies, some metaphors, little imagination: the public domain in intellectual space* (2010), Erickson, *supra* note 60 and Kop, *supra* note 47.

[71] Lipton, *supra* note 50. See also Lessig, *supra* note 58.

[72] Jeremy Waldron, *Property and Ownership*, THE STANFORD ENCYCLOPEDIA OF PHILOSOPHY (SUMMER 2020 EDITION); Edward N. Zalta (ed.), (Forthcoming), https://plato.stanford.edu/archives/sum2020/entries/property/

Hayek, Nozick and Demsetz, there can always be variations in the amount of control a proprietor (read: the data holder) has -be it an individual, a company, a university or the state- over his resource.[73] Additionally, the privatized sphere can overlap with the public sphere. In other words, there are variations in the degree of liberties, rights and powers that a private owner had over the assets assigned to him.[74] These freedoms can be assigned to legal subjects, or to society as a whole. Extrapolating to our contemporary infosphere: private, common and collective property of data are extremes within the property spectrum. Between these extremes there is room for interjacent solutions such as our proposed right to process data.

Another important lesson we can draw from the cultural legacy of the old philosophers is that some resources need to be governed by common property rules for the wider social good.[75] And that the individual interest -provided that principles such as fairness, proportionality and subsidiarity[76] are met- must sometimes make room for the general interest. Locke's natural law labor theory of property, including the Lockean proviso, can also be interpreted as providing a justification for common property.[77] Moreover, private control over data does not exclude the right to process data of the community.[78] If necessary and desirable for the progress of society, the state can implement these new states and modalities of property.[79]

According to Hughes, tangible and intangible forms of property should be justified with both the normative and incentive

---

[73] *Id. See also* Harold Demsetz, 57 *Towards a Theory of Property Rights,* THE AMERICAN ECONOMIC REVIEW 347 (1967); Stepanov, *supra* note 65 and Janeček, *supra* note 35, p.17

[74] Waldron, *supra* note 72.

[75] *Id.*

[76] According to Renda, subsidiarity will become a choice between centralised, distributed and decentralised governance, through a "Europe as a Platform" approach. *See* Renda, *supra* note 52.

[77] Jukka Gronow, "John Locke, Adam Smith and Karl Marx's Critique of Private Property." *On the Formation of Marxism: Karl Kautsky's Theory of Capitalism, the Marxism of the Second International and Karl Marx's Critique of Political Economy*, BRILL, LEIDEN; BOSTON, 225 (2016)*,* www.jstor.org/stable/10.1163/j.ctt1w8h23p.19. Note the difference between the individuation of property and private property. *See also* Adam Mossoff, *Saving Locke from Marx: The Labor Theory of Value in Intellectual Property Theory* (Jan. 11, 2012), 29 SOCIAL PHILOSOPHY AND POLICY (2012; GEORGE MASON LAW & ECONOMICS RESEARCH PAPER NO. 12-02. Locke and Marx's labor theories of value are relevant, in particular the moment that all products will be created by smart machines instead of human labor. *Compare with* Joshua Glawson*, Labor and Property: Locke vs Marx*, https://medium.com/@JoshuaGlawson/labor-and-property-locke-vs-marx-b1b79f34193b.

[78] *See* den Hartogh, G. A., *Tully's Locke*, 18 POLITICAL THEORY 656 (1990), https://doi.org/10.1177/0090591790018004013.

[79] Eric Engle, *Karl Marx's Intellectual Roots in John Locke* (September 15, 2008); 7 POSTMODERN OPENINGS 29 (2011). *See als*o den Hartogh, *supra* note 78, Drexl, *supra* note 6 and Geiger, *supra* note 122.

based Lockean theory of labor and the Hegelian personality theory, that legitimizes property as an expression of the self.[80]

Kant provides an adequate justification of limited rights to possession, not of private property rights.[81] Kant's principles and arguments can be used to justify general usufructuary rights as qualified choses in possession *i.e.* as the right to use something including constructive enjoyment.[82]

VI. A RIGHT TO PROCESS DATA

Back to data processing. Transformative technology asks for a more layered, yet more efficient, regulatory property regime by separating the traditional bundle of property rights into its different components.[83] What's missing is a clear legal basis for the primary and secondary use of input data for machine learning purposes. Thus, we should develop the law a bit further. It is paramount that we create legal rules and infrastructures that allow access, sharing and re-use of data. Logically, this asks for a right to data processing that works in the context of AI and the Internet of Things (IoT). A right to process data that fits in the EU *acquis communautaire*.

The next step is to think about the legal basis of this right, its justification and underlying principles and about the nature and scope of this right. We need to agree on values and principles embedded in a legal framework for data that contains this right and find a balance between its purpose and the legitimate interests of relevant stakeholders. We have to make decisions about legal and technical guarantees embedded in the design[84] of European data access[85] and sharing mechanisms in which freedom and control are mixed.[86]

---

[80] Hughes, *The Philosophy of Intellectual Property*, 77 GEO. L. J., 287, 288 (1988). For the lack of personal expression in machine learning, see: Lemley *supra* note 41. In general, IP's normative sources Welfare, Fairness, Culture and Social Planning Theory have difficulties justifying granting copyright protection to data from an economic, cultural and philosophical perspective, *see* William Fisher, *Theories of Intellectual Property, New Essays in the Legal and Political Theory of Property*, CAMBRIDGE UNIVERSITY PRESS 1 (2001). *See also* William Fisher, *Theories of IP*, http://ccb.ff6.mwp.accessdomain.com/Maps/IPTheories.html (2019)

[81] *Id.*

[82] Kenneth R. Westphal, *Do Kant's Principles Justify Property or Usufruct* 5 JAHRBUCH FÜR RECHT UND ETHIK / ANNUAL REVIEW OF LAW AND ETHICS 141 (1997)

[83] Eli M. Salzberger, *Economic Analysis of the Public Domain*, in Lucie Guibault & Bernt Hugenholtz (eds), *The Future of the Public Domain: Identifying the Commons in Information Law*, KLUWER LAW INTERNATIONAL 27–59. (2006)

[84] John Wilbanks & Stephen H. Friend, *First, design for data sharing*, NATURE (2016). For further reading on data protection by design practices, *see* Bryan Casey Ashkon Farhangi, and Roland Vogl, *Rethinking Explainable Machines: The GDPR's 'Right to Explanation' Debate and the Rise of Algorithmic Audits in Enterprise* 34 BERKELEY TECH. L. J. (2019).

[85] Drexl et al. *supra* note 6

[86] Lessig, *supra* note 58.

Important work must be done at the intersection between these disciplines. This exercise brings the solution to our problem closer.

The Nature of this Right

Let us learn from history once more. As is the case for ownership and IP on the treasures of AI output, parts of the multi-layered Roman property paradigm can be relevant for a novel right to process data. This codified, *sui generis* legal right should include 2 liberties: (1) an unrestricted freedom of (beneficial) use of primary digital resources including the right to alter, and (2) the freedom of taking or producing the fruits, profits and avails.[87] A *ius utendi (usus) et fruendi (fructus)* with regard to data. Without a *ius abutendi*, so without the right (or freedom) to deny use and proceeds to another person, or to impair and destroy. Consequently, no complete ownership or property, no *pars dominium* or *proprietas* will emerge from this property arrangement. Thus, I suggest that the EC implements a *sui generis* right to process data in the upcoming Data Governance Act, that resembles an imperfect or *quasi usufruct,*[88] ownership like right (*ius abutendi* decoupled) governed by common property rules. Modelled as a supreme right to commons-based production in the public interest.

The scope of this right could extend over all types of data, from industrial data to consumer data and from private and personal data to public data.[89] A user right with access, that includes a primary or secondary data usage right, while maintaining privacy and data protection.[90] Flanked -where necessary- by sector-specific regulations e.g. the Medical Device Regulation (MDR)[91] and the Machinery Directive,[92] certification and self-regulation via tools such as the AI Impact Assessment.[93] A broader scope means more data openness.

---

[87] Roscoe Pound, 2 *Jurisprudence,* 115 (2000).

[88] *See* A.N. Yiannopoulos, *Usufruct: General Principles - Louisiana and Comparative Law*, 27 LA. L. REV. (1967). *See also Usufruct | law. Encyclopedia Britannica,* https://www.britannica.com/topic/usufruct, https://legal-dictionary.thefreedictionary.com/usufruct

[89] *Compare with* Drexl, *supra* note 6.

[90] *See also* Drexl *supra* note 6.

[91] Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices, amending Directive 2001/83/EC, Regulation (EC) No 178/2002 and Regulation (EC) No 1223/2009 and repealing Council Directives 90/385/EEC and 93/42/EEC (MDR). For a legal perspective on machine vision and medical AI, *see* Zach Harned, Matthew P. Lungren & Pranav Rajpurkar, Comment*, Machine Vision, Medical AI, and Malpractice,* HARV. J.L. & TECH. DIG. (2019).

[92] Directive 2006/42/EC of The European Parliament and of the Council of 17 May 2006 on machinery, and amending Directive 95/16/EC (Machinery Directive).

[93] Mauritz Kop, AI Impact Assessment | Netherlands, ,AIRECHT (2018), https://airecht.nl/blog/2018/ai-impact-assessment-netherlands. *See also* Mauritz Kop, AI

At the same time, various European Directives need to be amended or clarified in order to solve the problems identified above, and to take out the associated legal uncertainty.[94] In case existing laws contradict with certain Data Governance Act provisions, private international law's (IPR) conflict rule mechanisms such as *lex specialis, lex superior* and *lex posterior* can offer a solution. This legislative effort will have a positive effect on innovation and societal progress, in the sense of pushing the boundaries of knowledge in the scientific, technological, aesthetic, cultural and social areas.[95] Machine learning training datasets are the domain of this novel right to process data. Training, testing and validation *corpora* qualify as subject matter (object). No exclusive control over training data should be allowed. It should be free as the air we breathe, for common use.[96] This applies equally to hand-labelled, refined, augmented or annotated *corpora*. To encourage openness, the subject matter's scope could be (gradually) increased.

Our next task is to identify the data process rightsholder. Is everybody included, or should we distinguish between types of actors/data processors and their role? Do these legal subjects have to be controllers/owners and data (re-)users? In other words, ought it be constructed as an inclusive right? Are machines (who are no legal subjects but can have legal agency) included, or just machine learning enterprises, data companies, research institutions and the government?

Another principal aspect is the identity of the data holder/owner, who must allow the data to be accessed, shared and re-used. Is this everybody who *de facto* controls data, or should this be specified or narrowed?[97] And what if the data subject i.e. the "data processee" is the same entity as the rightsholder/data processor?

It is critical that definitions are crystal clear about the specific purpose of this right. Should the right to process only be valid in the context of machine learning and adjacent transformative technologies such as synergies between AI and blockchain? In the context of all

Impact Assessment available in English! Code of Conduct and Roadmap included, European AI Alliance (2019), https://ec.europa.eu/futurium/en/european-ai-alliance/ai-impact-assessment-available-english-code-conduct-and-roadmap-included

[94] Kop, *supra* note 4.

[95] Mark P. McKenna and Brett M/ Frischmann., *Comparative Analysis of Innovation Failures and Institutions in Context*, 57 HOUS. L. REV., VOL. 57, (2019); NOTRE DAME LEGAL STUDIES PAPER NO. 191211. *See also* Camilla Hrdy, *Challenging what we think we know about "market failures" and "innovation"*, https://writtendescription.blogspot.com/2020/03/challenging-what-we-think-we-know-about.html.

[96] Benkler, *supra* note 70.

[97] For further reading about finding the appropriate balance between protection and access in intellectual property, real property and mutually beneficial information policies, *see* Yu, *supra* note 1, 798–799.

sorts, types and purposes of machine learning, across each economic area, in each industrial sector? Should sharing always be mandatory and un-remunerated, without an opt-out or override?[98] Or should we build in fair remuneration in the form of a statutory fee for parties that harvested, labelled, annotated or augmented the data?[99] Should we create/build a permissionless space or implement formal legal requirements and written documents? Should duration of the right to process be perpetual or should it be a temporary right, limited in time? Should the territory be Europe or the universe? Decisions on the described modalities of access to process data must be contextual, pragmatic, rational and responsible.

It is key that the EC makes normative and contextual choices about access, competition and innovation. Healthy competition is an important driver of innovation. Different blends of IP regimes, IP alternatives, property arrangements, data protection rules and competition laws shape the type of innovation we want to produce.[100] To incentivize data sharing, the EC should clearly delineate the subject matter of all of the regimes.[101] While optimizing the innovation operating system, it should principally and pragmatically exclude objects from regimes that belong in other domains.[102] The construction and interaction between these legal disciplines, and other branches of law such as human and civil rights, consumer law and international trade law, stands at the basis of sustainable exponential innovation policy.[103]

Since every society has unique relationships between sources of law and a unique hierarchy of legislative, executive, and judiciary powers, the same legal concept could be qualified as impactful in one specific institutional context (or country) and in-efficient in another context.[104] Against this background, it is fruitful to investigate whether a EU right to process data for machine learning purposes

---

[98] *Compare with* Drexl *supra* note 6. Drexl on the price of access: "Where public interest or competition law justifies access, a cost-based approach to assessing the royalty rates seems most appropriate."

[99] *See* Benjamin Sobel, *Artificial Intelligence's Fair Use Crisis*, COLUMBIA JOURNAL OF LAW & THE ARTS(2017); Hoeren, *supra* note 66 and Ramalho, *supra* note 66.

[100] McKenna, Mark P. and Sprigman, Christopher Jon, *What's In, and What's Out: How IP's Boundary Rules Shape Innovation* (February 19, 2016), HARVARD JOURNAL OF LAW AND TECHNOLOGY, VOL. 30, NO. 2, 2017.

[101] *Id.*

[102] *Id.*

[103] See also: Mauritz Kop, Beyond AI & Intellectual Property: Regulating Disruptive Innovation in Europe and the United States – A Comparative Analysis, https://law.stanford.edu/projects/beyond-ai-intellectual-property-regulating-disruptive-innovation-in-europe-and-the-united-states-a-comparative-analysis/

[104] Mauritz Kop, *Copyright, Machine Learning & Comparative Law* (June 28, 2020), *See also* Danny Pieters & Bert Demarsin, *Rechtsvergelijking, De uitdagende wereld van het recht*, ACCO, UITGEVERIJ (2019), Uwe Kischel, *Comparative Law*, OXFORD UNIVERSITY PRESS (2019).

could be easily transplanted to the U.S. And vice versa, to assess the impact of transplanting the fair use 4-factor test to the EU, as an innovation friendly limitation to copyright within the boundaries of the Berne three-step-test.[105] Contextual comparison of law offers legislators a powerful instrument to detect undesired consequences of legal reform in advance.[106]

VII. ARGUMENTS IN FAVOR

Arguments in favor of implementing a right to process data in the EU are numerous. First, for cohesion, clarity and speed, it is important to orchestrate data access regulation under 1 umbrella: The Data Governance Act. A specific legal regime achieves legal harmonization in the internal market.[107] Second, this novel right accomplishes primary EC Strategy for Data policy goals. It incentivizes access, enhances sharing and facilitates re-use of data. Third, even distribution of primary resources for machine learning creates a fair and level playing field where everybody has access to the same datasets. This encourages healthy competition and innovation. Fourth, legal certainty and trust are key incentives for incumbents to share data.[108] A clear legal framework for well-functioning, tailor made legal-technical system that enhances data access, sharing and re-use -including a technological infrastructure that embeds EU values, norms, principles and standards- will provide this much sought after legal certainty and trust.[109] Legal certainty and trusted technology also lowers transaction costs. Fifth, codification of a right to process data can be justified from a legal, economic and philosophical perspective. Sixth, besides exceptions and limitations to exclusive rights that create breathing room and freedom to operate, we could endorse and justify data access rights on the basis of the abuse of the market dominance doctrine (e.g. exploitative abuse in digital markets that makes fair competition impossible, lowers product quality and impedes on consumer welfare) an approach occasionally

---

[105] Kop, *supra* note 47. *See also* Niva Elkin-Koren & Neil Weinstock Netanel, *Transplanting Fair Use across the Globe: A Case Study Testing the Credibility of U.S. Opposition* (May 11, 2020), HASTINGS LAW JOURNAL, (Forthcoming), UCLA SCHOOL OF LAW, PUBLIC LAW RESEARCH PAPER NO. 20-15.

[106] Kop, *supra* note 104.

[107] Drexl et al., *supra* note 6.

[108] AI does not need incentives. IP incentives to share lead to overlapping rights and IP thickets that impair innovation. Inventing new layers of rights, be it IP, ownership, or data producer right, is not the right approach. *See also* Kop, *supra* note 47.

[109] For further reading on innovation incentives and allocation mechanisms, *see* Daniel J. Hemel & Lisa Larrimore Ouellette, *Innovation Policy Pluralism*, 128 YALE L.J. 544 (2019), https://digitalcommons.law.yale.edu/ylj/vol128/iss3/1. *See also Unboxing the Innovation Policy Toolkit with Professor Lisa Ouellette*, https://www.youtube.com/watch?v=fiZ7cD3AHwg.

employed as a concept of antitrust law.[110] Seventh, as machine
learning matures, data should be as open (freedom to operate) and
secure as possible. Open standards provide mutual benefit where
closed standards encourage monopoly of wealth. Openness and
security are not contradictory terms. They are sides of the same medal
similar to data protection and privacy.[111] More openness strengthens
the public domain, human rights and democracy and encourages
exponential innovation, prosperity and legal certainty. There is no
tragedy of the commons.[112] Lastly, a comprehensive open data policy
that includes a right to process data will result in a network effect,
which means even more incumbents will be comfortable to share,
access and re-use data.[113] Positive network externalities together with
trust and legal certainty will incentivize free, unencumbered data
circulation.[114]

VIII. ARGUMENTS AGAINST

The described advantages of an open data policy, such as
transparency and strengthening economic growth must be weighed
against risks such as misinterpretation and misuse of data, perceived
competitive disadvantages, security challenges and privacy
concerns.[115] Disadvantages, unintended consequences and costs must
be thoroughly investigated in the Data Governance Act impact
assessment. Impact assessments are mandatory for EC initiatives
expected to have significant economic, social or environmental
impacts.[116] The Data Governance Act impact assessment will also

---

[110] For the interplay between access to data and competition law, *see* Kop, *supra* note 4,
Marco Botta & Klaus Wiedemann, *Exploitative Conducts in Digital Markets: Time for a
Discussion after the Facebook Decision*, 10 JOURNAL OF EUROPEAN COMPETITION LAW &
PRACTICE, 465 (OCTOBER 2019), https://doi.org/10.1093/jeclap/lpz064 and Josef Drexl,
Politics, digital innovation, intellectual property and the future of competition law,
CONCURRENCES REVIEW 4, 2-5 (2019).
[111] Youtube, *supra* note 48
[112] *See also*, Hardin, *supra* note 60.
[113] For further reading about open data from a U.S perspective see Lawrence Lessig, *Code
and Other Laws of Cyberspace* (CODE 2.0 2006). For a European perspective, *see* Van
Eechoud, *A publisher's intellectual property right: Implications for freedom of expression,
authors and open content policies*, OFE ACADEMIC PAPER 41 (2017).
[114] Peter Stone et al., *"Artificial Intelligence and Life in 2030." One Hundred Year Study on
Artificial Intelligence: Report of the 2015-2016 Study Panel*, STANFORD UNIVERSITY,
STANFORD, CA (SEPTEMBER 2016), http://ai100.stanford.edu/2016-report. *See also* Kop,
supra note 47.
[115] Anneke Zuiderwijk and Marijn Janssen, *The negative effects of open government data -
Investigating the dark side of open data*, ACM INTERNATIONAL CONFERENCE PROCEEDING
SERIES (2014)
[116] *See* European Commission, The need for impact assessments,
https://ec.europa.eu/info/law/law-making-process/planning-and-proposing-law/impact-
assessments_en.

examine the correct application of the principles of subsidiarity and proportionality, according to which there should be no EU intervention when an issue can be dealt with effectively by EU Member States and EU action should not exceed what is necessary to achieve the objectives.[117]

In addition to inducing a Pareto improvement, modalities of mandatory data sharing may have detrimental effects.[118] Relevant parties may have spent significant amount of money in assembling high quality datasets and expect a return on investment. Data brokers will have to adapt and reinvent themselves. Certain stakeholders may have legitimate economic, privacy and security interests that do not rhyme properly with our novel right. For example, in the event a contribution to the creation of the data was made, in case of hi risk area data, or when there are concerns about identification, security, utility and competition.[119] To solve this, the judicial concept of the right to process data could develop doctrines to exclude subject matter from its domain because they belong to other regimes.[120] A right to process data should however be the main rule and not the exception. Another option -in exceptional cases duly justified- could be restricted data-sharing arrangements or data access mechanism that are less open.[121]

In case it applies to machine learning datasets, the Berne Convention's three-step-test can be used to determine whether limiting exclusive rights to commercial exploitation is proportional, reasonable and fair, in the light of user interests, public interest or other cultural, social or economic interests.[122] The outcome of that assessment could be that limiting exclusive rights, or introducing non-exclusive data user rights is not allowed in certain instances.[123]

A government regulation that introduces a right to process data could conflict with the principle of freedom of contract,

---

[117] *Id.*

[118] OECD Report, *supra* note 3.

[119] OECD Report, *supra* note 3.

[120] McKenna & Sprigman, *supra* note 100.

[121] OECD Report, *supra* note 3. *See also* Frederik Zuiderveen Borgesius, and Mireille M. M. van Eechoud, and Jonathan Gray, *Open Data, Privacy, and Fair Information Principles: Towards a Balancing Framework*, BERKELEY TECH. L. J. (2016); INSTITUTE FOR INFORMATION LAW RESEARCH PAPER NO. 2015-04; AMSTERDAM LAW SCHOOL RESEARCH PAPER NO. 2015-46.

[122] The three-step test is a flexible balancing tool. *See* Christophe Geiger, Daniel Gervais and Martin Senftleben, *Understanding the 'three-step test'*, in Daniel J. Gervais (ed), *International Intellectual Property: A Handbook of Contemporary Research*, EDWARD ELGAR PUBLISHING, 189 (2015).

[123] Intellectual property is however far from absolute and can be limited by the interest of society at large. *See* Christophe Geiger, *Reconceptualizing the Constitutional Dimension of Intellectual Property – An Update* Forthcoming in: P. Torremans (ed.), "Intellectual Propertyand Human Rights", 4TH ED., AUSTIN/ BOSTON/ CHICAGO/ NEW YORK, THE NETHERLANDS,KLUWER LAW INTERNATIONAL (2020).

especially in B2G and B2B settings. Freedom of contract is the process in which stakeholders conclude contracts, based on mutual agreement and free choice.[124] This doctrine states that legal subjects have the right to legally bind them without government restrictions.

Settled market players -either from Europe, USA or Asia-who benefit the most from status quo will lobby against open data policy.[125] I suggest caution in having the most ambitious parts of the Data Governance Act strand on a negotiation between stakeholders with completely different viewpoints. This will lead to economic stagnation. It is vital that the EC does not linger in an endless weighing of interests but that it is assertive and future-oriented in achieving its policy targets. The described advantages of an open data policy evidently outweigh the drawbacks. By doing nothing Europe will fall behind globally.

IX. ALTERNATIVES

The European legislature must develop new concepts and metaphors for law and policy that promote quick access to machine learning training data in a pro-active and pro-competitive way.[126] In addition to making good laws, policymakers can help to remove barriers for data sharing by setting up a number of concrete initiatives. These are especially relevant until the Data Governance Act comes upon us. First, the state can issue knowledge vouchers to early AI start-ups so that they acquire the budget to be properly advised and become GDPR compliant.[127] By way of government funding for the purpose of healthy competition, innovation and a fair level playing field. Second, the state can stimulate knowledge transfer in public-private partnerships such as the NL AIC.[128] Think of workshops on data sharing and access or drawing up and automating license contracts for data sharing. Contracts offer no universal salvation in this setting, because of differences in bargaining power. This involves a risk of parties to data sharing agreements being forced to accept unbalanced standard clauses. Third, data exchange protocols and standards for interoperability must be implemented, together with

---

[124] *See, e.g.* Péter Cserne, *Freedom of Contract and Paternalism: Prospects and Limits of an Economic Approach* (excerpt), in *Freedom of Contract and Paternalism: Prospects and Limits of an Economic Approac*h, PALGRAVE MACMILLAN (2012)
and Micklitz, Hans-W., *On the Intellectual History of Freedom of Contract and Regulation,* EUI DEPARTMENT OF LAW RESEARCH PAPER NO. 2015/09.

[125] Kop, *supra* note 44.

[126] Stone et al., *supra* note 114. *See also* Drexl, *supra* note 6.

[127] Mauritz Kop*, Data delen als voorwaarde voor een succesvol AI-ecosysteem*, AIRECHT (2020), https://airecht.nl/blog/2020/data-delen-voorwaarde-voor-succesvol-ai-ecosysteem.

[128] *See* Nederlandse AI Coalitie (NL AIC).

associated IEC, ISO and NEN standards.[129] Fourth, policymakers can take citizens and businesses by hand and provide guidance on the current possibilities with regard to data processing of the FFD and the GDPR in practice.

Finally, data altruism is an inventive solution to a shortage of relevant data. It can be applied alongside our novel right to process data. A government institution that takes care of questions and formalities regarding the GDPR and the FFD could facilitate and register data donorship by subjects that possess or control data, e.g. citizens, universities, companies and the government. Data altruism enables data donation of mobility data, energy data or health information for open data in the public interest, based -*inter alia*- on legal concepts such as creative commons and the public domain, as well as opt in & opt out mechanisms. In practice, by means of an accessible, streamlined data donor codicil. "Yes, No or Someone else decides for you."

Synchronous to harmonized legislation and sustainable innovation-stimulating initiatives, the social impact of digital transformation can be regulated by the architecture of digital systems. Imagine a collaboration of blockchain and AI that creates synergetic effects. For example, through a combination of centralized and decentralized data management systems that connect cloud and edge.[130] Responsible tech and European ethically aligned design should become the norm. Embedding values in the design of systems should become a fundamental starting point of our data paradigm.

X. Conclusion

For the European data-driven economy to function as efficiently as possible, a new and as yet unused term must be introduced to the field of AI & law: the right to process data for machine learning purposes.
Europe is now at a crucial juncture in deciding how to deploy data driven technologies in ways that encourage democracy, prosperity and the well-being of European citizens. Normative preferences about how related technology laws ought to be designed should define the best sustainable exponential innovation policy. These preferences are dynamic and contextual. The upcoming Data Governance Act provides a major window of opportunity to change the story. In this respect, it is key that the EC takes firm action, removes overbearing policy and regulatory obstacles, strenuously harmonizes relevant

---

[129] Kop, *supra* note 4.
[130] Compare to Jurcys et al. *supra* note 6.

legislation, and provides concrete incentives and mechanisms for access, sharing and re-use of data.[131]

To make AI and machine learning thrive, we should critically re-examine the applicability and scope of intellectual property rights to data, including copyrights, *sui generis* database rights and trade secrets. The article argues that exclusive *de facto* possession or control over machine learning input training, testing and validation datasets hinders healthy competition, a fair level playing field and rapid European innovation. The article rejects exclusive legal ownership rights over autonomously machine generated non-personal data, including AI made creations and inventions: this output belongs to the public domain. Machines do not need incentives, people need freedom of expression and businesses need freedom to operate.

Synchronous to harmonized legislation, the social impact of digital transformation can be balanced and regulated by the architecture of digital systems.[132] Embedding values in design should become a fundamental starting point of our data paradigm.

Data has become a primary resource that should not be enclosed or commodified *per se*, but used for the common good. Commons based production and data for social good initiatives should be stimulated by the state. We need not to think in terms of exclusive, private property on data, but in terms of rights and freedoms to use, (modalities of) access, process and share data. Whether or not data as digital assets are ultimately admitted to the *numerus clausus* of legal objects i.e. acknowledged as subject matter eligible for private ownership, or whether other modalities of property are being developed, it is clear that there should also be exceptions to (*de facto*, economic or legal) ownership claims on data that provide user rights and freedom to operate in the setting of AI model training.

The article concludes that this exception is conceivable as a legal concept analogous to a *quasi*, imperfect *usufruct* in the form of a right to process data for machine learning purposes. A combination of *usus* and *fructus* (*ius utendi et fruendi*), not for land but for primary resource data. A right to process data that works within the context of

---

[131] Stone et al., *supra* note 114. Contextual comparison of law and impact assessments offer legislators a powerful instrument to detect undesired consequences of legal reform in advance.

[132] *See also* Mauritz Kop, *Shaping the Law of AI: Transatlantic Perspectives*, TTLF WORKING PAPERS NO. 65, STANFORD-VIENNA TRANSATLANTIC TECHNOLOGY LAW FORUM (2020), https://law.stanford.edu/publications/no-65-shaping-the-law-of-ai-transatlantic-perspectives/ and Mauritz Kop, *Regulating Transformative Technology in The Quantum Age: Intellectual Property, Standardization & Sustainable Innovation*, 2 TTLF NEWSLETTER ON TRANSATLANTIC ANTITRUST AND IPR DEVELOPMENTS STANFORD-VIENNA TRANSATLANTIC TECHNOLOGY LAW FORUM, STANFORD UNIVERSITY 2020, https://law.stanford.edu/publications/regulating-transformative-technology-in-the-quantum-age-intellectual-property-standardization-sustainable-innovation/.

AI and the Internet of Things (IoT), and that fits in the EU *acquis communautaire*. Such a right makes access, sharing and re-use of data possible, and helps to fulfil the European Strategy for Data's desiderata.